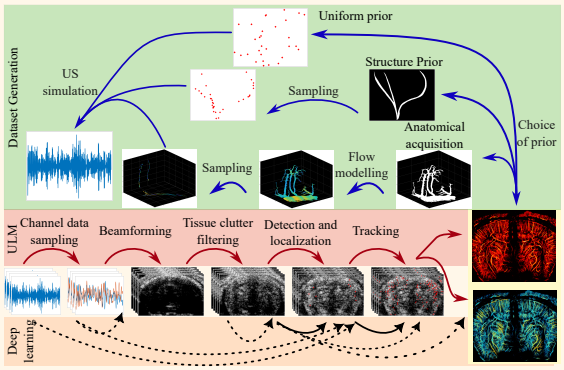
Deep Learning in Ultrasound Localization Microscopy: Applications and Perspectives

Brice Rauby, *student member, IEEE*, Paul Xing, *student member, IEEE*, Maxime Gasse, and Jean Provost *Member, IEEE*

***Abstract*— Ultrasound Localization Microscopy (ULM) is a novel super-resolution imaging technique that can image the vasculature *in vivo* at depth with resolution far beyond the conventional limit of diffraction. By relying on the localization and tracking of clinically approved microbubbles injected in the blood stream, ULM can provide not only anatomical visualization but also hemodynamic quantification of the microvasculature of different tissues. Various deep-learning approaches have been proposed to address challenges in ULM including denoising, improving microbubble localization, estimating blood flow velocity or performing aberration correction. Proposed deep learning methods often outperform their conventional counterparts by improving image quality and reducing processing time. In addition, their robustness to high concentrations of microbubbles can lead to reduced acquisition times in ULM, addressing a major hindrance to ULM clinical application. Herein, we propose a comprehensive review of the diversity of deep learning applications in ULM focusing on approaches assuming a sparse microbubbles distribution. We first provide an overview of how existing studies vary in the constitution of their datasets or in the tasks targeted by deep learning model. We also take a deeper look into the numerous approaches that have been proposed to improve the localization of microbubbles since they differ highly in their formulation of the optimization problem, their evaluation, or their network architectures. We finally discuss the current limitations and challenges of these methods, as well as the promises and potential of deep learning for ULM in the future.**



***Index Terms*— Deep Learning, Neural Network, Super-resolution, Ultrasound Localization Microscopy**

I. Introduction

Inspired from development of super resolution imaging techniques in optical microscopy [1, 2], Ultrasound Localization Microscopy (ULM) leverages the detection and localization of individual microbubbles injected into the bloodstream to overcome the diffraction limit in ultrasound imaging [3, 4]. ULM can reconstruct portions of the vascular tree at depth, *in vivo*, and with a resolution on the order of a tenth of the imaging wavelength, thus partially alleviating the trade-off between penetration depth and resolution [5]. ULM has also been extended to 3D imaging either using fully addressed matrix arrays [6, 7, 8], multiplexed matrix arrays [9] or row-column arrays [10, 11, 12]. ULM proof-of-concepts in pathological animal models have shown, e.g., the characterization of vascular function impairments in AD mice models [13] and between early phases of ischemic and hemorrhagic strokes in mice models [14]. Novel sequences also enable the extraction of dynamic quantities in ULM such as pulsatility imaging in the brain [15, 16], cardiac imaging [17, 18] or functional imaging [19, 20] by using high microbubble detection rates and retrospective gating. Singular microbubble behaviors have also been leveraged to highlight specific structures, such as glomeruli [21, 22]. Applications in humans have also been proposed for aneurysm imaging in the brain [23], breast [24, 25] or pancreas cancer imaging [26], lymph node metastatic cancer [27], kidney [21, 26, 28], prostate [29], lower limb muscle

This work was supported in part by the Institute for Data Valorization (IVADO), in part by the Canada Foundation for Innovation under Grant 38095, in part by the Canadian Institutes of Health Research (CIHR) under Grant 452530, in part by the New Frontiers in Research Fund under Grant NFRFE-2018-01312 and in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant RGPIN-2019-04982. Further support came from TransMedTech Institute, Fonds de recherche du Québec - Nature et technologies, Quebec Bio-Imaging Network, and Canada First Research Excellence Fund.

B. Rauby is the Department of Engineering Physics, Polytechnique Montréal, Montréal, QC H3T 1J4, Canada, and Mila-Quebec artificial intelligence institute, Montréal, QC H2S 3H1, Canada (email: brice.rauby@polymtl.ca)

P. Xing is with the Department of Engineering Physics, Polytechnique Montréal, Montréal, QC H3T 1J4, Canada (e-mail: paul.xing@polymtl.ca)

M. Gasse is with ServiceNow, Montréal, QC H2S 3G9, Canada, the Department of Computer Engineering and Software Engineering, Polytechnique Montréal, Montréal, QC H3T 1J4, Canada, and Mila-Quebec artificial intelligence institute, Montréal, QC H2S 3H1, Canada (e-mail: maxime.gasse@servicenow.com)

J. Provost is with the Department of Engineering Physics, Polytechnique Montréal, Montréal, QC H3T 1J4, Canada, and the Montreal Heart Institute, Montréal, QC H1T 1C8, Canada (email: jean.provost@polymtl.ca)

Highlights

- **Existing deep learning approaches in ULM vary in the constitution of their dataset and in the problem they address. We review these approaches and compare their dataset constitution.**
- **Several approaches have focused on improving microbubble localization. We compare these studies in their evaluation, their formulation of the optimization problem and their network architecture.**
- **This review compares existing deep learning approaches in ULM and provides insights on future research directions.**

[30], liver imaging [26], vasa vasorum of the carotid wall [31], and testicular microcirculation [32].

However, ULM also faces several inherent challenges. First, imaging the entire vascular tree with current methods requires impractically long acquisition times, since tens of minutes would be needed to have a single microbubble flow in each capillary at typical concentrations [33]. Second, ULM is degraded by skull aberration in brain imaging, clutter and cardiac motion in cardiac imaging, and tissue motion in general [34]. Finally, clinical translation of ULM can be challenging due to the large amount of data to acquire and process, often in the range of hundreds of gigabytes, and the processing time that can take several hours for a single acquisition [34].

Deep learning algorithms excel at signal processing tasks, driven by the increasing availability of computational power and large-scale datasets [35]. Since AlexNet [36] reduced the top-5 error rate on the ImageNet Large Scale Visual Recognition Challenge from 26.1% to 15.3% in 2012 [37], subsequent deep learning models have further decreased this error to only 3.6% within three years [38], increasing the popularity of deep learning for computer vision. Larger datasets and advances in model architectures and training procedures have since enabled deep learning methods to address more challenging tasks such as object detection[39], multi-instance segmentation [40] or image generation [41, 42]. Advances in one specific domain often translate into other domains as well, with several key components often reusable across different tasks, like optimization algorithms [43, 44], normalization layers [45, 46], activation functions [47, 48] or backbone building blocks [49, 50]. For example, transformers originally developed for Natural Language Processing (NLP) [50] have later been applied to image processing tasks with great success [51]. Foundation models, which are also originating from the field of NLP, [52], are very large deep learning models that are pre-trained on vast amounts of data, which can be re-used either as-is or with little fine-tuning to address new tasks in related or even different domains. Such foundation models are now widespread in computer vision [53], and can be applied to segment new images even from unrelated distributions [54]. In medical image analysis, foundation models trained on a sufficiently large dataset combining different modalities can segment regions of interest with better generalization and accuracy than specialized, domain-specific models [55]. The current performance of deep learning models makes them attractive for processing and analyzing medical images, and the application of future methods from other domains further enhances their potential.

In recent years, several works have investigated deep learning methods as a way to improve ultrasound imaging, with notable successes in beamforming [56, 57, 58, 59, 60], and clutter suppression in Contrast-Enhanced Ultrasound (CEUS) [61]. For recent reviews of deep learning methods in a general ultrasound settings, see [62, 63]. In ULM, deep learning methods have served several purposes such as reducing processing [64, 65, 66] or acquisition times [64, 65, 67], enhancing image quality [20, 64, 66, 68], improving blood velocity estimation [20, 69], and increasing robustness to challenging experimental settings such as increased microbubble concentrations [20, 67, 70] or phase aberrations [71]. In this review, we focus on deep learning methods specific to ULM, which leverage the presence and the sparsity of microbubble echoes in the ultrasound signal. Our objective is to cover and put into perspective three essential aspects of deep learning algorithms in ULM: dataset constitution, the range of targeted tasks, and, using the example of microbubble localization, the variations in formalism for a single task.

Deep learning methods can be integrated at various stages of the ULM processing pipeline. To provide an overview of such deep learning applications, we consider the following pipeline for ULM (also depicted in Fig 2):

- Channel data sampling,
- Beamforming,
- Tissue clutter filtering,
- Microbubble detection,
- Localization,
- Tracking,
- Accumulation of trajectory statistics to form vascular maps

Additional steps such as aberration correction, motion correction, additional filtering, denoising, or post-processing of the trajectories are also discussed in this review when deep learning approaches specific to ULM have been proposed.

In section II, we discuss the different existing approaches related to generating labelled datasets used either in training or evaluation of deep learning-based ULM methods. In section III, we review the different application stages of deep learning in the ULM pipeline. In section IV, we focus on the localization stage of the ULM pipeline, where most deep learning approaches have been applied. Finally, in section V, we summarize this review with an overview of the successes, limitations, and open challenges for deep learning-based ULM methods.

II. Generation of labelled datasets

Dataset constitution is a critical step that impacts both model parameter optimization and performance evaluation.

Existing acoustic field simulators enable the creation of realistic ultrasound echoes from microbubble positions, thereby facilitating the creation of *in silico* datasets, which partially mitigate the limited availability and the lack of ground truth of in vivo datasets. Literature on domain adaptation suggests that more realistic simulations, leading to reduced domain shift, may facilitate *in vivo* applications [72]. Driven by these theoretical insights, the ULM community has strived to produce highly realistic simulations allowing for *in vivo* applications of models trained *in silico*. Simulations can also be used for the evaluation of ULM methods, as done in recent benchmarking efforts such as the Ultra-SR challenge [73], and PALA [74], described hereafter. Compared to other computer vision domains, where dataset collection require costly manual annotations and often results in the creation of large-scale publicly available datasets [37, 75, 76], ULM training datasets tend to be smaller in scale and designed to match the imaging parameters of one or a few studies. Diversity in simulation models and their underlying hypotheses leads to discrepancies between datasets used in different studies, making model comparisons challenging. Larger-scale ULM datasets that target more diverse applications and a broader scope, could reduce redundant efforts in dataset generation while enhancing *in vivo* model performance and facilitating inter-study comparisons. The datasets used for deep learning in ULM present comparable challenges and characteristics regardless of the task addressed by the proposed model. In this section, we review and compare existing approaches for dataset generation based on simulations. We also review methods that directly learn from *in vivo* data, which address the domain shift that can exist between training simulations and *in vivo* applications.

A. Formalism

ULM processing can be formulated as recovering multiple microbubble positions, y, from an ultrasound signal, x. In a supervised learning context, x represents the input data used by the model, while y denotes the target labels that we aim to estimate. Hence, ULM can be defined as the estimation of the probability of microbubble positions from the given ultrasound data, which corresponds to modeling the posterior distribution $p(y|x)$. With the same notation, the constitution of a dataset can be formulated as sampling a collection of ultrasound signals with corresponding microbubble positions $D = \{(x_i, y_i)\}$ from the joint distribution, $(x_i, y_i) \sim p(x, y)$. Prior knowledge regarding microbubble positions can be expressed by formulating assumptions on the marginal distribution $p(y)$, referred as prior distribution. Ultrasound physics and simulation models describe the conditional probability $p(x|y)$, which represents the likelihood of an ultrasound signal, x, given microbubble positions, y. Using the Bayes rule, the joint probability $p(x, y)$ can be decomposed to highlight the roles of the prior and the conditional probability:

$$p(x, y) = p(x|y)p(y).$$

An essential assumption of supervised learning is that the training set, $D_{\text{train}} = \{(x_{\text{train},i}, y_{\text{train},i})\}$, which is sampled from the distribution $p_{\text{train}}(x, y)$, and the test set, $D_{\text{test}} = \{(x_{\text{test},i}, y_{\text{test},i})\}$, which is sampled from $p_{\text{test}}(x, y)$, are independent and identically distributed (i.i.d.). Thus, the i.i.d. hypothesis implies that $p_{\text{train}} \sim p_{\text{test}}$. In practice, training on simulation and testing on *in vivo* data causes p_{train} and p_{test} to differ, which limits the validity of the i.i.d. hypothesis. Some level of realism in the generation of p_{train} is crucial and depends on both the assumptions regarding the prior distribution $p_{\text{train}}(y)$, and the validity of the underlying simulation model of $p_{\text{train}}(x|y)$. It is also important to evaluate deep learning models not only *in vivo* but also on i.i.d. simulated test data to disentangle the impact of realistic datasets from model expressive power. Since *in vivo* evaluation measures both the dataset quality and its expressive power, i.i.d. evaluation is critical to assess the model's capacity to learn and address the targeted task independently of the simulation quality.

B. Prior probability p(y): label generation

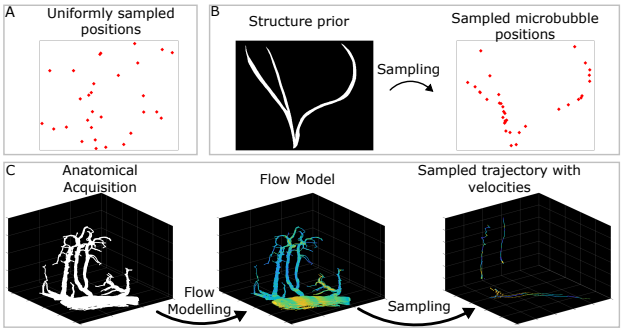


Fig. 1. Representation of the different types of label generation strategies illustrating sampling from different prior distribution. A: Uniform sampling of independent frames. B: Structure based sampling of independent frames. C: Trajectory sampling based on anatomical acquisition and simulated flow from [77]

In this section, we examine existing methodologies for modeling microbubble distributions, the assumptions regarding the simulation prior $p_{\text{train}}(y)$ and the implications regarding the evaluation on i.i.d. datasets and *in vivo*. Existing studies are listed in Table I with the corresponding prior distribution and temporal context. Depending on the deep learning model task (e.g., single-frame localization, tracking, velocity estimation), and the use of temporal context, y can either represent positions in single, independent frames [64, 88] or trajectories across multiple frames [67, 81, 89].

1) Single-frame simulations: In single-frame simulations, some approaches have used spatially uniform sampling [64, 68, 69, 79, 84, 87, 90], while others have employed a spatial distribution of scatterers conditioned by a given structure, often mimicking vasculature [64, 66, 79, 88]. The structure used to condition the prior have been handcrafted [66, 74], simulated [88], based on ULM acquisitions [64, 79], or derived from acquisitions from other modalities [67, 79, 89]. Using a uniform spatial prior benefits from being high-entropy, which may reduce biases when working with unseen vascular distributions. It also allows for greater diversity and larger dataset sizes. However, a uniform distribution does not accurately

TABLE I
COMPARISON OF SIMULATED DATASET PROPERTIES FOR DEEP LEARNING APPROACHES IN ULM

Deep learning method	Temporal context	Training MB prior	Testing MB prior	Structural Prior	Ultrasound Simulation
Blanken et al. [78]	Independent frames	Uniform	Uniform	N.A.	Non-linear propagation and response
Chen et al. [79]	Independent frames	Uniform and Structure	Structure	CAM	Field II [80]
Chen et al. [81]	Simulated flow	Structure	Structure	CAM	Field II [80] and convolution with Gaussian PSF
Gu et al. [82]	Independent frames	Structure	Structure	Artificial Vasculature	Convolution with Gaussian PSF
Hahne et al. [66]	Independent frames from simulated flow	PALA [6]	PALA [6]	Artificial structure	Verasonics Research Ultrasound Simulator
Liu and Almekkawy [83]	Independent frames	Uniform	Structure	Artificial vasculature	Convolution with Gaussian PSF
Liu et al. [64]	Independent frame	Uniform	Structure	Vasculature from *in vivo* acquisition	Convolution with Gaussian PSF
Luan et al. [84]	Independent Frames	Uniform	Structure	Artificial Vasculature	High-resolution convolution with Gaussian PSF
Milecki et al. [67]	Simulated flow	Structure	Structure	Vasculature from *ex vivo* acquisition	SIMUS [85]
Pustovalov et al. [86]	Displacement map	Structure	Structure	CAM and Artificial Structure	Convolution with experimental PSF
Shin et al. [20]	Simulated flow	Uniform	Uniform	N.A.	Convolution with learned distribution of PSF
van Sloun et al. [62]	Independent frames	Uniform	Structure	Artificial Vasculature	Convolution with Gaussian PSF
Youn et al. [69]	Independent frames	Uniform	Uniform	N.A.	Field II [80]
Yu et al. [87]	Independent frames	Uniform	Uniform	N.A.	Convolution with Gaussian PSF
Zhang et al. [68]	Independent frames	Uniform	Uniform	N.A.	Convolution with Gaussian PSF
Zhang et al. [70]	Independent frames	Uniform	Structure	Artificial Vasculature	Convolution with Gaussian PSF

model microbubble positions constrained to blood vessels, potentially reducing the validity of the i.i.d. hypothesis when testing *in vivo*. Datasets based on a given structure allow evaluation using standard ULM metrics typically used for *in vivo* evaluation, such as separation power or full width half maximum, often reported in literature [74, 91]. To combine the advantages of reduced bias in training while allowing the use of standard ULM metrics in evaluation, some approaches have used uniform sampling for the generation of the training set and vasculature-based sampling for *in silico* testing and evaluation.

2) Multi-frame simulations: To generate realistic microbubble trajectories, multi-frame simulations often use frameworks composed of several stages. This typically involves a defined structure and a flow model conditioned by the structure properties to generate realistic trajectories and velocities. Microbubbles are randomly seeded within the structure, and their trajectories are computed using the physical flow model. The underlying structures and flow models can vary depending on the approach. For example, Belgharbi et al. proposed a simulation framework based on mouse brain vascular structures acquired by 2-photon microscopy (2PM) [77]. The 2PM-acquired vascular structure was segmented and converted into a graph model using an existing framework [92]. Vessel radii from the segmentation were stored as features of the graph nodes, and a Poiseuille flow model was used to determine the velocity of randomly generated microbubbles. Chen et al. [81] used a binarized chorioallantoic membrane (CAM) dataset of chicken embryos obtained through optical microscopy, and mouse and rat brains obtained through ULM, to generate graphs and simulate flowing microbubbles. Using the same CAM dataset, Pustovalov et al. [86] proposed to use displacement maps to generate microbubble motions. To address the dataset size limitations inherent in methods relying on *in vivo* acquisitions of vasculature, Lerendegui et al. proposed a framework to generate vascular structure [93] using a recursively-generated simulation framework. Flow and pressure were then simulated based on Navier-stokes equations for an incompressible New-

tonian fluid and Hagen-Poiseuille flow model. Microbubble positions and trajectories were then randomly generated with probabilities proportional to the vascular flow. This trajectory simulation process has notably been used to generate the Ultra-SR challenge dataset [73].

Alternatively to the low-entropy prior based on anatomical structures, Shin et al. proposed a multi-frame simulation approach with a high-entropy prior [20]. This approach was based on a uniform distribution of initial positions and speed. Microbubble motion was simulated using stochastic perturbations of their directions added at each time step.

3) Considerations on choosing a prior: Using a low-entropy prior based on realistic anatomical structures can improve the overall performance of the trained model on similar data distributions [79]. However, exact modeling of in vivo microbubble positions and trajectories is challenging due to anatomical differences between animal models, organs, and the scale of observed blood vessels. The exact position and velocity of microbubbles cannot be easily measured in vivo, and potentially biased velocity measurements used in the flow model may propagate to the training set. Underrepresented trajectory patterns, such as spinning microbubbles in glomeruli [21], might be undetected by a model trained with existing flow models. Similar to domain randomization in reinforcement learning, which has been proven efficient for transferring from simulation to real-world applications [94, 95], using high-entropy priors could improve the generalization capability of deep learning methods in ULM. Regardless of the choice of prior used for the training set, evaluation on unseen tests data is crucial to distinguish between models underfitting, overfitting, and out-of-distribution generalization. Evaluation datasets that allow computation of widely adopted imaging metrics are also critical for comparison with conventional ULM and other modalities.

C. Conditional probability $\mathrm{p(x|y)}$: ultrasound simulation

In this section, we focus on the simulation of ultrasound signals based on microbubble positions. This simulation process can involve deterministic steps, such as acoustic wave propagation computation, and stochastic steps, such as speckle noise addition. The entire simulation process can be formulated as sampling from the conditional distribution $p(x|y)$. We review various simulation methods employed to model microbubble echoes and techniques for noise addition, with the aim of generating realistic ultrasound samples.

Several studies worked under the assumption of translational invariance and linearity of the imaging system, allowing for simulations based on the convolution of the scatterer distribution with an estimated PSF of the system [64, 82, 84, 87, 90]. To enrich datasets, the PSF parameters were randomized to account for variations in size, intensity, and shape. The parameter variation ranges were estimated based on *in vivo* data. Alternatively, Shin et al. employed generative modeling to sample a wide range of different PSFs based on *in vivo* acquisitions [20].

Several ultrasound simulators have been proposed in the literature to model acoustic wave propagation. Some are mesh-based, like k-wave [96], which allows for simulation of non-linear acoustics with multiple scattering and heterogeneous media. To allow for fine positioning of scatterers and short computation time, most of the surveyed studies have used particle-based simulators such as Field II [80] and SIMUS [85], which are based on stronger prior assumptions of linearity, weak scattering and the homogeneity of the medium.

Adding non-linear effect to capture the full response of microbubbles [97] would allow the application of deep learning methods to process ULM data relying on the non-linear response of microbubbles [98] and increase the realism of simulation of transducers with lower central frequencies. Transducers with central frequencies of 10 MHz and above often have a bandwidth that captures only the fundamental response of microbubbles, which can be modeled with linear simulators. In contrast, lower frequency transducers, such as those used in clinical settings, often capture the non-linear response of microbubbles, allowing for sub-harmonic or harmonic imaging applications but requiring more sophisticated models [97]. Inspired by the combination of k-wave and the Marmottant model of microbubble response by Brown et al. [99], Lerendegui et al. [93] proposed integrating the non-linear microbubble response into a linear simulator, Field II, using a two-step process. First, BuFF [93] uses Field II to estimate the pressure at the microbubble position. The microbubble response is then derived using the Rayleigh–Plesset equation from the Marmottant model [97]. Finally, this response is used to compute the signal received by the transducer using Field II. This approach allows for fast computation using a widely available simulator while modeling the full response of the microbubble. In addition to modelling the full response of microbubbles [97], Blanken et al. [78] proposed to also account for the non-linear propagation of ultrasound [100]. Blanken et al. also showed that localization performance degraded when using a polydisperse distribution of microbubbles, suggesting a more challenging learning problem. Additionally, the non-linear response of microbubbles varies with their parameters (size, coating, manufacturers) [101, 102], making it difficult to accurately model the large diversity of microbubble responses. Accurately modeling $p(x|y)$ for non-linear imaging remains a significant challenge for the application of deep learning approaches in non-linear ULM.

Adding to the diversity of simulators, different noise distributions have been used. For example, some approaches have employed white noise on B-mode or radiofrequency (RF) data with varying SNR [64, 79, 84, 90], or Rice distribution on B-mode, which is more specific to ultrasound data [20]. Aiming to produce realistic noise, Xing et al. [71] used the SIMUS simulator to generate speckle noise from a dense distribution of scatterers. This approach is computationally intensive but provides convincing results in vivo (see fig. 3).

Despite the availability of many simulators, determining which one is best suited for specific needs remains unclear. Empirical comparison between simulators and noise distributions is a challenging task. Their effects are intertwined and entangled with other factors such as model performance and the choice of prior, and require comprehensive evaluation of their impact on the *in vivo* performance of the models.

This is essential before creating large-scale datasets that meet most of the needs of existing studies, thereby enabling better reusability and a wider scope of applications.

D. Learning using in vivo data

In this section, we review how existing approaches have been able to learn directly from *in vivo* data, even in the absence of ground truth for microbubble positions [20, 68, 87]. This ensures that the training and test sets are i.i.d., leading to improved *in vivo* performance [20].

Blind deconvolution is an image processing technique, which, when applied to conventional ULM, jointly estimates the PSF and the position of isolated scatterers [103]. The blind deconvolution algorithm estimates the scatterer distribution that produces the original signal when convolved with the estimated PSF. This distribution is constrained with a sparse prior and is estimated alternately with the PSF. In practice, blind deconvolution can distinguish scatterers only if they are separated by more than one FWHM of the local PSF [103]. Improving on this method, Zhang et al. proposed training two networks concurrently to estimate the PSF and the scatterer distribution [68]. When including regularization and constraints, this approach can be trained directly on in vivo data and account for PSF variability through the trained model. Li et al. [104] proposed a similar approach, leveraging self-supervised learning on *ex vivo* CAM dataset. Shin et al. [20] proposed LOcalization with Context Awareness ULM (LOCA-ULM) that uses Generative Adversarial Networks (GANs) to learn the distribution of *in vivo* PSFs. PSFs can be extracted from *in vivo* datasets using conventional ULM, and a generative model is trained to mimic these extracted PSFs while another model discriminates the generated PSFs from the real ones, providing a loss to train the generative model. By modeling $p(x|y)$ in the neighborhood of microbubbles from *in vivo* data, LOCA-ULM reduces the domain shift between the training distribution and the target distribution. A potential caveat identified by the authors is that, when extracting the PSF from *in vivo* data with conventional ULM, the localization errors on microbubble positions may propagate into the training dataset, which might inherently limit the localization accuracy of LOCA-ULM [20]. Lok et al. [89] have also used labels from conventional ULM to generate realistic datasets, which were acquired *in vitro* in water tank or *in vivo* in CAM datasets or from patient liver.

Leveraging both the precision of physics-based simulators (i.e., exact match between the microbubble position and the simulated echo) and the availability of *in vivo* data has been explored by Yu et al. [87]. They proposed a method aiming to accelerate an existing block matching algorithm for denoising of ULM data before localization [105]. With a few *in vivo* samples labelled with the block matching algorithm and many labeled *in silico* samples, Yu et al. proposed using Domain Specific Projection (DSP) to enable supervised learning while accounting for domain variation, and self-supervised learning to leverage the numerous unlabeled *in vivo* samples. The requirement for labeled *in vivo* data limits the direct transfer of this method from denoising to localization, but it paves the way for other domain adaptation or domain generalization approaches in ULM.

III. Deep learning in ULM processing stages

Deep learning has been utilized at various stages of the processing pipeline that forms ULM images from ultrasound signals. This section is motivated by the diverse applications of these approaches, and illustrates the range of opportunities and formulations that can be employed. We examine each approach in relation to the corresponding steps in the ULM pipeline presented in Section I, identifying which conventional limitations are addressed. This provides a framework for evaluating proposed deep learning methods independently of their application stage. We classify existing applications into one or several processing steps of the ULM pipeline and summarized this view in Fig 2. Additionally, we discuss steps such as aberration correction and denoising, which may enhance image quality when incorporated into the ULM pipeline.

A. Aberration correction

Ultrasound applications in brain imaging are hindered by the presence of the skull, which creates aberrations of the ultrasound wavefront. ULM is similarly affected by aberrations that may, e.g., impede microbubble detection, degrade the PSF, or even cause vessel duplications. Aberration correction has been extensively studied in the ultrasound literature [108]. Existing approaches often utilize speckle brightness [109], measurements cross-correlation from neighboring transducer elements [110], or iterative time reversal [111, 112] to estimate phase differences. Specific methods for plane wave imaging have also been proposed and leveraging the possibility of correcting transmission aberration in postprocessing by, e.g., coherently compounding a large number of angled plane waves [113, 114, 115]. Deep learning-based approaches have also been developed [116, 117, 118, 119], though they are not exclusively limited to ULM.

For pre-clinical studies using ULM, the skull is often thinned or removed to create an imaging window for ultrasound imaging [3, 120]. In small animals, such as young rodents, direct imaging through the skull is possible because its impact on imaging quality is minimal [9, 23, 121]. To simplify the experimental set-ups of pre-clinical studies and allow for clinical applications, recent works have focused on correcting aberrations specifically for ULM [122, 123, 124]. By leveraging the theoretical RF echoes of isolated microbubbles, these methods estimate a phase aberration function using either iterative estimation and virtual focusing [122] or by solving the inverse problem of the imaging process [123].

Similarly, leveraging individual microbubble echoes, the use of complex-valued neural networks (CVNN) has been proposed to estimate the phase aberration function based on microbubble IQ signals [71]. After detecting microbubbles using a standard ULM pipeline, the IQ signal near the microbubbles is isolated and realigned to serve as input to a CVNN, which predicts the aberration function for this region. As shown in Fig. 3, this approach demonstrates convincing *in vivo* results in older mice (6 months), outperforming coherence-based

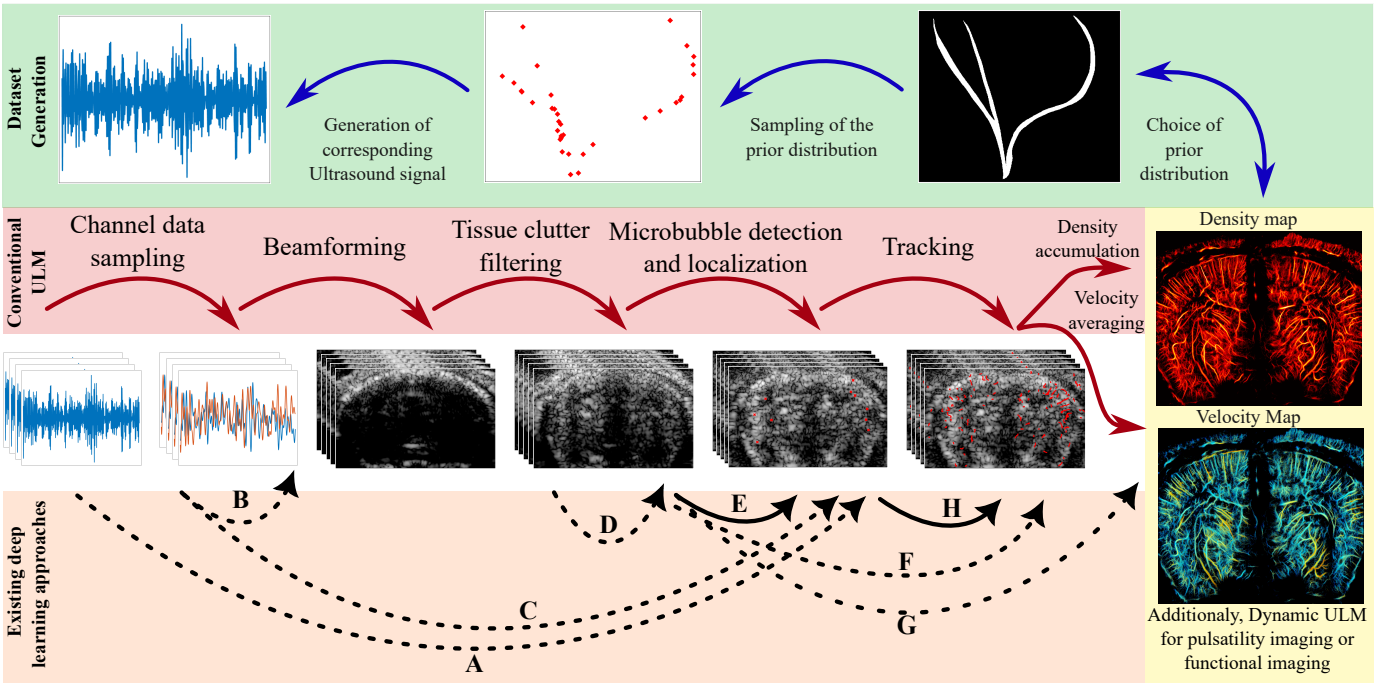


Fig. 2. Overview of ULM processing and simulation pipeline, and the existing deep learning approaches. Dashed lines represent imperfect correspondence with pipeline steps, full lines represent perfect replacement of one or more steps. We note that some pipeline stages can be performed in different orders: such as the beamforming and the clutter filtering [66] or the tracking and localization [106]. A: Blanken et al. localize directly the microbubble in on RF data in channel/fast-time space (i.e., the space where the data lies before beamforming with dimension for the elements of the transducer, the transmits, and the fast-time), beamforming is needed after localization [78]. B: Xing et al. correct for aberration based on IQ data [71]. C: Youn et al., and Hahne et al. use Singular Value Decomposition (SVD) clutter filter and localize the microbubble on IQ in channel/fast-time space [66, 69]. D: Yu et al. enhance the Signal-to-Noise Ratio (SNR) denoising post SVD filtering[87]. E: The localization step has been investigated by several studies [20, 64, 68, 70, 79, 83, 84, 86, 88]. F: Milecki et al. [67] and Lok et al. [89] proposed approaches that temporally project the localization and detect trajectory in spatio-temporal domain, which merges localization and tracking step. G: Chen et al. proposed Deep-SMV that directly estimate the velocity map skipping localization, tracking, and accumulation steps[81]. H: Zhang et al. [107] introduced GRU-MT to solve the assignment problem to form microbubbles trajectories from detected positions.

correction approaches, especially in the presence of a larger number of microbubbles. These results suggest that utilizing CVNN alongside complex-valued IQ data, are relevant for accurately modeling phase relations, and also contributes to the increased robustness to high microbubbles concentrations, exhibited by deep learning localization approaches [67, 79].

B. Beamforming

While initial ULM proof-of-concepts detected microbubble echoes in channel/fast-time space prior to beamforming [128, 129, 130], more recent ULM studies performed beamforming, typically using the delay-and-sum algorithm [131], before microbubble localization [74]. Recent efforts in developing deep learning approaches for microbubble localization have focused on utilizing the full information from either uncompressed RF data or complex-valued IQ data [66, 67, 69, 78, 79]. To do so, some studies used the signal in channel/fast-time space and performed localization either in this space [78] or directly in the image space with specific projection layers [66, 69]. Since only the signal corresponding to microbubble positions needs to be projected into spatial coordinates, the beamforming operation can be simplified [66], which can save processing time and limit beamforming issues like grid artifacts.

Youn et al. [69] argued that overlapping PSFs caused by high scatterer density induce a loss of information. To alleviate this issue, they used a CNN that directly processes RF data from every channel and every transmit to predict scatterer positions in the beamformed space. To stabilize training by reducing output map sparsity, they introduced a confidence map prediction, from which the exact positions of scatterers can be extracted in a second step. This CNN integrates both localization and beamforming operations and includes absolute positional embedding layers, specifically CoordConv [132], which add channels encoding the pixel absolute positions in the image. This model approach was validated *in silico* and *in vitro*, and could resolve overlapping PSFs with better performance than conventional ULM.

Driven by a similar motivation to utilize the entire RF information, Blanken et al. proposed using a 1D CNN to recover the time of arrival of microbubble echoes in RF data from a single channel [78]. Microbubble responses and acoustic wave propagation are not assumed to be linear. The network was trained on the full response of monodisperse microbubbles, making it suitable for lower frequency setups. This approach performs the deconvolution of the RF signal, which yields a super-resolved image after DAS beamforming. Evaluating the importance of modeling the nonlinear response of microbubbles is crucial for clinical applications, which typically use lower frequencies than small animal studies. As highlighted by the authors [78], applying a 2D end-to-end approach (i.e., using all RF channels and producing an

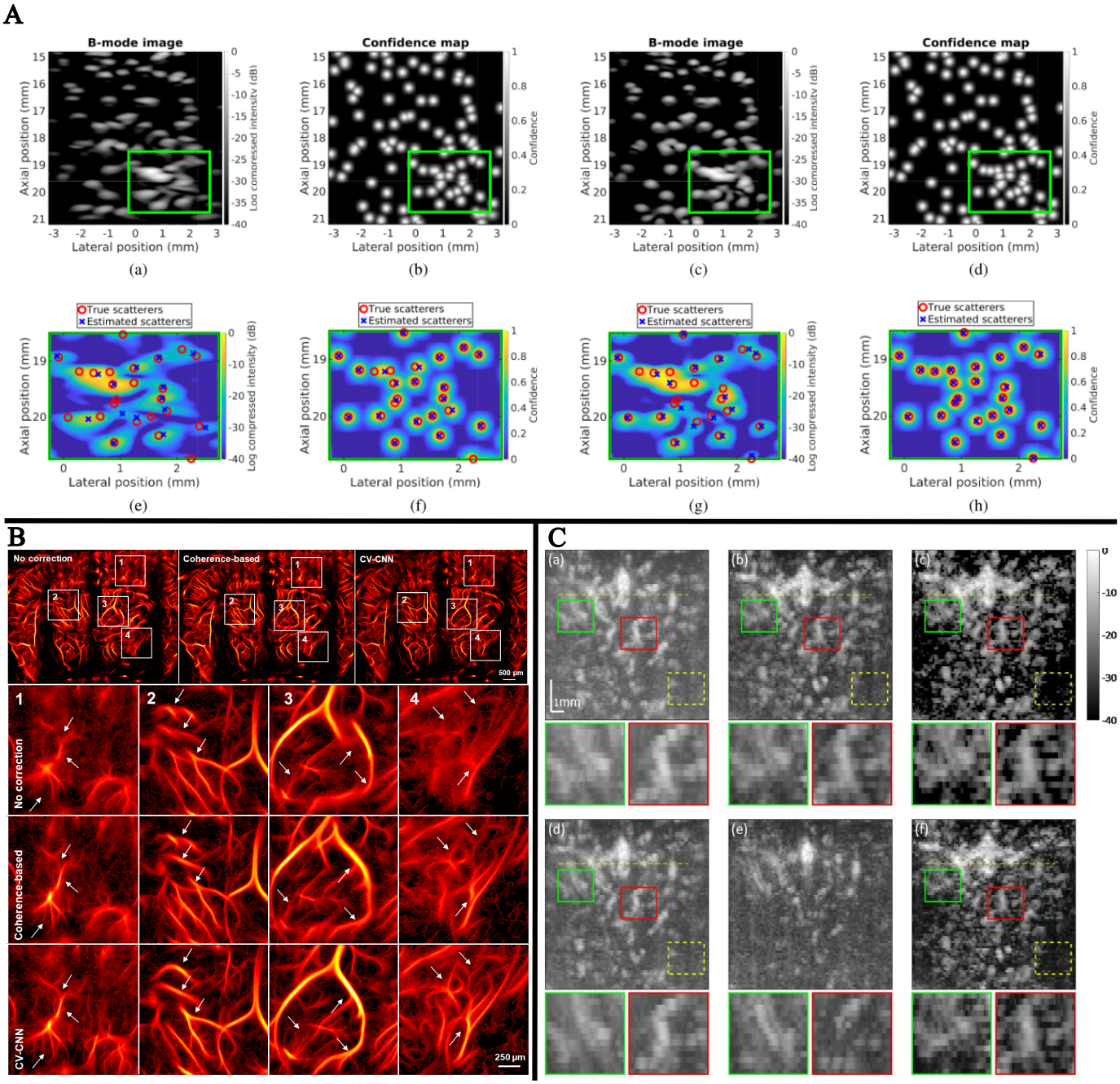


Fig. 3. A: In silico results from [69] ©2020, IEEE. B-mode image and local maxima detection compared with the confidence map from the deep learning model [69]. B : Illustration of results obtained with deep learning based aberration correction from [71]. *In vivo* imaging on 6 months old mouse through intact skull and skin showing improved vascular reconstruction with the deep learning based approach using a CVNN in comparison without aberration correction or with coherence based correction adapted from [110]. White arrows are pointing to disconnected or duplicated vessels that are corrected using the CVNN correction. ©2024, IEEE C: Example of clutter filtering using deep learning of a CEUS acquisition of a rat brain vasculature from [61]. The Deep unfolded model-based method, CORONA (c) from [61] is compared with SVD clutter filter [125] (a), the model based approach, FISTA [126], (b),6th order Butterworth filter (d and e), and a ResNet [127] trained to perform clutter removal. Color bar is in dB. Reused with the authorization of the authors ©2020, IEEE

output in beamformed space) is both promising and essential for further applications but poses complexity issues during training.

Building on these approaches, Hahne et al. [66] recently proposed a method performed in channel/fast-time space and successfully applied it *in vivo*, surpassing conventional ULM and B-mode-based deep learning ULM. Their model utilized complex-valued IQ data, with the real and imaginary parts represented in the channel dimension, whereas previous approaches used uncompressed real-valued RF data. To perform localization in channel/fast-time space on *in vivo* data, Hahne et al. applied clutter filtering on RF data and used an affine projection model to map ground truth from image space to channel/fast-time space for network training and invert this projection to recover predictions in image space. This affine transformation replaces the beamforming operation while fully

leveraging the sparsity of the localization prediction. Finally, this approach demonstrates increased robustness to the domain shift between training simulations and *in vivo* predictions [66].

C. Clutter filtering and denoising

To detect microbubbles, various strategies have been explored to remove tissue clutter, including non-linear imaging [18, 98, 133, 134] and post-processing filtering [125]. In ULM, SVD clutter filtering [125] is widely used to remove tissue clutter due to its simplicity and efficiency. Alternative filtering methods such as high-pass filtering [21, 91] and mean removal [135] have also been applied, either complementing or replacing the SVD filter. Several approaches have been proposed to enhance the SVD clutter filtering step using deep learning, primarily to address computation time issues [61, 136, 137]. Among these, Solomon et al. [61] proposed using Robust Principal Component Analysis (RPCA) instead of SVD to leverage the spatial sparsity of microbubbles alongside spatio-temporal information. To mitigate computation time issues, they employed deep learning to enhance the convergence rate of the iterative algorithm used for RPCA decomposition.

In practice, clutter filtering algorithms can be sufficient to perform ULM in murine brain imaging. However, in less ideal conditions (e.g., deeper field of view, thicker skull), the lower SNR of microbubble echoes hinders both the localization and tracking processes. To facilitate application on larger animal models or clinical translation, studies have investigated the impact of adding a denoising step after clutter removal. Model-based approaches have used non-local means filtering [138] or block matching [105] to improve SNR after clutter removal. Being based on the assumptions of microbubble sparsity, these approaches are specific to CEUS and ULM. Yu et al. [87] proposed a data-driven approach to learn the denoising step performed by block matching [105] from *in silico* and *in vivo* data. This approach uses a domain adaptation method, namely Domain Specific Projection from [139], to adapt the representation learned by the network to the training domain *in silico* and the testing domain *in vivo*. The training set comprises *in silico* labeled data, and a large amount of *in vivo* data, with only a limited number labeled with only a limited number labeled with the predictions of the block matching denoising algorithm. Semi-supervised learning leverages the unlabeled *in vivo* data to enhance the model's performance. This approach improves the processing time for the denoising step, making it more usable and enhancing the downstream ULM image quality.

D. Localization

After beamforming and filtering, conventional ULM often performs a simple detection step based on local maxima of intensity [6], SNR [9], or local correlation with the PSF [15, 67, 138]. After detecting local maxima, sub-resolution localization aims to determine the precise position of microbubbles within small regions of interest centered on the local maxima. This process relies on the assumption of having a single, isolated scatterer in the region of interest and can be performed with radial symmetry or Gaussian fitting, among other methods [74]. When microbubble trajectories get closer or cross, causing their PSFs to overlap, this assumption no longer holds, leading to missed detections or increased localization errors. This effect can be mitigated by injecting a lower concentration of microbubbles, but this increases the acquisition time, limiting the application and translation of ULM to clinical settings [33]. Deep learning approaches have been proposed to address the issue of overlapping PSFs and to allow for higher microbubble concentrations and reduced acquisition times. By learning more complex patterns in simulations or by adding temporal context to localization, deep learning methods have enabled increased microbubble concentrations both *in silico* and *in vivo* [20, 67]. Numerous approaches have been proposed, and a more focused and exhaustive review is provided in section IV.

E. Tracking

Tracking algorithms are commonly used to remove microbubble detections that cannot be tracked across several frames. Additionally, they can identify microbubbles over multiple frames and allow for track interpolation to compensate for high velocity or missed detections. The tracking step is often performed using the Hungarian or nearest neighbor algorithms and may incorporate Kalman filtering to refine the assignment solution [16, 26, 140, 141]. Deep learning improved Kalman filter [142] has been applied jointly with the Hungarian algorithm [73, 143], but designing deep learning alternatives to conventional assignment algorithm remains challenging. Tracking can be computationally intensive, and commonly used algorithms struggle with high concentration areas and microbubbles of variable velocities. The combined use of temporal context, uncertainty in localization, and trajectory dynamics is difficult to model, making it an attractive application case for data-driven approaches. Recently, Zhang et al. [107] have proposed an approach that solves the assignment problem and improves the position estimation based on a Gated Recurrent Unit [144] and using the positions predicted in the 4 preceding frames. The position estimations and predicted trajectories are merged in post-processing, allowing to recover full length trajectories. Sui et al. [145] proposed to use GANs to perform the tracking step from localization maps. Some approaches have merged localization and tracking to directly predict more downstream results. Milecki et al. [67] proposed Deep-stULM, an architecture incorporating temporal context through 3D convolution. This method outputs the projection of all microbubble positions for a short period, allowing the CNN to learn temporal information and distinguish neighboring microbubbles based on their trajectories. Similar output formulation has been used while encoding temporal information within channel dimensions [89]. These approaches were able to reconstruct high quality density maps *in vivo* using high microbubble concentration [67, 89]. To extend these applications to velocity estimation, Chen et al. [81] proposed Deep-SMV, a Long Short-Term Memory (LSTM) based approach that directly outputs velocity maps. Incorporating the tracking step in Deep-SMV is particularly relevant when estimating

velocity maps, as errors in tracking can significantly impact velocity values.

IV. A FOCUS ON MICROBUBBLE LOCALIZATION

Given the large diversity of methods tackling the localization stage in the ULM pipeline, this section provides a more thorough focus on deep learning methods for localization. Some examples of *in vivo* application on various animal models and organs are displayed in Fig. 4. With an increasing number of studies targeting performance improvements rather than new applications, evaluating and comparing these approaches is becoming more critical. We aim to provide insights on this topic and review how current approaches differ in terms of evaluation, formalism, and architectures.

A. Evaluation

Performance evaluation of localization algorithms, whether deep learning-based or not, is crucial for comparison and further improvements in ULM. The ULTRA-SR challenge [73] received 38 submissions and enabled fair comparisons of several state-of-the-art approaches. Submitted approaches were evaluated on both *in silico* test sets and *in vivo* data, with the latter assessed by an expert panel. The datasets used for evaluation were simulated using BUFF [93]. Heiles et al. also proposed a benchmark to evaluate localization methods [74] both *in vivo* and *in silico*. *In vivo* evaluation is performed with objective metrics such as gridding or saturation, which can be measured when evaluating new methods in future studies. RF data were also made available [150], allowing the comparisons of methods that utilize the full RF information in channel/fast-time space [66, 69, 78] or in image space [67, 79]. This benchmark has been used in recent studies [66, 106, 151].

When evaluating deep learning methods, the choice of training simulation parameters, such as dataset size, prior distribution, and simulation model, can greatly influence a model's performance. Thus, proposing fair evaluations for deep learning methods also requires a companion training set. Using a test set sampled from the same distribution as the training set allows for i.i.d. evaluation, providing a fair quantification of the model's expressive power. The generalization ability of deep learning methods also needs to be evaluated on out-of-distribution examples, such as *in vivo* datasets and *in silico* data sampled from a different prior distribution or using a different simulation model. Since replicating deep learning baselines requires the original training dataset, method implementation, and careful hyperparameter tuning, benchmarks are important to avoid costly baseline comparisons. In this section, we review how existing deep learning localization methods have been evaluated in the absence of such deep learning-specific benchmarks.

1) In silico evaluation: Many studies have performed evaluations on *in silico* datasets, as such test datasets are easily obtained when the simulation process is already implemented for the training set generation. Since the simulation process is based on known microbubble positions, it allows for an evaluation that penalizes false positives and precisely measures localization error. In practice, *in silico* evaluation can focus on two aspects: i.i.d. evaluation and out-of-distribution evaluation.

Model evaluation on i.i.d. datasets is particularly relevant for studies proposing new architectures, as it allows for comparing the representational power against pre-existing deep learning methods. For example, this approach has been used to show that transformer blocks could improve the model expressive power [83] or to quantify the impact of using Sparse Tensor Neural Networks in ULM [146].

By changing the simulation parameters, several studies have explored the robustness of their proposed models on specific out-of-domain generalizations, such as increased noise [68, 78], aberrations [20], and high microbubble concentration [67, 68, 152]. A key benefit of using deep learning models for localization is their robustness in high-concentration scenarios. Consequently, several studies have evaluated the robustness of their methods under increasing concentrations and compared them to conventional localization techniques [67, 68, 152].

In silico evaluations often employ pre-existing metrics to measure the precision and recall of the models [64]. To measure localization precision, distance-based metrics such as RMSE can be used for detected localizations [64]. These are standard evaluation metrics also found in PALA and ULTRA-SR benchmarks [73, 74]. More aggregated metrics, such as the Jaccard index or the Dice coefficient, have also been proposed to evaluate the overlap between a ground truth angiogram and its estimation from the model [67].

2) In vivo evaluation: Even though deep learning approaches for ULM can be trained and evaluated on *in silico* datasets, their targeted application is to perform well on *in vivo* datasets, which, in that case, corresponds to out-of-domain generalization. Microbubble ground truth positions are not available for *in vivo* datasets, making the evaluation of *in vivo* performance challenging. This inherent limitation in ULM evaluation can be mitigated by using anatomic validation with other modalities [20, 79, 81], qualitative image assessments[67], or evaluation metrics that do not require ground truth, such as Fourier Ring Correlation (FRC) [91] or full-width half maximum on arbitrarily selected regions.

The first deep learning approaches for microbubble localization were evaluated *in vivo* and compared against conventional ULM using full-width-half maxima [64, 67, 81, 84, 88]. These comparisons provide convincing proof of feasibility and out-of-domain generalization of the proposed methods.

To improve robustness and replicability, more recent approaches have used FRC [91] to evaluate the resolution of deep learning methods *in vivo* [20]. FRC measures spatial resolution without depending on selecting specific blood vessels, which improves reproducibility and robustness. Additionally, Shin et al. correlated the number of detections with Power Doppler intensity to evaluate microbubble detection power [20].

To provide evaluations against anatomical ground truth, Song and collaborators have used ex-ovo Chicken Embryo CAM and optical imaging to obtain a ground truth of the vasculature [20, 79, 81]. Measuring the overlap between the vascular network estimated by the model and that imaged by optical microscopy provides an evaluation method that is robust to false detections or vessel duplicates due to aber-

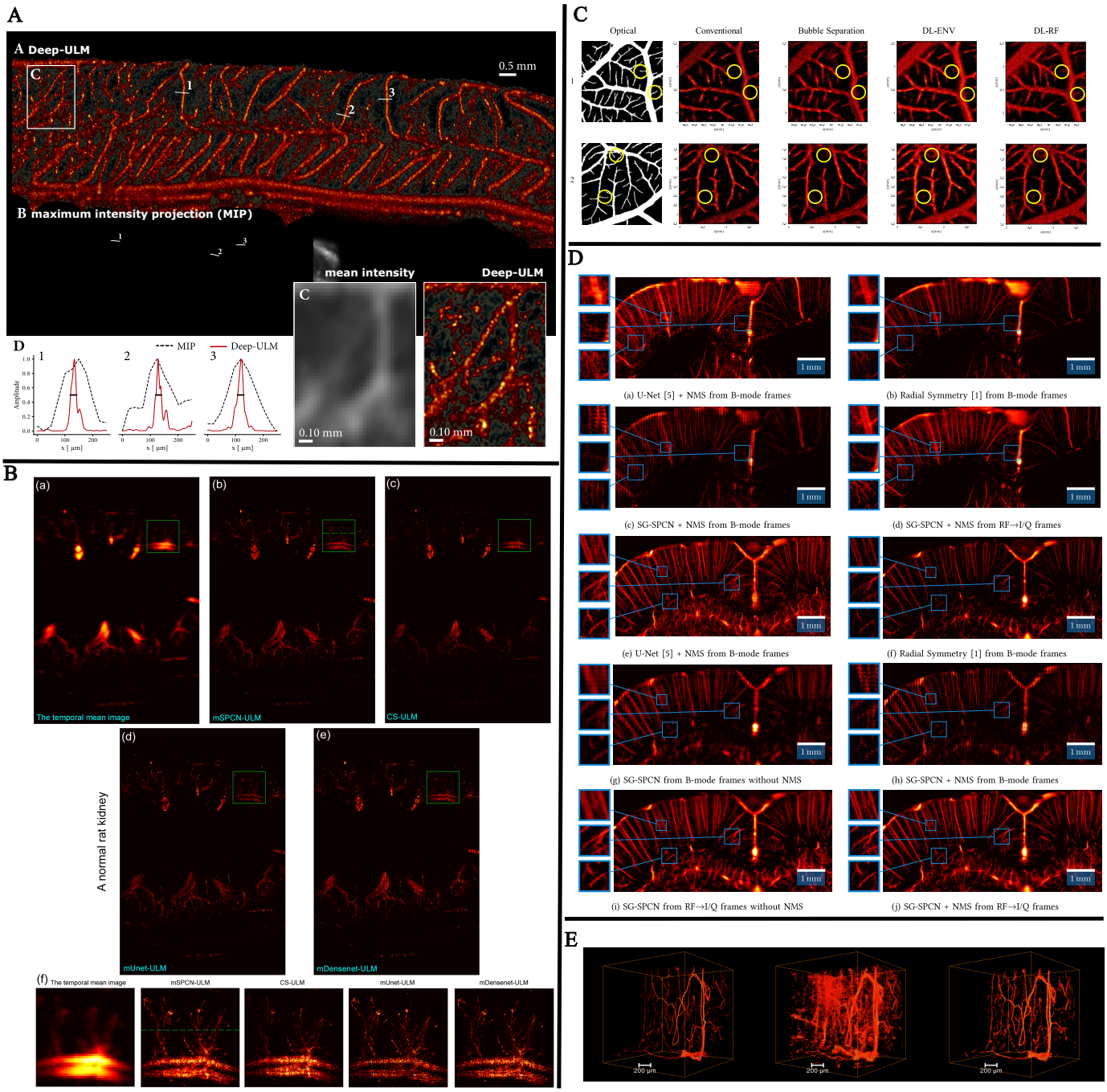


Fig. 4. Illustration of *in vivo* results in 2D from different deep learning localization approach in various organ and animal models (A, B, C, and D) and *in silico* results in 3D from [146] (E). A: Rat spinal cord processed with Deep ULM [88] compared to Maximum Intensity Persitence (MIP) image from [88] ©2020, IEEE. Acquisition performed with a frame rate of 400Hz and a transmit frequency of 15MHz. The black horizontal lines on the intensity profile indicate full width half maximum and measure respectively $\mathbf{21}\boldsymbol{\mu}$ m, $\mathbf{19}\boldsymbol{\mu}$ m, and $\mathbf{20}\boldsymbol{\mu}$ m for profiles 1, 2, and 3. B: Rat Kidney from [64] ©2020, IEEE. Acquisition with a frame rate of 400Hz and a transmit frequency of 15.625MHz processed mSPCN-ULM [64], mUnet-ULM adapted from [62], mDensenet- ULM adapted from [147], and CS-ULM [148]. C: Chicken embryo CAM image from [79]. Acquisition made with a frame rate of 1000Hz and a transmit frequency of 20MHz and processed with DL-ENV, DL-Rf and optical reference from [79] compared with Fourier based microbubble separation from [149] and conventional ULM. D: Rat brain image from [66] ©2024, IEEE. Data from PALA [74] processed with SG-SPCN with and without Non-Maximum Suppression (NMS) on B-mode or IQ in channel/fast-time space [66] and compared with U-Net from [88] and Radial Symmetry [74] on B-mode. E: In silico comparison in 3D for conventional ULM (center) and sparse tensor neural network (right) with ground truth (left) from [146].

rations. Validation using MRI or CT in sheep brain, human brain, and heart has been performed by conventional ULM studies [18, 23, 135]. Such validation could be of interest for deep learning as a more challenging evaluation in presence of aberration, reverberation and strong attenuation. It has been reported that registration can be challenging for such validation [22].

3) Future directions: Due to their relatively recent introduction, existing deep learning-based approaches have focused on identifying the potential benefits of using deep learning for localization or on demonstrating feasibility *in vivo*. These kinds of contribution are extremely valuable as ULM is gradually applied to more complex imaging challenges, and the ability of deep learning approaches to enhance ULM localization in such setups remains an open question. Improving existing ULM applications is equally important and requires fair and repeatable evaluation procedures. Evaluation on publicly available datasets [74], with reproducible metrics [91], facilitates the comparison with conventional ULM, existing deep learning methods, and future approaches. Since such comparisons also depend on the training dataset used, it is important to distinguish the impact of a new simulation framework from using a new model architecture. Models can be evaluated with a given training set and evaluation set, as is done in other fields, but this requires publicly accessible benchmarks with a companion training set. Such benchmarks should compare models in the case of i.i.d distribution as well as *in silico* and *in vivo* out-of-domain generalization to independently evaluate expressive power and generalization ability. *In silico* comparison with conventional techniques, which do not benefit from small distances between the training and test distributions, requires realistic noise levels and precautions to avoid test set contamination, ensuring unbiased evaluation in favor of data-driven approaches. The evaluation of the intrinsic quality of training datasets in ULM remains an open question and could be an impactful research direction.

B. Training formulation

In the literature, the localization task in ULM has been formulated as an optimization problem in various forms to train models. The choice of input and output representations directly influences the design of loss functions, which are crucial for model performance and robustness. This section details existing formulations, focusing on how they differ in terms of ultrasound representation, temporal context size, and output formats, along with the implications for the loss function.

1) Ultrasound representation: Many approaches directly use B-mode images [62, 64, 79, 81, 83, 153]. RF data can also be used either before [78] or after beamforming [79] as they allow leveraging the full RF information, which contains both the B-mode and the phase information. Chen et al. demonstrated that using the full RF data rather than B-mode images, with adequate spatial sampling, is beneficial [79].

To benefit from lossless IQ compression and reduce computation cost, some approaches [66, 67] use the complex valued IQ data, with the real and imaginary parts integrated as channel data within real valued neural networks both prior to and post beamforming. Complex Valued Neural Networks (CVNN) [154], leveraging complex arithmetic and complex model parameters, can also be used to process the complex-valued IQ data in ultrasound image reconstruction [155] or for aberration correction in ULM data [71].

2) Temporal context: In addition to ultrasound data representation format, localization deep learning methods also differ in the input temporal context size. Most localization deep learning approaches in ULM process frames independently and output a map of pixels containing microbubbles [64, 88]. Relying on temporal context to enhance detection, some methods consider multiple frames simultaneously, raising challenges in input handling and output formulation. DeepstULM [67], for instance, used a 3D spatio-temporal CNN to process 512 frames, outputting a temporal projection of microbubble positions. Lok et al. [89] encoded temporal information in the channel dimension with a smaller context size, reducing memory constraints. Gu et al. proposed predicting accumulated ULM images directly from averages of 20 B-mode images, bypassing localization maps [82]. To provide further information, Deep-SMV [81] used 16-frame temporal context and sequence modelling, specifically Long Short Term Memory (LSTM) [156] to predict velocity maps. Lee et al. proposed to use optical flow estimation to represent temporal context [157]. Shin et al. proposed LOCA-ULM inspired by Single Molecule Localization Microscopy (SMLM) architecture [158] to incorporate temporal context from adjacent frames for improved localization in single frame [20]. Using a U-net to process three frames independently, and then aggregating features in another U-net, their approach predicts microbubble positions in the center frame. Obtaining microbubble positions for each frame can facilitate integration into ULM pipeline and future applications. LOCA-ULM effectively incorporates adjacent frame information, replacing the conventional localization step, though it has been applied to only small context sizes (3 frames). Recently, Pustovalov et al. [86] also proposed an architecture inspired by DECODE [158], which uses 3D convolutions to incorporate the temporal context.

3) Loss formulation and output format: Intrinsically linked to the output format, the loss function formulation, or training objective, is key to the performance and robustness of deep learning approaches. Most localization methods encode the microbubble presence probability for each pixel of a grid that typically matches the desired ULM image resolution. Consequently, microbubble positions in these high-resolution localization maps are very sparse, and training with naive loss functions often leads to predicting zeroed outputs [67, 88].

To improve training stability, van Sloun et al. proposed convolving the output maps with a Gaussian kernel to obtain soft labels, while constraining the solution with a L_1 sparse prior [88]. Similarly, Liu et al. applied the convolution kernel to both the network output and the labels [64]. Reducing the Gaussian kernel's standard deviation during training can also promote sparser predictions [66], allowing for the use of simple L_2-based loss functions, such as Mean Squared Error. Incorporating more complex loss functions based on SSIM [79] or focal loss [84] can smooth the loss function and improve training stability. Inspired by medical image segmentation [159], some works have used the dice coefficient either alone [67] or in combination with L_1 applied to soft labels [78]. While the dice coefficient automatically accounts for class imbalance, it is not continuous with respect to spatial

translation, which can reduce training stability. Zhao et al. [160] leveraged Wasserstein GANs [161] to improve their loss formulation and enhance localization performance. Shin et al. [162] recently used a loss function based on the total count of microbubbles and the prediction likelihood under a probability density derived from the labels through a Gaussian Mixture Model. Originally developed for SMLM [158], this loss function is also adapted for sub-pixel localization, reducing the impact of the output grid resolution and making it highly relevant for localization tasks. It requires changing the output format from a probability localization grid to a multichannel output that encodes not only the probability of the presence of a microbubble, but also its relative position with respect to the center of the pixel. Changing the output to a denser format, such as velocity maps [81] or non-overlapping Gaussian confidence maps [69], also reduces training instability and permits the use of simpler loss functions. Alternative approaches using object detection formulations have also been proposed [163, 164].

C. Architectures

Due to the significant diversity in problem formulations for localization tasks, deep learning approaches tackling this challenge often have distinct architectures. Some components or properties are shared among these architectures, and in certain cases, their specificity is limited to a few layers. Many localization methods use CNNs with an encoder-decoder structure [64, 67, 79, 88, 89], but typically differ in their modeling of the upsampling techniques used to achieve super resolution or the building blocks used for feature encoding. In this section, we review the different building blocks employed in these architectures. We also explore the architectural specificities of approaches operating in the channel/fast-time space. Given that most of these architectures are limited to 2D imaging, we review the few studies that focus on scalability to 3D imaging.

1) Upsampling, super-resolution and grid artifact: Since most existing architectures use encoder-decoder architectures, upsampling is required to project low-resolution encoded representation back to the original input resolution. Furthermore, in sub-resolution localization, the output is often projected at a finer resolution than the input, necessitating additional upsampling. Upsampling operations are also common in natural image super-resolution architectures and have been known for leading to checkerboard artifacts [165], resulting in gridding in the ULM image. Upsampling has been done using methods such as nearest neighbor upsampling [67] or transposed convolution [62]. A more efficient approach is to use sub-pixel convolution [166], which has been applied for localization in ULM [64]. Sub-pixel convolution is empirically more robust to checkerboard artifacts and is computationally more efficient.

It is also possible to reduce the number of upsampling layers required with an improved output format, as done in the DECODE architecture [20, 158], which allows for sub-pixel resolution with limited dependence on the grid. However, grid artifacts can still occur in high-concentration areas [158] but can be mitigated by interpolating the model's input [20]. Performing the localization in channel/fast-time space can also be key in eliminating grid artifacts [66].

2) Architecture in channel/fast-time space: Performing localization prior to beamforming presents additional technical challenges in terms of architectures. Unlike post beamforming localization, where the microbubble echo is spatially limited, the echo originating from a single microbubble in channel/fast-time space reaches many or all of the elements of the transducer at different times. Depending on the temporal sampling and the use of RF data rather than IQ, the receptive field required to encode a microbubble echo with a single element can already span up to 125 grid points [78]. When information from several elements is considered, the receptive field needs to be further increased to account for the time differences depending on the field of view.

To increase the receptive field of their networks, Youn et al. used an encoder-decoder structure with additional downsampling blocks [69]. Alternatively, Blanken et al. utilized dilated convolutions [78]. Hahne et al. proposed adding a semi-global bottleneck block to ensure a sufficiently large receptive field [66]. These methods can also handle the beamforming operation, which can be done implicitly through position embedding [69] or with a learned projection of the detection [66].

3) Attention-based architecture: Inspired by recent advances in NLP and vision from attention-based architectures and transformers, transformer-based models have been applied to the ULM localization task [65, 70, 83, 90, 164, 167, 167, 168, 169, 170]. For example, Liu et al. proposed SR-MT (Super-Resolution Modified Transformer) [90], based on the Swin transformer block [171], which employs shifted windows to improve the efficiency of attention computation and allows for modeling at various scales through its hierarchical architecture. Similarly, Luan et al. [84] leveraged attention mechanisms to enhance the representational power of localization models, addressing concerns about the efficiency and scaling complexity of self-attention. They proposed a cascade-axial-attention (CAA) block to capture global context with reduced complexity. Gharamaleki et al. [164] modified the problem formulation to an end-to-end object detection framework and applied DEtection TRansformers (DETR) [172]. To address inherent limitations of DETR in small object detection, Gharamaleki et al. [168] also proposed using Deformable DETR [173] within a similar framework. Transformer-based architectures have also been used to improve ULM reconstruction from averaged B-mode images [65], showing improvements both *in silico* and *in vivo* compared to previously introduced GAN-based methods [82]. Zhang et al. [70] have also incorporated Transformer Self Attention into a modified U-net architecture, called ULM-TransUNet. ULM-TransUNet outperformed conventional ULM and other deep learning ULM approaches based on U-net or transformers in several settings, including at high concentration *in vivo*.

4) Scaling to 3D imaging: ULM and DULM have been extended to 3D imaging [6, 9, 16, 174] to offer a better understanding of vascular anatomy and function, as well as to reduce user-dependence on measurements due to the choice of the imaging plane. 3D imaging is particularly challenging due to the associated computational cost of an additional

dimension, which is even more problematic for DULM as the temporal dimension needs to be preserved. This computational cost has significantly hindered the application of deep learning approaches to 3D ULM, as both training and inference can become problematic. Some solutions have been proposed [146, 175]. Piepenbrock et al. proposed using a 3D CNN to perform localization and process each volume independently [175]. Rauby et al. [146] used Sparse Tensor Neural Networks [176] and tensor pruning based on intermediate predictions to improve the scalability of CNN-based architectures for ULM by leveraging temporal context and high-resolution projection grids. *In silico* results are displayed in Figure 4.

V. Perspectives

In this section, we detail current limitations either partly or not addressed by existing approaches and that we find to be critical for future improvements of deep learning in ULM. Improvements in deep learning in ULM also requires better evaluation and comparison to established baseline on widely accepted dataset. We explore the prerequisite for such evaluations.

A. Limitations and future challenges

1) Improving deep learning approaches: After forming tracks by pairing localizations across frames, conventional ULM approaches further leverage microbubble trajectory information. These trajectories can be used to compute clinically relevant statistics at trajectory level. For example, the Sum Of Angles Metric (SOAM) is computed directly on each trajectory and has been used to estimate vascular tortuosity in the mouse brain, showing significant differences between young and aged mice [177]. Sensing ULM (sULM) utilized trajectories metrics such as normalized distance, remanence and dispersity [21, 22] to identify microbubbles travelling in glomeruli. Trajectories can also be used to estimate blood flow pulsatility [178] which has been link to AD and cognitive decline [179, 180, 181, 182]. In addition to its clinical relevance, using prior assumptions on microbubble trajectories can improve position estimates and filter out detections based on track length or global displacement. Inverting the tracking and localization steps, to better leverage the temporal context and continuity prior of trajectory, has also been shown to improve localization performance [106]. Despite the importance of microbubble trajectories in conventional ULM, current deep learning approaches do not fully utilize such prior information. They either perform single-frame localization with tracking as a post-processing step or estimate temporal projections of tracks [67, 89], which prevents downstream utilization of the tracks. Formulating tracks as deep learning model outputs and developing a differentiable loss to enable end-to-end learning of trajectory-level predictions is currently an open problem in ULM. Such formulations would allow for the joint estimation of sub-pixel localizations and tracking across several frames, using temporal context to disambiguate spatial measurement uncertainty and vice versa.

Current deep learning approaches have mostly focused on 2D imaging, which heavily depends on the selection of the imaging plane for reproducibility. 2D imaging is inherently limited for velocity measurements due to planar projection of trajectories and is sensitive to tissue motion caused by out-of-plane displacements. Extending existing architectures to 3D imaging poses challenges in terms of memory usage and computational cost required for model training and inference. Dataset construction is also more challenging in 3D due to the increased computational and storage demands. Some solutions have been proposed to reduce these computations by leveraging the sparsity of microbubble trajectories [146], but they have not yet been applied *in vivo*.

2) Evaluation and Benchmarking: With the growing number of deep learning methods for the processing of ULM, standard methods to fairly compare deep learning approach become more and more critical for the community. Akin to previous comparison efforts between ULM methods [73, 74], a widely adopted benchmark would benefit the community given that it addresses the limitations of existing works. More explicitly, we believe that such a benchmark have to fulfill several desiderata:

- **Determined training set**, in order to disentangle improvements in model architecture or problem formulation from improvement in simulation procedure or impact of dataset size. This training set can contain equivalent signal in various forms (IQ/RF/B-mode) to allow for flexibility in input representation as well as improvement in several steps of the ULM pipeline. It also needs to be large enough to allow for discrimination of representation power with limited risk overfitting. Finally, the simulation process needs to be realistic enough to limit domain shift when transferring to *in vivo* application.
- **Diverse evaluation datasets** on several *in vivo* applications (organs, experimental set-up, animal model) but also *in silico* to evaluate localization error on several various microbubbles velocities, unseen trajectory pattern. PALA benchmark [74] and ULTRA-SR challenge [73] datasets are relevant and could be included in such a set of evaluations datasets. However, careful considerations have to be taken to propose evaluation datasets with both similar and different simulation parameters to the training set in order to evaluate generalization and robustness of deep learning model.
- **Collection of robust metrics** with known value for conventional approaches as well as existing deep learning approaches. ULM performance metrics such as Jaccard index, lateral and axial error, gridding, FRC, and saturation have already been proposed and used in the literature [74, 91]. Works in biomedical image analysis [183] have provided valuable recommendations on the pitfalls to consider when designing and selecting these metrics.

If the constitution of such benchmark would benefit the community in terms of results reproducibility and comparison between method, quality control and best practice guideline from other fields[183, 184] should be adapted and implemented to ensure that progress enabled could translate to

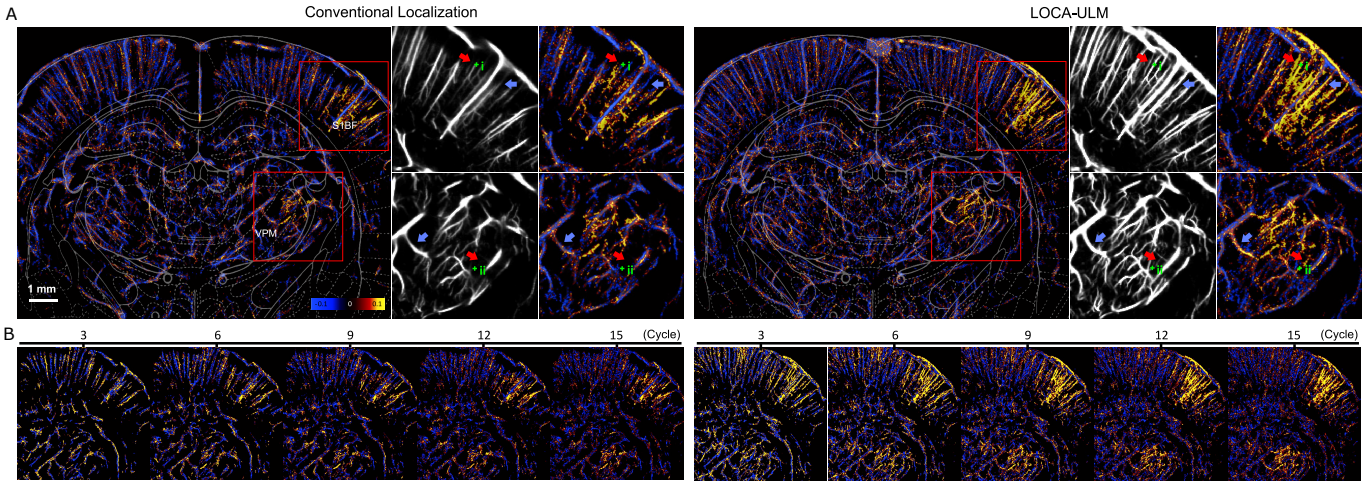


Fig. 5. fULM comparison between conventional localization and LOCA-ULM from [20] showing the improved microbubble detection ability of LOCA-ULM, allowing to reduce the required number of stimulation cycles. Reused with the authorization of the authors

robust improvements. In addition to the datasets and metrics, other factors might prevent comparison between deep learning approaches, such as limited availability of the code and dataset to reproduce results, varying computation times, or outputs format too different from microbubble trajectories. When comparing deep learning models, especially with larger datasets available, model complexity and computational costs can become the limiting factor of performance, which makes their comparison and reporting critical. Several studies have reported computation times and comparison with existing approaches [64] to highlight reduced computational costs. As noted by the authors, these reports provide insights on computational efficiency rather than offering an absolute comparison, which prevents comparison across different studies. Moreover, computational efficiency can vary greatly depending on the hardware used and the level of implementation optimization. For example, deep learning library leverage highly optimized convolution algorithm and parallel computing on GPU, which can bias time-based comparisons of complexity. Analysis of computational complexity, reporting the relation between the output image size and the number of operations required to process it, is crucial for further understanding and improvement of processing time. Conventional ULM currently relies on representing microbubble trajectories as tracks, and in a near future downstream application and biomarkers are likely to rely on such structure. A unified output format is essential for fair comparison between approaches. Thus, when trying to improve on existing applications and changing output format, one needs to be cautious as it might hinder comparisons with existing and future approach or limit the application to future ULM development. However, these considerations should not discourage proof of concept or feasibility study such as velocity prediction [81] or ULM at high concentration [67, 89].

B. Successes and promises

The development of the aforementioned deep learning approaches in ULM in recent years has shown repeated successes, proving their relevance in certain applications. Multiple approaches have demonstrated good performance *in vivo* while being trained *in silico*. This provides reasonable evidence that current acoustic simulators are realistic enough to enable deep learning models to generalize to real *in vivo* data.

Several studies [20, 64, 69, 70, 79, 84, 88, 167] have provided consistent results showing that deep learning approaches are better suited for modeling ULM signals in high concentrations of microbubbles. *In silico* comparisons have been conducted under varying concentrations, showing better localization performance for deep learning-based methods at high concentration in 2D [20, 64, 69, 79, 84, 88, 167] and in 3D [146]. *In vitro* studies have also shown similar results [69]. *In vivo* comparison studies with conventional ULM have also reported improvements when using deep learning methods for ULM at high concentrations [20, 67, 70, 79]. The improvements were measured either in FRC, FWHM, or number of detections/vessel saturations. Such findings pave the way for faster ULM acquisition with higher microbubble concentrations, which has already been applied to facilitate functional ULM (see Fig. 5). Recent developments in model architectures have been demonstrated that performance in high concentration of microbubbles could be further improved [70].

Deep learning has also been very successful in accelerating ULM processing. For certain processing steps, such as clutter filtering or denoising, conventional algorithms with high performance exist but are limited in application due to long computation times [61, 105]. Using these algorithms to generate labeled datasets *in vivo* and training deep learning models allows for either direct approximation of the operation [87] with faster computation times or acceleration of the convergence of an iterative process [61]. Efforts to accelerate the processing of ULM have also taken the form of directly forming the ULM image with averaged B-mode images [65, 82, 153]. In addition to drastically reducing acquisition and processing time [65, 153], such approaches could also be facilitates ULM application in clinical setting where available echographs are limited to frame rates much lower than typical ULM settings. You et al. [185] have also leveraged ULM data and adversarial

learning *in vivo* to enhance the resolution of Contrast Free Power Doppler, suggesting that the range of deep learning in ULM application could be broader than expected.

Along with the successes demonstrating the potential of deep learning in ULM, further developments are anticipated to further enhance ULM's capabilities and applicability. As the number of publicly available datasets and the interest of the research community increase, the performance of deep learning approaches is expected to improve accordingly. Leveraging larger datasets from multiple sources could enable the training of ULM deep learning models that are robust to varying experimental setups and consistently outperform conventional ULM without intensive parameter tuning. The creation of larger training sets and improved deep learning formulations could also allow ULM to perform effectively on noisy or aberrated acquisition, enhancing the reproducibility of ULM imaging and its impact on pre-clinical and clinical studies. Applying deep learning to steps of the ULM pipeline that are currently handled by conventional methods could further enhance image quality, practicality, or robustness. For instance, deep learning techniques for motion correction, extensively studied in elastography [186] using supervised [187, 188, 189, 190], semi-supervised [191, 192], or unsupervised learning [193, 194], could inspire the development of motion correction methods tailored specifically for ULM.

Deep learning in ULM may also have an impact on the adaptability of ULM to new experimental settings or transmission sequences. Application to *in vivo* data requires data-dependent tuning of conventional ULM, whereas some deep learning approaches required limited parameter tuning. Indeed, Liu et al. [64] have noted that their proposed mSPCN-ULM was robust to training parameters and provided better flexibility in implementation of ULM in comparison to conventional ULM. The authors mentioned that mSPCN-ULM still required preprocessed inputs, depending on some external parameters. Shin et al. [20] reported that their proposed LOCA-ULM needed to be retrained when ultrasound imaging settings were altered. Exploring ULM deep learning approaches robustness across different acquisitions, imaging settings, or organs could improve the adaptability of ULM to new experimental set-up and reduce the user input needed to form an ULM image. With the development of new transmission sequences [135, 195], deep learning approaches may better detect PSF with varying shape or microbubbles with "silenced" signal, given that the training set incorporated such patterns. Such benefits are more hypothetical, as constitution of training sets can be costly and including enough variability to reach a sufficient level of robustness might be unrealistic.

The current robustness of deep learning in ULM at high concentrations, combined with efforts in task compression within the ULM pipeline—such as training models to perform multiple stages simultaneously and advancements in embedded deep learning, could enable near real-time and online ULM in the near future, greatly improving the practicality of ULM.

References

[1] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nature Methods*, vol. 3, no. 10, pp. 793–795, Oct. 2006.

[2] E. Betzig *et al.*, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, no. 5793, pp. 1642–1645, Sep. 2006.

[3] C. Errico *et al.*, "Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging," *Nature*, vol. 527, no. 7579, pp. 499–502, Nov. 2015.

[4] K. Christensen-Jeffries, R. J. Browning, M. Tang, C. Dunsby, and R. J. Eckersley, "In Vivo Acoustic Super-Resolution and Super-Resolved Velocity Mapping using microbubbles," *IEEE Transactions on Medical Imaging*, vol. 34, no. 2, pp. 433–440, Feb. 2015.

[5] O. Couture, V. Hingot, B. Heiles, P. Muleki-Seya, and M. Tanter, "Ultrasound localization microscopy and super-resolution: A state of the art," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 8, pp. 1304–1320, Aug. 2018.

[6] B. Heiles *et al.*, "Ultrafast 3D ultrasound localization microscopy using a 32×32 matrix array," *IEEE Transactions on Medical Imaging*, vol. 38, no. 9, pp. 2005–2015, Sep. 2019.

[7] O. Demeulenaere *et al.*, "In Vivo Whole Brain Microvascular Imaging in mice using transcranial 3D Ultrasound Localization Microscopy," *eBioMedicine*, vol. 79, May 2022.

[8] B. Heiles *et al.*, "Volumetric Ultrasound Localization Microscopy of the Whole Rat Brain Microvasculature," *IEEE Open Journal of Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 2, pp. 261–282, 2022.

[9] A. Chavignon, B. Heiles, V. Hingot, C. Orset, D. Vivien, and O. Couture, "3D transcranial Ultrasound Localization Microscopy in the rat brain with a multiplexed matrix probe," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 7, pp. 2132–2142, Jul. 2022.

[10] J. Hansen-Shearer *et al.*, "Ultrafast 3-D transcutaneous super resolution ultrasound using row-column array specific coherence-based beamforming and rolling acoustic sub-aperture processing: In vitro, in rabbit and in human study," *Ultrasound in Medicine and Biology*, vol. 50, no. 7, pp. 1045–1057, Jul. 2024.

[11] J. A. Jensen *et al.*, "Anatomic and Functional Imaging Using Row–Column Arrays," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 10, pp. 2722–2738, Oct. 2022.

[12] A. Wu *et al.*, "3D transcranial Dynamic Ultrasound Localization Microscopy in the mouse brain using a Row-Column Array," Jun. 2024, arXiv:2406.01746.

[13] M. R. Lowerison *et al.*, "Super-Resolution Ultrasound Reveals Cerebrovascular Impairment in a Mouse Model of Alzheimer's Disease," *Journal of Neuroscience*, vol. 44, no. 9, Feb. 2024.

[14] A. Chavignon, V. Hingot, C. Orset, D. Vivien, and O. Couture, "3D transcranial Ultrasound Localization Microscopy for discrimination between ischemic and hemorrhagic stroke in early phase," *Scientific Reports*, vol. 12, no. 1, p. 14607, Aug. 2022.

[15] C. Bourquin, J. Porée, F. Lesage, and J. Provost, "In

vivo pulsatility measurement of cerebral microcirculation in rodents using dynamic ultrasound localization microscopy," *IEEE Transactions on Medical Imaging*, vol. 41, no. 4, pp. 782–792, Apr. 2022.

[16] C. Bourquin *et al.*, "Quantitative pulsatility measurements using 3D dynamic ultrasound localization microscopy," *Physics in Medicine & Biology*, vol. 69, no. 4, p. 045017, Feb. 2024.

[17] P. Cormier, J. Porée, C. Bourquin, and J. Provost, "Dynamic Myocardial Ultrasound Localization Angiography," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2021.

[18] J. Yan *et al.*, "Transthoracic ultrasound localization microscopy of myocardial vasculature in patients," *Nature Biomedical Engineering*, pp. 1–12, May 2024.

[19] N. Renaudin, C. Demené, A. Dizeux, N. Ialy-Radio, S. Pezet, and M. Tanter, "Functional ultrasound localization microscopy reveals brain-wide neurovascular activity on a microscopic scale," *Nature Methods*, vol. 19, no. 8, pp. 1004–1012, Aug. 2022.

[20] Y. Shin *et al.*, "Context-aware deep learning enables high-efficacy localization of high concentration microbubbles for super-resolution ultrasound localization microscopy," *Nature Communications*, vol. 15, no. 1, p. 2932, Apr. 2024.

[21] L. Denis *et al.*, "Sensing Ultrasound Localization Microscopy for the visualization of Glomeruli in living rats and humans," *eBioMedicine*, vol. 91, p. 104578, Apr. 2023.

[22] G. Chabouh *et al.*, "Whole organ volumetric sensing Ultrasound Localization Microscopy for characterization of kidney structure," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2024.

[23] C. Demené *et al.*, "Transcranial ultrafast Ultrasound Localization Microscopy of brain vasculature in patients," *Nature biomedical engineering*, vol. 5, no. 3, pp. 219–228, Mar. 2021.

[24] T. Opacic *et al.*, "Motion model ultrasound localization microscopy for preclinical and clinical multiparametric tumor characterization," *Nature Communications*, vol. 9, no. 1, p. 1527, Apr. 2018.

[25] C. Porte *et al.*, "Ultrasound Localization Microscopy for Breast Cancer Imaging in Patients: Protocol Optimization and Comparison with Shear Wave Elastography," *Ultrasound in Medicine & Biology*, vol. 50, no. 1, pp. 57–66, Jan. 2024.

[26] C. Huang *et al.*, "Super-resolution ultrasound localization microscopy based on a high frame-rate clinical ultrasound scanner: An in-human feasibility study," *Physics in Medicine & Biology*, vol. 66, no. 8, p. 08NT01, Apr. 2021.

[27] J. Zhu *et al.*, "Super-Resolution Ultrasound Localization Microscopy of Microvascular Structure and Flow for Distinguishing Metastatic Lymph Nodes – An Initial Human Study," *Ultraschall in der Medizin - European Journal of Ultrasound*, vol. 43, no. 6, pp. 592–598, Dec. 2022.

[28] S. Bodard *et al.*, "Ultrasound localization microscopy of the human kidney allograft on a clinical ultrasound scanner," *Kidney International*, vol. 103, no. 5, pp. 930–935, May 2023.

[29] O. Solomon, R. J. G. van Sloun, H. Wijkstra, M. Mischi, and Y. C. Eldar, "Exploiting Flow Dynamics for Superresolution in Contrast-Enhanced Ultrasound," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 10, pp. 1573–1586, Oct. 2019.

[30] S. Harput *et al.*, "Two-Stage Motion Correction for Super-Resolution Ultrasound Imaging in Human Lower Limb," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 5, pp. 803–814, May 2018.

[31] G. Goudot *et al.*, "Assessment of Takayasu's arteritis activity by ultrasound localization microscopy," *eBioMedicine*, vol. 90, Apr. 2023.

[32] M. Li *et al.*, "Super-resolution ultrasound localization microscopy for the non-invasive imaging of human testicular microcirculation and its differential diagnosis role in male infertility," *VIEW*, vol. 5, no. 2, p. 20230093, 2024.

[33] V. Hingot, C. Errico, B. Heiles, L. Rahal, M. Tanter, and O. Couture, "Microvascular flow dictates the compromise between spatial resolution and acquisition time in Ultrasound Localization Microscopy," *Scientific Reports*, vol. 9, no. 1, p. 2456, Feb. 2019.

[34] P. Song, J. M. Rubin, and M. R. Lowerison, "Super-resolution ultrasound microvascular imaging: Is it ready for clinical use?" *Zeitschrift für Medizinische Physik*, vol. 33, no. 3, pp. 309–323, Aug. 2023.

[35] R. S. Sutton. (2019) The bitter lesson. [Online]. Available: www.incompleteideas.net/IncIdeas/BitterLesson.html

[36] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[37] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[38] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *2015 IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile: IEEE, Dec. 2015, pp. 1026–1034.

[39] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[40] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2961–2969.

[41] I. Goodfellow *et al.*, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020.

[42] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion

Probabilistic Models," in *Advances in Neural Information Processing Systems*, vol. 33. Curran Associates, Inc., 2020, pp. 6840–6851.

[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[44] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2019.

[45] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," 2016.

[46] Y. Wu and K. He, "Group normalization," 2018.

[47] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, J. Fürnkranz and T. Joachims, Eds., 2010, pp. 807–814.

[48] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," 2017.

[49] F. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: continual prediction with lstm," in *1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)*, vol. 2, 1999, pp. 850–855 vol.2.

[50] A. Vaswani *et al.*, "Attention is All you Need," in *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc., 2017.

[51] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *International Conference on Learning Representations*, Oct. 2020.

[52] R. Bommasani *et al.*, "On the opportunities and risks of foundation models," Jul. 2022, arXiv:2108.07258.

[53] M. Awais *et al.*, "Foundational models defining a new era in vision: A survey and outlook," Jul. 2023, arXiv:2307.13721.

[54] A. Kirillov *et al.*, "Segment Anything," Apr. 2023, arXiv:2304.02643.

[55] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature Communications*, vol. 15, no. 1, p. 654, Jan. 2024.

[56] D. Hyun, L. L. Brickson, K. T. Looby, and J. J. Dahl, "Beamforming and Speckle Reduction Using Neural Networks," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 5, pp. 898–910, May 2019.

[57] M. Gasse, F. Millioz, E. Roux, D. Garcia, H. Liebgott, and D. Friboulet, "High-Quality Plane Wave Compounding Using Convolutional Neural Networks," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 64, no. 10, pp. 1637–1639, Oct. 2017.

[58] 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, Sep. 2018.

[59] H. Strohm, S. Rothlübbers, K. Eickel, and M. Günther, "Deep learning-based reconstruction of ultrasound images from raw channel data," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 9, pp. 1487–1490, 2020.

[60] J. Kim *et al.*, "Deep Learning-based 3D Beamforming on a 2D Row Column Addressing (RCA) Array for 3D Super-resolution Ultrasound Localization Microscopy," in *2022 IEEE International Ultrasonics Symposium (IUS)*, Oct. 2022, pp. 1–4.

[61] O. Solomon *et al.*, "Deep Unfolded Robust PCA With Application to Clutter Suppression in Ultrasound," *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, pp. 1051–1063, Apr. 2020.

[62] R. J. G. van Sloun, R. Cohen, and Y. C. Eldar, "Deep Learning in Ultrasound Imaging," *Proceedings of the IEEE*, vol. 108, no. 1, pp. 11–29, Jan. 2020.

[63] B. Luijten, N. Chennakeshava, Y. C. Eldar, M. Mischi, and R. J. G. van Sloun, "Ultrasound Signal Processing: From Models to Deep Learning," *Ultrasound in Medicine and Biology*, vol. 49, no. 3, pp. 677–698, Mar. 2023.

[64] X. Liu, T. Zhou, M. Lu, Y. Yang, Q. He, and J. Luo, "Deep Learning for Ultrasound Localization Microscopy," *IEEE Transactions on Medical Imaging*, vol. 39, no. 10, pp. 3064–3078, Oct. 2020.

[65] G. Zhang *et al.*, "ULM-MbCNRT: In vivo Ultrafast Ultrasound Localization Microscopy by Combining Multibranch CNN and Recursive Transformer," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, pp. 1–1, 2024.

[66] C. Hahne, G. Chabouh, A. Chavignon, O. Couture, and R. Sznitman, "RF-ULM: Ultrasound Localization Microscopy Learned from Radio-Frequency Wavefronts," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2024.

[67] L. Milecki *et al.*, "A Deep Learning Framework for Spatiotemporal Ultrasound Localization Microscopy," *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1428–1437, May 2021.

[68] Z. Zhang, M. Hwang, T. J. Kilbaugh, and J. Katz, "Improving sub-pixel accuracy in ultrasound localization microscopy using supervised and self-supervised deep learning," *Measurement Science and Technology*, vol. 35, no. 4, p. 045701, Apr. 2024.

[69] J. Youn, M. L. Ommen, M. B. Stuart, E. V. Thomsen, N. B. Larsen, and J. A. Jensen, "Detection and Localization of Ultrasound Scatterers Using Convolutional Neural Networks," *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 3855–3867, Dec. 2020.

[70] G. Zhang *et al.*, "*In Vivo* ultrasound localization microscopy for high-density microbubbles," *Ultrasonics*, vol. 143, p. 107410, Sep. 2024.

[71] P. Xing *et al.*, "Phase Aberration Correction for In Vivo Ultrasound Localization Microscopy Using a Spatiotemporal Complex-Valued Neural Network," *IEEE Transactions on Medical Imaging*, vol. 43, no. 2, pp. 662–673, Feb. 2024.

[72] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Machine Learning*, vol. 79, no. 1-2, pp. 151–175, May 2010.

[73] M. Lerendegui *et al.*, "ULTRA-SR Challenge: Assess-

ment of Ultrasound Localization and TRacking Algorithms for Super-Resolution Imaging," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2024.

[74] B. Heiles, A. Chavignon, V. Hingot, P. Lopez, E. Teston, and O. Couture, "Performance benchmarking of microbubble-localization algorithms for ultrasound localization microscopy," *Nature Biomedical Engineering*, vol. 6, no. 5, pp. 605–616, May 2022.

[75] B. H. Menze *et al.*, "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[76] O. Bernard *et al.*, "Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved?" *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2514–2525, Nov. 2018.

[77] H. Belgharbi *et al.*, "An anatomically realistic simulation framework for 3D ultrasound localization microscopy," *IEEE Open Journal of Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 3, pp. 1–13, 2023.

[78] N. Blanken, J. M. Wolterink, H. Delingette, C. Brune, M. Versluis, and G. Lajoinie, "Super-resolved microbubble localization in single-channel ultrasound rf signals using deep learning," *IEEE Transactions on Medical Imaging*, vol. 41, no. 9, pp. 2532–2542, Sep. 2022.

[79] X. Chen, M. R. Lowerison, Z. Dong, A. Han, and P. Song, "Deep Learning-Based Microbubble Localization for Ultrasound Localization Microscopy," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 4, pp. 1312–1325, Apr. 2022.

[80] J. A. Jensen, "Field: A Program for Simulating Ultrasound Systems: 10th Nordic-Baltic Conference on Biomedical Imaging," *Medical & Biological Engineering & Computing*, vol. 34, no. sup. 1, pp. 351–353, 1997.

[81] X. Chen, M. R. Lowerison, Z. Dong, N. V. Chandra Sekaran, D. A. Llano, and P. Song, "Localization free super-resolution microbubble velocimetry using a long short-term memory neural network," *IEEE Transactions on Medical Imaging*, vol. 42, no. 8, pp. 2374–2385, Aug. 2023.

[82] W. Gu, Z. Yan, B. Li, C. Liu, D. Ta, and X. Liu, "GAN-Based Ultrasound Localization Microscopy," in *2022 IEEE International Ultrasonics Symposium (IUS)*, Oct. 2022, pp. 1–4.

[83] X. Liu and M. Almekkawy, "Ultrasound Localization Microscopy Using Deep Neural Network," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 70, no. 7, pp. 625–635, Jul. 2023.

[84] S. Luan *et al.*, "Deep learning for fast super-resolution ultrasound microvessel imaging," *Physics in Medicine & Biology*, vol. 68, no. 24, p. 245023, Dec. 2023.

[85] D. Garcia, "SIMUS: An open-source simulator for medical ultrasound imaging. Part I: Theory & examples," *Computer Methods and Programs in Biomedicine*, vol. 218, p. 106726, May 2022.

[86] V. Pustovalov, D.-H. Pham, and D. Kouamé, "Enhanced Localization in Ultrafast Ultrasound Imaging through Spatio-Temporal Deep Learning," in *32nd European Signal Processing Conference (EUSIPCO 2024)*, vol. TU1.SC1: Advances in Computational Ultrasound Imaging, Lyon, France, Aug. 2024, p. TU1.SC1.6.

[87] X. Yu *et al.*, "Deep learning for fast denoising filtering in ultrasound localization microscopy," *Physics in Medicine & Biology*, vol. 68, no. 20, p. 205002, Oct. 2023.

[88] R. J. G. van Sloun *et al.*, "Super-Resolution Ultrasound Localization Microscopy Through Deep Learning," *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 829–839, Mar. 2021.

[89] U.-W. Lok *et al.*, "Fast super-resolution ultrasound microvessel imaging using spatiotemporal data with deep fully convolutional neural network," *Physics in Medicine & Biology*, vol. 66, no. 7, p. 075005, Mar. 2021.

[90] X. Liu and M. Almekkawy, "Ultrasound Super Resolution using Vision Transformer with Convolution Projection Operation," in *2022 IEEE International Ultrasonics Symposium (IUS)*, Oct. 2022, pp. 1–4.

[91] V. Hingot, A. Chavignon, B. Heiles, and O. Couture, "Measuring Image Resolution in Ultrasound Localization Microscopy," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3812–3819, Dec. 2021.

[92] R. Damseh *et al.*, "Automatic Graph-Based Modeling of Brain Microvessels Captured With Two-Photon Microscopy," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2551–2562, Nov. 2019.

[93] M. Lerendegui, K. Riemer, B. Wang, C. Dunsby, and M.-X. Tang, "BUbble Flow Field: A Simulation Framework for Evaluating Ultrasound Localization Microscopy Algorithms," Nov. 2022, arXiv:2211.00754.

[94] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2017, pp. 23–30.

[95] B. Mehta, M. Diaz, F. Golemo, C. J. Pal, and L. Paull, "Active Domain Randomization," in *Proceedings of the Conference on Robot Learning*. PMLR, May 2020, pp. 1162–1176.

[96] B. E. Treeby, J. Budisky, E. S. Wise, J. Jaros, and B. T. Cox, "Rapid calculation of acoustic fields from arbitrary continuous-wave sources," *The Journal of the Acoustical Society of America*, vol. 143, no. 1, pp. 529–537, Jan. 2018.

[97] P. Marmottant *et al.*, "A model for large amplitude oscillations of coated bubbles accounting for buckling and rupture," *The Journal of the Acoustical Society of America*, vol. 118, no. 6, pp. 3499–3505, Dec. 2005.

[98] J. N. Harmon, Z. Z. Khaing, J. E. Hyde, C. P. Hofstetter, C. Tremblay-Darveau, and M. F. Bruce, "Quantitative tissue perfusion imaging using nonlinear ultrasound

localization microscopy," *Scientific Reports*, vol. 12, no. 1, p. 21943, Dec. 2022.

[99] J. Brown, K. Christensen-Jeffries, S. Harput, C. Dunsby, M. X. Tang, and R. J. Eckersley, "Investigation of microbubble detection methods for super-resolution imaging of microvasculature," in *2017 IEEE International Ultrasonics Symposium (IUS)*, Sep. 2017, pp. 1–4.

[100] D. T. Blackstock, "Generalized Burgers equation for plane waves," *The Journal of the Acoustical Society of America*, vol. 77, no. 6, pp. 2050–2053, Jun. 1985.

[101] J. Jiménez-Fernández, "Nonlinear response to ultrasound of encapsulated microbubbles," *Ultrasonics*, vol. 52, no. 6, pp. 784–793, Aug. 2012.

[102] W. T. Shi and F. Forsberg, "Ultrasonic characterization of the nonlinear properties of contrast microbubbles," *Ultrasound in Medicine & Biology*, vol. 26, no. 1, pp. 93–104, Jan. 2000.

[103] J. Zhang, Q. He, C. Wang, H. Liao, and J. Luo, "A General Framework for Inverse Problem Solving using Self-Supervised Deep Learning: Validations in Ultrasound and Photoacoustic Image Reconstruction," Oct. 2021, arXiv:2110.14970.

[104] Y. Li, L. Huang, J. Zhang, C. Huang, S. Chen, and J. Luo, "Localization of High-concentration Microbubbles for Ultrasound Localization Microscopy by Self-Supervised Deep Learning," in *2021 IEEE International Ultrasonics Symposium (IUS)*, Sep. 2021, pp. 1–4.

[105] S. Lei *et al.*, "In Vivo Ultrasound Localization Microscopy Imaging of the Kidney's Microvasculature With Block-Matching 3-D Denoising," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 2, pp. 523–533, Feb. 2022.

[106] A. Leconte *et al.*, "A Tracking prior to Localization workflow for Ultrasound Localization Microscopy," Aug. 2023, arXiv:2308.02724.

[107] Y. Zhang *et al.*, "Efficient Microbubble Trajectory Tracking in Ultrasound Localization Microscopy Using a Gated Recurrent Unit-Based Multitasking Temporal Neural Network," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, pp. 1–1, 2024.

[108] G. Ng, S. Worrell, P. Freiburger, and G. Trahey, "A comparative evaluation of several algorithms for phase aberration correction," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 41, no. 5, pp. 631–643, Sep. 1994.

[109] L. Nock, G. E. Trahey, and S. W. Smith, "Phase aberration correction in medical ultrasound using speckle brightness as a quality factor," *The Journal of the Acoustical Society of America*, vol. 85, no. 5, pp. 1819–1833, May 1989.

[110] M. O'Donnell and S. Flax, "Phase-aberration correction using signals from point reflectors and diffuse scatterers: Measurements," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 35, no. 6, pp. 768–774, Nov. 1988.

[111] G. Montaldo, M. Tanter, and M. Fink, "Time Reversal of Speckle Noise," *Physical Review Letters*, vol. 106, no. 5, p. 054301, Feb. 2011.

[112] B.-F. Osmanski, G. Montaldo, M. Tanter, and M. Fink, "Aberration correction by time reversal of moving speckle noise," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 59, no. 7, pp. 1575–1583, Jul. 2012.

[113] W. Lambert, L. A. Cobus, M. Couade, M. Fink, and A. Aubry, "Reflection Matrix Approach for Quantitative Imaging of Scattering Media," *Physical Review X*, vol. 10, no. 2, p. 021048, Jun. 2020.

[114] W. Lambert, L. A. Cobus, T. Frappart, M. Fink, and A. Aubry, "Distortion matrix approach for ultrasound imaging of random scattering media," *Proceedings of the National Academy of Sciences*, vol. 117, no. 26, pp. 14 645–14 656, Jun. 2020.

[115] W. Lambert, L. A. Cobus, J. Robin, M. Fink, and A. Aubry, "Ultrasound Matrix Imaging—Part II: The Distortion Matrix for Aberration Correction Over Multiple Isoplanatic Patches," *IEEE Transactions on Medical Imaging*, vol. 41, no. 12, pp. 3921–3938, Dec. 2022.

[116] M. Feigin, D. Freedman, and B. W. Anthony, "A Deep Learning Framework for Single-Sided Sound Speed Inversion in Medical Ultrasound," *IEEE transactions on bio-medical engineering*, vol. 67, no. 4, pp. 1142–1151, Apr. 2020.

[117] M. Sharifzadeh, H. Benali, and H. Rivaz, "Phase Aberration Correction: A Convolutional Neural Network Approach," *IEEE Access*, vol. 8, pp. 162 252–162 260, 2020.

[118] W. A. Simson, M. Paschali, V. Sideri-Lampretsa, N. Navab, and J. J. Dahl, "Investigating pulse-echo sound speed estimation in breast ultrasound with deep learning," *Ultrasonics*, vol. 137, p. 107179, Feb. 2024.

[119] Z. Tian, M. Olmstead, Y. Jing, and A. Han, "Transcranial Phase Correction Using Pulse-Echo Ultrasound and Deep Learning: A 2-D Numerical Study," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 71, no. 1, pp. 117–126, Jan. 2024.

[120] V. Hingot, C. Errico, M. Tanter, and O. Couture, "Subwavelength motion-correction for ultrafast ultrasound localization microscopy," *Ultrasonics*, vol. 77, pp. 17–21, May 2017.

[121] J. R. McCall, F. Santibanez, H. Belgharbi, G. F. Pinton, and P. A. Dayton, "Non-invasive transcranial volumetric ultrasound localization microscopy of the rat brain with continuous, high volume-rate acquisition," *Theranostics*, vol. 13, no. 4, pp. 1235–1246, Feb. 2023.

[122] J. Robin *et al.*, "In vivo adaptive focusing for clinical contrast-enhanced transcranial ultrasound imaging in human," *Physics in Medicine & Biology*, vol. 68, no. 2, p. 025019, Jan. 2023.

[123] P. Xing, A. Malescot, E. Martineau, R. Rungta, and J. Provost, "Inverse Problem Based on a Sparse Representation of Contrast-enhanced Ultrasound Data for in vivo Transcranial Imaging," Jan. 2024, arXiv:2401.10389.

[124] P. Xing, V. Perrot, A. U. Dominguez-Vargas, S. Quessy, N. Dancause, and J. Provost, "Towards Transcranial 3D Ultrasound Localization Microscopy of the Nonhuman

Primate Brain," Apr. 2024, arXiv:2404.03547.

[125] C. Demené *et al.*, "Spatiotemporal Clutter Filtering of Ultrafast Ultrasound Data Highly Increases Doppler and fUltrasound Sensitivity," *IEEE Transactions on Medical Imaging*, vol. 34, no. 11, pp. 2271–2285, Nov. 2015.

[126] A. Beck and M. Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, Jan. 2009.

[127] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 770–778.

[128] O. M. Viessmann, R. J. Eckersley, K. Christensen-Jeffries, M. X. Tang, and C. Dunsby, "Acoustic super-resolution with ultrasound and microbubbles," *Physics in Medicine and Biology*, vol. 58, no. 18, pp. 6447–6458, Sep. 2013.

[129] Y. Desailly, O. Couture, M. Fink, and M. Tanter, "Sono-activated ultrasound localization microscopy," *Applied Physics Letters*, vol. 103, no. 17, p. 174107, Oct. 2013.

[130] M. A. O'Reilly and K. Hynynen, "A super-resolution ultrasound method for brain vascular mapping," *Medical Physics*, vol. 40, no. 11, p. 110701, Nov. 2013.

[131] G. Montaldo, M. Tanter, J. Bercoff, N. Benech, and M. Fink, "Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, no. 3, pp. 489–506, Mar. 2009.

[132] R. Liu *et al.*, "An intriguing failing of convolutional neural networks and the CoordConv solution," in *Advances in Neural Information Processing Systems*, vol. 31. Curran Associates, Inc., 2018.

[133] J. Foiret, H. Zhang, T. Ilovitsh, L. Mahakian, S. Tam, and K. W. Ferrara, "Ultrasound localization microscopy to image and assess microvasculature in a rat kidney," *Scientific Reports*, vol. 7, no. 1, p. 13662, Oct. 2017.

[134] T. M. Kierski *et al.*, "Super harmonic ultrasound for motion-independent localization microscopy: Applications to microvascular imaging from low to high flow rates," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 5, pp. 957–967, May 2020.

[135] A. Coudert *et al.*, "3D Transcranial Ultrasound Localization Microscopy reveals major arteries in the sheep brain," Mar. 2024.

[136] K. G. Brown, D. Ghosh, and K. Hoyt, "Deep Learning of Spatiotemporal Filtering for Fast Super-Resolution Ultrasound Imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 9, pp. 1820–1829, Sep. 2020.

[137] W. Han, Y. Zhang, Y. Zhao, A. Luo, and B. Peng, "3D U-Net3+ Based Microbubble Filtering for Ultrasound Localization Microscopy," in *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct. 2023, pp. 3974–3979.

[138] P. Song *et al.*, "Improved Super-Resolution Ultrasound Microvessel Imaging With Spatiotemporal Nonlocal Means Filtering and Bipartite Graph-Based Microbubble Tracking," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 2, pp. 149–167, Feb. 2018.

[139] R. Gu *et al.*, "Contrastive Semi-Supervised Learning for Domain Adaptive Segmentation Across Similar Anatomical Structures," *IEEE Transactions on Medical Imaging*, vol. 42, no. 1, pp. 245–256, Jan. 2023.

[140] S. Tang *et al.*, "Kalman Filter-Based Microbubble Tracking for Robust Super-Resolution Ultrasound Microvessel Imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 9, pp. 1738–1751, Sep. 2020.

[141] I. Taghavi *et al.*, "Ultrasound super-resolution imaging with a hierarchical Kalman tracker," *Ultrasonics*, vol. 122, p. 106695, May 2022.

[142] G. Revach, N. Shlezinger, X. Ni, A. L. Escoriza, R. J. G. van Sloun, and Y. C. Eldar, "KalmanNet: Neural Network Aided Kalman Filtering for Partially Known Dynamics," *IEEE Transactions on Signal Processing*, vol. 70, pp. 1532–1547, Jan. 2022.

[143] T. S. Stevens *et al.*, "A Hybrid Deep Learning Pipeline for Improved Ultrasound Localization Microscopy," in *2022 IEEE International Ultrasonics Symposium (IUS)*, Oct. 2022, pp. 1–4.

[144] K. Cho *et al.*, "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, A. Moschitti, B. Pang, and W. Daelemans, Eds. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1724–1734.

[145] Y. Sui, X. Guo, J. Yu, D. Ta, and K. Xu, "Generative Adversarial Nets for Ultrafast Ultrasound Localization Microscopy Reconstruction," in *2022 IEEE International Ultrasonics Symposium (IUS)*, Oct. 2022, pp. 1–4.

[146] B. Rauby, P. Xing, J. Porée, M. Gasse, and J. Provost, "Pruning Sparse Tensor Neural Networks Enables Deep Learning for 3D Ultrasound Localization Microscopy," Feb. 2024, arXiv:2402.09359.

[147] T. Tong, G. Li, X. Liu, and Q. Gao, "Image Super-Resolution Using Dense Skip Connections," in *2017 IEEE International Conference on Computer Vision (ICCV)*. Venice: IEEE, Oct. 2017, pp. 4809–4817.

[148] Y. Shu, C. Han, M. Lv, and X. Liu, "Fast Super-Resolution Ultrasound Imaging With Compressed Sensing Reconstruction Method and Single Plane Wave Transmission," *IEEE Access*, vol. 6, pp. 39 298–39 306, 2018.

[149] C. Huang *et al.*, "Short Acquisition Time Super-Resolution Ultrasound Microvessel Imaging via Microbubble Separation," *Scientific Reports*, vol. 10, no. 1, p. 6007, Apr. 2020.

[150] B. Heiles, A. Chavignon, V. Hingot, P. Lopez, E. Teston, and O. Couture, "Addendum: Performance benchmark-

ing of microbubble-localization algorithms for ultrasound localization microscopy," *Nature Biomedical Engineering*, pp. 1–1, Oct. 2023.

[151] G. Tuccio, S. Afrakhteh, G. Iacca, and L. Demi, "Time Efficient Ultrasound Localization Microscopy Based on A Novel Radial Basis Function 2D Interpolation," *IEEE Transactions on Medical Imaging*, vol. 43, no. 5, pp. 1690–1701, May 2024.

[152] J. Youn, M. L. Ommen, M. B. Stuart, E. V. Thomsen, N. B. Larsen, and J. A. Jensen, "Ultrasound Multiple Point Target Detection and Localization using Deep Learning," in *2019 IEEE International Ultrasonics Symposium (IUS)*, Oct. 2019, pp. 1937–1940.

[153] W. Gu, B. Li, J. Luo, Z. Yan, D. Ta, and X. Liu, "Ultrafast Ultrasound Localization Microscopy by Conditional Generative Adversarial Network," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 70, no. 1, pp. 25–40, Jan. 2023.

[154] C. Trabelsi *et al.*, "Deep Complex Networks," in *International Conference on Learning Representations*, Feb. 2018.

[155] J. Lu, F. Millioz, D. Garcia, S. Salles, D. Ye, and D. Friboulet, "Complex Convolutional Neural Networks for Ultrafast Ultrasound Imaging Reconstruction From In-Phase/Quadrature Signal," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 2, pp. 592–603, Feb. 2022.

[156] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[157] H. Lee, S.-H. Oh, M.-G. Kim, Y.-M. Kim, G. Jung, and H.-M. Bae, "Optical Flow Assisted Super-Resolution Ultrasound Localization Microscopy using Deep Learning," in *2022 IEEE International Ultrasonics Symposium (IUS)*, Oct. 2022, pp. 1–4.

[158] A. Speiser *et al.*, "Deep learning enables fast and dense single-molecule localization with high accuracy," *Nature Methods*, vol. 18, no. 9, pp. 1082–1090, Sep. 2021.

[159] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, Oct. 2016, pp. 565–571.

[160] Y. Zhao, S. Liu, A. Luo, and B. Peng, "Dual Generative Adversarial Network For Ultrasound Localization Microscopy," in *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct. 2022, pp. 3125–3130.

[161] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein Generative Adversarial Networks," in *Proceedings of the 34th International Conference on Machine Learning*. PMLR, Jul. 2017, pp. 214–223.

[162] Y. Shin *et al.*, "Context-Aware Deep Learning Enables High-Efficacy Localization of High Concentration Microbubbles for Super-Resolution Ultrasound Localization Microscopy," p. 2023.04.21.536599, Apr. 2023.

[163] X. Liu and M. Almekkawy, "Ultrasound Microbubbles Localization Using Object Detection Model," in *2023 IEEE International Ultrasonics Symposium (IUS)*, Sep. 2023, pp. 1–4.

[164] S. K. Gharamaleki, B. Helfield, and H. Rivaz, "Transformer-Based Microbubble Localization," in *2022 IEEE International Ultrasonics Symposium (IUS)*, Oct. 2022, pp. 1–4.

[165] A. Aitken, C. Ledig, L. Theis, J. Caballero, Z. Wang, and W. Shi, "Checkerboard artifact free sub-pixel convolution: A note on sub-pixel convolution, resize convolution and convolution resize," Jul. 2017, arXiv:1707.02937.

[166] W. Shi *et al.*, "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 1874–1883.

[167] X. Liu and M. Almekkawy, "Ultrasound Super Resolution Using Deep Learning Based on Attention Mechanism," in *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI), 18-21 April 2023*, ser. 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). Piscataway, NJ, USA: IEEE, 2023, pp. 1–5.

[168] S. K. Gharamaleki, B. Helfield, and H. Rivaz, "Deformable-Detection Transformer for Microbubble Localization in Ultrasound Localization Microscopy," in *2023 IEEE International Ultrasonics Symposium (IUS)*, Sep. 2023, pp. 1–4.

[169] G. Zhang, Y. Yue, F. Dai, X. Liu, and D. Ta, "Transformer for Ultrafast Ultrasound Localization Microscopy," in *2022 IEEE International Ultrasonics Symposium (IUS)*, Oct. 2022, pp. 1–4.

[170] T. Zhou, M. Lu, Y. Yang, Q. He, J. Luo, and X. Liu, "qULM-DL: Quantitative Ultrasound Localization Microscopy via Deep Learning," in *2020 IEEE International Ultrasonics Symposium (IUS)*. Las Vegas, NV, USA: IEEE, Sep. 2020, pp. 1–4.

[171] Z. Liu *et al.*, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, QC, Canada: IEEE, Oct. 2021, pp. 9992–10 002.

[172] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-End Object Detection with Transformers," in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham: Springer International Publishing, 2020, vol. 12346, pp. 213–229.

[173] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable DETR: Deformable Transformers for End-to-End Object Detection," in *International Conference on Learning Representations*, Oct. 2020.

[174] U.-W. Lok *et al.*, "Three-Dimensional Ultrasound Localization Microscopy with Bipartite Graph-Based Microbubble Pairing and Kalman-Filtering-Based Tracking on a 256-Channel Verasonics Ultrasound System with a 32×32 Matrix Array," *Journal of Medical and Bio-*

logical Engineering, vol. 42, no. 6, pp. 767–779, Dec. 2022.

[175] M. Piepenbrock, D. Koretskaia, G. Schmitz, and S. Dencks, "3D Microbubble Localization with a Convolutional Neural Network for Super-Resolution Ultrasound Imaging," in *2021 IEEE International Ultrasonics Symposium (IUS)*. Xi'an, China: IEEE, Sep. 2021, pp. 1–4.

[176] C. Choy, J. Gwak, and S. Savarese, "4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA: IEEE, Jun. 2019, pp. 3070–3079.

[177] M. R. Lowerison *et al.*, "Aging-related cerebral microvascular changes visualized using ultrasound localization microscopy in the living mouse," *Scientific Reports*, vol. 12, no. 1, p. 619, Dec. 2022.

[178] M. Wiersma, B. Heiles, D. Kalisvaart, D. Maresca, and C. S. Smith, "Retrieving Pulsatility in Ultrasound Localization Microscopy," *IEEE Open Journal of Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 2, pp. 283–298, 2022.

[179] M. van den Kerkhof *et al.*, "Impaired damping of cerebral blood flow velocity pulsatility is associated with the number of perivascular spaces as measured with 7T MRI," *Journal of Cerebral Blood Flow & Metabolism*, vol. 43, no. 6, pp. 937–946, Jun. 2023.

[180] A. E. Roher *et al.*, "Transcranial Doppler ultrasound blood flow velocity and pulsatility index as systemic indicators for Alzheimer's disease," *Alzheimer's & Dementia*, vol. 7, no. 4, pp. 445–455, 2011.

[181] R. G. Gosling and D. H. King, "Arterial Assessment by Doppler-shift Ultrasound," *Proceedings of the Royal Society of Medicine*, vol. 67, no. 6 Pt 1, pp. 447–449, Jun. 1974.

[182] C.-P. Chung, H.-Y. Lee, P.-C. Lin, and P.-N. Wang, "Cerebral Artery Pulsatility is Associated with Cognitive Impairment and Predicts Dementia in Individuals with Subjective Memory Decline or Mild Cognitive Impairment," *Journal of Alzheimer's Disease*, vol. 60, no. 2, pp. 625–632, Jan. 2017.

[183] A. Reinke *et al.*, "Understanding metric-related pitfalls in image analysis validation," *Nature Methods*, vol. 21, no. 2, pp. 182–194, Feb. 2024.

[184] L. Maier-Hein *et al.*, "Why rankings of biomedical image analysis competitions should be interpreted with care," *Nature Communications*, vol. 9, no. 1, p. 5217, Dec. 2018.

[185] Q. You *et al.*, "Contrast-Free Super-Resolution Power Doppler (CS-PD) Based on Deep Neural Networks," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, pp. 1355–68, 2023.

[186] M. Ashikuzzaman, A. Héroux, A. Tang, G. Cloutier, and H. Rivaz, "Displacement Tracking Techniques in Ultrasound Elastography: From Cross Correlation to Deep Learning," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 71, no. 7, pp. 842–871, Jul. 2024.

[187] B. Peng, Y. Xian, and J. Jiang, "A Convolution Neural Network-Based Speckle Tracking Method for Ultrasound Elastography," in *2018 IEEE International Ultrasonics Symposium (IUS)*, Oct. 2018, pp. 206–212.

[188] B. Peng, Y. Xian, Q. Zhang, and J. Jiang, "Neural-network-based Motion Tracking for Breast Ultrasound Strain Elastography: An Initial Assessment of Performance and Feasibility," *Ultrasonic Imaging*, vol. 42, no. 2, pp. 74–91, Mar. 2020.

[189] M. G. Kibria and H. Rivaz, "GLUENet: Ultrasound Elastography Using Convolutional Neural Network," in *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation*, D. Stoyanov *et al.*, Eds. Cham: Springer International Publishing, 2018, pp. 21–28.

[190] A. K. Z. Tehrani and H. Rivaz, "Displacement Estimation in Ultrasound Elastography Using Pyramidal Convolutional Neural Network," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 12, pp. 2629–2639, Dec. 2020.

[191] A. K. Z. Tehrani, M. Sharifzadeh, E. Boctor, and H. Rivaz, "Bi-Directional Semi-Supervised Training of Convolutional Neural Networks for Ultrasound Elastography Displacement Estimation," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 4, pp. 1181–1190, Apr. 2022.

[192] A. K. Z. Tehrani, M. Mirzaei, and H. Rivaz, "Semi-supervised Training of Optical Flow Convolutional Neural Networks in Ultrasound Elastography," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, A. L. Martel *et al.*, Eds. Cham: Springer International Publishing, 2020, pp. 504–513.

[193] R. Delaunay, Y. Hu, and T. Vercauteren, "An unsupervised learning approach to ultrasound strain elastography with spatio-temporal consistency," *Physics in Medicine & Biology*, vol. 66, no. 17, p. 175031, Sep. 2021.

[194] X. Wei *et al.*, "Unsupervised Convolutional Neural Network for Motion Estimation in Ultrasound Elastography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 7, pp. 2236–2247, Jul. 2022.

[195] J. Kim *et al.*, "Improved Ultrasound Localization Microscopy Based on Microbubble Uncoupling via Transmit Excitation," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 3, pp. 1041–1052, Mar. 2022.