| | |
|---|---|
| **Titre:** Title: | Apprentissage profond pour la microscopie de localisation ultrasonore : imagerie volumique, robustesse in vivo et jeu de données de grande échelle |
| **Auteur:** Author: | Brice Rauby |
| **Date:** | 2026 |
| **Type:** | Mémoire ou thèse / Dissertation or Thesis |
| **Référence:** Citation: | Rauby, B. (2026). Apprentissage profond pour la microscopie de localisation ultrasonore : imagerie volumique, robustesse in vivo et jeu de données de grande échelle [Thèse de doctorat, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/73180/ |

**POLYTECHNIQUE MONTRÉAL**
affiliée à l'Université de Montréal

# Apprentissage profond pour la microscopie de localisation ultrasonore : imagerie volumique, robustesse in vivo et jeu de données de grande échelle

**BRICE RAUBY**
Institut de génie biomédical

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*
Génie biomédical
Février 2026

**POLYTECHNIQUE MONTRÉAL**
affiliée à l'Université de Montréal

Cette thèse intitulée :

**Apprentissage profond pour la microscopie de localisation ultrasonore : imagerie volumique, robustesse in vivo et jeu de données de grande échelle**

présentée par **Brice RAUBY**
en vue de l'obtention du diplôme de *Philosophiæ Doctor*
a été dûment acceptée par le jury d'examen constitué de :

**Lucien WEISS**, président
**Jean PROVOST**, membre et directeur de recherche
**Maxime GASSE**, membre et codirecteur de recherche
**Hervé LOMBAERT**, membre
**Hassan RIVAZ**, membre externe

# DÉDICACE

*À mes grands-parents dont la gentillesse,*
*l'abnégation et l'obstination m'inspirent chaque jour.*

# REMERCIEMENTS

En premier lieu, je tiens à exprimer ma profonde gratitude à mon directeur de thèse, Jean. C'est en arrivant au terme de ce long périple doctoral que je mesure pleinement l'ampleur de ce que tu m'as apporté. J'espère que ce manuscrit sera à la hauteur de tes attentes et que tu jugeras que c'est une "bonne thèse". Je dois reconnaître que tu as réussi le tour de force de me réconcilier avec la physique, une discipline que j'appréhendais pourtant avec une certaine réserve avant mon arrivée au laboratoire. Merci pour tes conseils scientifiques, tes orientations de recherche, tes relectures — parfois jusqu'à la dernière minute — et pour avoir toujours su monter la barre juste assez haut. Finalement, plus qu'un directeur, tu auras été pour moi un véritable mentor.

Ce travail n'aurait pas eu la même saveur sans l'appui précieux de Maxime, mon co-directeur. Merci pour les discussions sur les dernières nouveautés du deep learning et ton apport technique qui aura été précieux. Nos réunions en prenant un petit café à ServiceNow et tes séances de coaching d'escalade resteront des moments privilégiés de ces dernières années (même si, en fin de compte, je pense que je préfère monter en courant).

La maturation des idées présentes (ou non) dans cette thèse doit aussi beaucoup à mes collègues de laboratoire. Jonathan, Alice, Paul, Alexis, Samuel, Nin, Gerardo, Vincent et Stephen, merci pour vos lumières sur ces sujets dont la clé résidait, pour moi, à mi-chemin entre magie noire et connaissances élémentaires d'ultrasons. Cependant, nos interactions ne se sont pas limitées à des discussions scientifiques. Jonathan, merci pour ton humour et tes vannes caustiques. Si tu t'étais dépêché de jouer au tarot, cette thèse serait finie depuis un an ; si tu n'avais pas été là, elle n'aurait certainement pas commencé. Nin, merci de ton entrain et de ton soutien pour toutes les tâches de nettoyage. De la vaisselle du lab au jeu de données ULM sur le nearline en passant par pipULM : on en aura remué de la poussière. La team espresso, Alice et Alexis, la grande et le petit, vous aurez été une source d'inspiration et de réconfort dans les moments les plus bas du début d'avril ou de la fin août. Alors un grand merci pour ces cafés partagés et les appels à la cafét millimétrés. C'est bientôt votre tour de leur montrer de quel bois les doctorants se chauffent. Mich-Mich, Sam, Hatim, Chloé, Oleksandra et Olivier, merci d'avoir égayé ces années covid et post-covid avec des bons desserts, du ski de fond, des conseils de présentation, de la convivialité, du pragmatisme déconcertant et un tapis moush-moush.

Merci à Amal, pour sa compréhension et son aide précieuse dans toutes les étapes administratives de la poursuite d'un doctorat.

Merci aussi à ma famille, en particulier à mes parents d'avoir tout fait pour me mettre dans les meilleures conditions, et à mon frère d'avoir montré l'exemple sur bien des aspects.

On dit souvent que la thèse ce n'est pas un sprint mais un marathon ; dans mon cas, c'était quand même un poil plus long. Alors merci à tous les amis qui m'ont aidé, défié, inspiré ou soutenu pour courir toujours plus loin et toujours plus vite : Matt, Clém, Marjo, François, Jon, Albin, Marvin, Guillaume, Jess, Julien, Maxime, Valentin, Julio ainsi que tous les amis du Lundi1000 et de RIM.

Parce que s'époumoner à la Montagne Coupée ou se mettre capot ouvert sur le Mont-Royal restera le meilleur moyen de se remonter le moral quand la recherche n'avance pas : un grand merci à Nico et Ray, mais surtout à tous les amis du CSFUM. En particulier Vincent, Lauriane, Aurel, Raph, Paul, Sandrine, Noémie, Michou, Fab, Kristina, Justine et Koldo, merci de votre amitié et de votre énergie.

Un grand merci à Lesly, que ce soit pour un ravitaillement à La Chouette, à St-Tite ou au Pain dans les Voiles entre deux tomates à Thèsez-vous, ton soutien m'aura bien aidé à arriver au bout (cette fois, au moins).

Finalement, merci Mélanie, de m'avoir guidé dans mes choix, écouté et soutenu dans les moments difficiles. C'est fini, je ne t'embêterai plus avec l'alignement de mes simulations ou l'allure de mon bruit.

# RÉSUMÉ

**Contexte et Problématique** L'imagerie de la microvascularisation cérébrale est fondamentale pour le diagnostic précoce et la compréhension de nombreuses pathologies, notamment les maladies neurodégénératives comme Alzheimer, où les atteintes capillaires précèdent souvent les dommages tissulaires. Les modalités cliniques actuelles (IRM, CT, échographie classique) échouent à résoudre ces échelles microscopiques en profondeur. Grâce à la localisation de microbulles injectées dans le sang, la Microscopie de Localisation Ultrasonore (ULM) permet l'observation du réseau vasculaire profond à des échelles de l'ordre de la dizaine de microns *in vivo*. Néanmoins, l'adoption généralisée de l'ULM reste entravée par des limites majeures : des temps d'acquisition prohibitifs nécessaires pour capturer les microbulles diluées, un volume de données massif difficile à gérer, et une sensibilité critique à certains paramètres expérimentaux.

**Hypothèse et Méthodologie** Cette thèse repose sur l'hypothèse que l'apprentissage profond (*Deep Learning*) peut surmonter ces compromis physiques en modélisant les signaux complexes de microbulles à haute concentration, là où les méthodes conventionnelles échouent. Cependant, elle identifie un obstacle central : l'écart de distribution (*domain shift*) entre les simulations utilisées pour l'entraînement et la réalité expérimentale *in vivo*, qui limite la généralisation des modèles. Cette thèse se découpe en quatre contributions majeures.

**Structuration du domaine** Une analyse critique de la littérature a permis de cibler les applications de l'apprentissage profond ayant le plus fort impact potentiel, et d'identifier formellement la rupture de l'hypothèse i.i.d. (indépendante et identiquement distribuée) comme un frein majeur à l'application *in vivo* des modèles existants. Cette analyse a également souligné l'absence de solutions pour l'ULM 3D ainsi que le manque de méthodes d'évaluation standardisées pour les approches d'apprentissage.

**Passage à l'échelle 3D par parcimonie** Pour rendre l'ULM 3D accessible, cette thèse introduit l'utilisation de réseaux de neurones sur tenseurs parcimonieux (*Sparse Tensor Neural Networks*). Cette architecture exploite la parcimonie spatio-temporelle des microbulles pour réduire la consommation mémoire de deux ordres de grandeur en 3D par rapport aux méthodes d'apprentissage profond denses conventionnelles. Elle permet non seulement l'apprentissage profond sur des données volumiques, mais étend également à la 3D la capacité de résoudre implicitement le chevauchement des microbulles à haute concentration, ouvrant

ainsi la voie à des acquisitions volumiques rapides.

**Apprentissage *In Vivo* robuste (*Teacher-Student*)** Pour combler l'écart entre données d'entraînement simulées et application *in vivo*, une approche *Teacher-Student* a été développée pour entraîner les modèles directement sur des données expérimentales, sans recours aux simulations. Cette méthode a démontré une robustesse accrue au bruit et permet de maintenir une qualité d'image élevée tout en réduisant la complexité des sondes ultrasonores par un facteur 4, allégeant considérablement les contraintes matérielles.

**Partage des données (*ULMShare*)** Afin de permettre l'extension des méthodes précédemment proposées et de surmonter les limitations d'évaluation identifiées, la thèse présente *ULMShare*, la plus grande base de données publique d'ULM *in vivo* (99 acquisitions sur 61 souris pour 30 To). Ce jeu de données offre enfin à la communauté le socle nécessaire pour développer et comparer objectivement de nouveaux algorithmes, qu'ils soient basés sur l'apprentissage ou conventionnels.

**Conclusion et Perspectives** De la simulation vers la donnée réelle et de l'image au volume, ces travaux posent les fondations méthodologiques pour que l'apprentissage profond contribue à rendre l'ULM plus rapide, plus robuste et plus accessible. En proposant des solutions à des limites inhérentes à l'ULM, ils ouvrent des perspectives concrètes pour une adoption préclinique plus large et, à terme, vers un transfert clinique.

# ABSTRACT

**Context and Motivation** Imaging cerebral microvasculature is fundamental for the early diagnosis and understanding of numerous pathologies, particularly neurodegenerative diseases such as Alzheimer's, where capillary damage often precedes tissue damage. Current clinical modalities (MRI, CT, conventional ultrasound) fail to resolve these microscopic scales at depth. Relying on the localization of microbubbles injected into the bloodstream, Ultrasound Localization Microscopy (ULM) enables the deep observation of the vascular network at scales of tens of microns *in vivo*. However, the widespread adoption of ULM remains hindered by major bottlenecks: prohibitive acquisition times required to capture dilute microbubbles, massive data volumes that are difficult to manage, and critical sensitivity to experimental parameters.

**Hypothesis and Methodology** This thesis relies on the hypothesis that Deep Learning can overcome these physical trade-offs by modeling complex signals of high-concentration microbubbles, where conventional methods fail. However, it identifies a central obstacle: the distribution shift (domain shift) between the simulations used for training and the experimental reality *in vivo*, which limits model generalization. This thesis is composed of four major contributions.

**Structuring the Field** A critical analysis of the literature identified Deep Learning applications with the highest potential impact and formally pinpointed the violation of the i.i.d. (independent and identically distributed) assumption as a major barrier to the *in vivo* application of existing models. This analysis also highlighted the lack of solutions for 3D ULM and the absence of standardized evaluation methods for learning-based approaches.

**Scaling to 3D via Sparsity** To make 3D ULM accessible, this thesis introduces the use of Sparse Tensor Neural Networks. This architecture exploits the spatiotemporal sparsity of microbubbles to reduce memory consumption by two orders of magnitude in 3D compared to standard dense deep learning methods. It not only enables deep learning on volumetric data but also extends to 3D the ability to implicitly resolve overlapping microbubbles at high concentrations, thereby paving the way for rapid volumetric acquisitions.

**Robust *In Vivo* Learning (*Teacher-Student*)** To bridge the gap between simulated training data and *in vivo* application, a Teacher-Student approach was developed to train

models directly on experimental data, without reliance on simulations. This method demonstrated increased robustness to noise and maintains high image quality while reducing ultrasound probe complexity by a factor of 4, significantly alleviating hardware constraints.

**Open access dataset (*ULMShare*)** To enable the extension of the previously proposed paradigm and address the identified evaluation limitations, this thesis presents *ULMShare*, the largest public *in vivo* ULM database to date (99 acquisitions on 61 mice, totaling 30 TB). This dataset finally offers the community the necessary foundation to objectively develop and compare new algorithms, whether learning-based or conventional.

**Conclusion and Perspectives** From simulation to real data and from image to volume, this work lays the methodological foundations for Deep Learning to contribute to making ULM faster, more robust, and more accessible. By proposing solutions to inherent limitations of ULM, it opens concrete perspectives for broader preclinical adoption and, ultimately, towards clinical translation.

# TABLE DES MATIÈRES

DÉDICACE . . . . . iii

REMERCIEMENTS . . . . . iv

RÉSUMÉ . . . . . vi

ABSTRACT . . . . . viii

LISTE DES TABLEAUX . . . . . xiv

LISTE DES FIGURES . . . . . xv

LISTE DES SIGLES ET ABRÉVIATIONS . . . . . xvii

LISTE DES ANNEXES . . . . . xix

CHAPITRE 1 INTRODUCTION . . . . . 1
1.1 Contexte . . . . . 1
1.1.1 Importance et limites de l'imagerie vasculaire cérébrale . . . . . 1
1.1.2 Microscopie de localisation ultrasonore (ULM) . . . . . 2
1.1.3 Limitations actuelles et perspectives . . . . . 2
1.1.4 Rôle et limitations de l'apprentissage profond en ULM . . . . . 3
1.2 Contributions et impact . . . . . 4

CHAPITRE 2 MICROSCOPIE PAR LOCALISATION ULTRASONORE : PRINCIPES, APPLICATIONS ET POSITIONNEMENT DANS L'IMAGERIE BIOMÉDICALE . . . . . 6
2.1 L'imagerie microvasculaire : cruciale mais hors de portée . . . . . 6
2.1.1 Limites des modalités cliniques actuelles . . . . . 6
2.1.2 Limitations des approches précliniques . . . . . 8
2.1.3 Dépasser la diffraction : de la microscopie par localisation optique à l'ULM . . . . . 9
2.2 Principes physiques et implémentation pratique de l'ULM . . . . . 10
2.2.1 Acquisition et formation d'image . . . . . 11
2.2.2 Filtrage du signal (Clutter filtering) . . . . . 12
2.2.3 Localisation et Suivi des microbulles . . . . . 13
2.2.4 Mesure de la qualité de reconstruction . . . . . 14

2.3 Applications et extensions . . . . . 15
2.3.1 Applications précliniques et translation clinique . . . . . 15
2.3.2 ULM dynamique, fonctionnelle et 3D . . . . . 15
2.4 Limites inhérentes et directions d'optimisation . . . . . 17
2.4.1 Compromis entre résolution temporelle, volume de données et concentration . . . . . 17
2.4.2 Limites de la localisation conventionnelle : formalisme et intractabilité 17

CHAPITRE 3 ARTICLE 1 : DEEP LEARNING IN ULTRASOUND LOCALIZATION MICROSCOPY : APPLICATIONS AND PERSPECTIVES . . . . . 20
3.1 Abstract . . . . . 21
3.2 Introduction . . . . . 22
3.3 Generation of labeled datasets . . . . . 24
3.3.1 Formalism . . . . . 25
3.3.2 Prior probability $p(y)$: label generation . . . . . 26
3.3.3 Conditional probability $p(x|y)$: ultrasound simulation . . . . . 29
3.3.4 Learning using in vivo data . . . . . 31
3.4 Deep learning in ULM processing stages . . . . . 32
3.4.1 Aberration correction . . . . . 32
3.4.2 Beamforming . . . . . 34
3.4.3 Clutter filtering and denoising . . . . . 36
3.4.4 Localization . . . . . 37
3.4.5 Tracking . . . . . 38
3.5 A focus on microbubble localization . . . . . 38
3.5.1 Evaluation . . . . . 39
3.5.2 Training formulation . . . . . 43
3.5.3 Architectures . . . . . 45
3.6 Perspectives . . . . . 47
3.6.1 Limitations and future challenges . . . . . 48
3.6.2 Successes and promises . . . . . 50

CHAPITRE 4 ARTICLE 2 : PRUNING SPARSE TENSOR NEURAL NETWORKS ENABLES DEEP LEARNING FOR 3D ULTRASOUND LOCALIZATION MICROSCOPY . . . . . 53
4.1 Abstract . . . . . 54
4.2 Introduction . . . . . 55
4.3 Theory . . . . . 57

4.4 Method . . . 59
4.4.1 Simulations . . . 59
4.4.2 Model training and evaluation . . . 61
4.4.3 Additional studies . . . 63
4.5 Results . . . 66
4.5.1 Processing time, memory reduction and performance comparison in 2D 66
4.5.2 3D feasibility study . . . 68
4.5.3 Additional studies . . . 71
4.6 Discussion . . . 72
4.6.1 Reducing memory usage of existing methods in 2D . . . 72
4.6.2 Scaling to 3D imaging . . . 74
4.6.3 Further reductions of memory and performance trade-off . . . 75
4.6.4 Limitations and perspectives . . . 75
4.7 Conclusion . . . 76

CHAPITRE 5 ARTICLE 3 : TEACHER-STUDENT MODELS FOR ROBUST IN VIVO DEEP-LEARNING IN ULTRASOUND LOCALIZATION MICROSCOPY 78
5.1 Abstract . . . 79
5.2 Introduction . . . 80
5.3 Methods . . . 82
5.3.1 Dataset constitution . . . 82
5.3.2 Deep Learning training . . . 84
5.3.3 Evaluation . . . 84
5.3.4 Finetuning with input perturbation and self-distillation . . . 87
5.4 Results . . . 88
5.4.1 Performance in i.i.d. setting . . . 89
5.4.2 Out-Of-Distribution (OOD) generalization . . . 91
5.4.3 Impact of input perturbation and distillation . . . 93
5.5 Discussion . . . 95
5.5.1 Improved image quality in i.i.d. settings . . . 95
5.5.2 Robustness to limited change in distribution . . . 96
5.5.3 Finetuning for enhanced robustness under severe perturbations . . . 96
5.6 Conclusion . . . 97

CHAPITRE 6 ARTICLE 4 : ULMSHARE : A LARGE-SCALE IN VIVO ULTRASOUND LOCALIZATION MICROSCOPY DATASET FOR MICROVASCULAR IMAGING . . . 98

6.1 Abstract . . . 99
6.2 Background & Summary . . . 100
6.3 Methods . . . 101
6.3.1 ULM acquisition . . . 101
6.3.2 ULM Processing . . . 103
6.4 Data Records . . . 104
6.4.1 Data directory . . . 105
6.4.2 Code directory . . . 105
6.5 Technical Validation . . . 107
6.5.1 Vascular saturation . . . 108
6.5.2 FRC - Spatial coherence . . . 108
6.5.3 Track length - Temporal coherence . . . 109
6.5.4 Qualitative inspection . . . 109
6.6 Data Availability . . . 110
6.7 Code Availability . . . 110

CHAPITRE 7 CONCLUSION . . . 111
7.1 Synthèse des travaux . . . 111
7.1.1 Analyse critique et structuration du domaine . . . 111
7.1.2 Passage à l'échelle pour l'ULM 3D par parcimonie . . . 111
7.1.3 Apprentissage in vivo et robustesse par distillation . . . 112
7.1.4 Base de donnée massive : ULMShare . . . 112
7.2 Limitations et recherches futures . . . 112
7.2.1 Limitations principales . . . 113
7.2.2 Perspectives de recherche . . . 114
7.3 Conclusion générale . . . 115

RÉFÉRENCES . . . 116

ANNEXES . . . 139

# LISTE DES TABLEAUX


Table 3.1 Dataset properties for deep learning approaches in ULM . . . . . 27

Table 4.1 Comparison of the memory usage, inference time and angiogram reconstruction performance . . . . . 66

Table 4.2 Comparison of the memory usage and performance of the different additions to Sparse DeepST-ULM architecture. . . . . 74

Table 5.1 Fourier Ring Correlation and saturation comparison across methods and datasets . . . . . 89

Table 6.1 Comparison of Acquisition Protocols used in ULMShare. . . . . 102

Table 6.2 Detailed composition of the ULMShare dataset grouped by experimental protocol . . . . . 103

Table 6.3 Summary of fixed parameters used in ULM processing for each probe type. . . . . 105

# LISTE DES FIGURES


Figure 2.1 Comparaison résolution/profondeur des modalités d'imagerie . . . . . 7
Figure 2.2 Principe de la super-résolution par localisation . . . . . 10
Figure 2.3 Principe du Beamforming Delay-and-Sum . . . . . 12
Figure 2.4 Impact de la densité de sources sur l'imagerie ULM 2D. . . . . 19
Figure 3.1 Representation of the different types of label generation strategies illustrating sampling from different prior distribution. . . . . 26
Figure 3.2 Overview of ULM processing and simulation pipeline, and the existing deep learning approaches. . . . . 33
Figure 3.3 Results from deep learning approaches targeting stages beyond localization on beamformed data . . . . . 35
Figure 3.4 Illustration of *in vivo* results in 2D from different deep learning localization approach in various organ and animal models and *in silico* results in 3D from [1]. . . . . 40
Figure 3.5 Functional ULM (fULM) comparison between conventional localization and LOCA-ULM from [2] . . . . . 47
Figure 4.1 Simplified representation of localisation and tracking in ULM . . . . . 56
Figure 4.2 Method overview of the proposed Sparse Tensor Neural Network in ULM 58
Figure 4.3 Comparison of performance under increasing concentration between ULM, Dense Deep-stULM . . . . . 67
Figure 4.4 Performance study under varying conditions for the Sparse Deep-stULM model . . . . . 69
Figure 4.5 3D Comparison of performance under increasing concentration between ULM and sparse DeepST-ULM . . . . . 70
Figure 4.7 Failure cases of the *dense-to-sparse* . . . . . 73
Figure 5.1 Representation of the *Teacher-Student-ULM* (TS-ULM) framework . 81
Figure 5.2 Transcranial ULM of a mouse brain processed with TS-ULM . . . . . 89
Figure 5.3 Evolution of TS-ULM loss value on the test sest under varying training perturbation. . . . . 90
Figure 5.4 Transcranial ULM acquisition with synthetic noise addition with TS-ULM . . . . . 92
Figure 5.5 TS-ULM reconstruction of a mouse brain with synthetic reduction of the number of channels . . . . . 92

Figure 5.6 TS-ULM reconstruction of a mouse brain with a probe (GE L8-18iD) not used in the training set . . . . . 93
Figure 5.7 TS-ULM reconstuction with strong synthetic perturbation . . . . . 94
Figure 6.1 Overview of animal-specific information in ULMShare. . . . . 103
Figure 6.2 Directory structure of the raw data stored in the FRDR repository. . 106
Figure 6.3 Directory structure of ULM examples, summary, and helper codes hosted in the GitHub repository. . . . . 107
Figure 6.4 Distribution of ULM validation quantitative metrics across probe types. 108
Figure 6.5 Representative examples of ULM reconstructions illustrating the range of acquisition quality in the ULMShare dataset . . . . . 109

# LISTE DES SIGLES ET ABRÉVIATIONS

| | |
|---|---|
| 2PM | Microscopie Biphotonique (*Two-Photon Microscopy*) |
| AD | Maladie d'Alzheimer (*Alzheimer's Disease*) |
| BUFF | Champ d'Écoulement de Bulles (*BUbble Flow Field*) |
| CAM | Membrane Chorioallantoïdienne (*Chorioallantoic Membrane*) |
| CEUS | Échographie de Contraste (*Contrast-Enhanced Ultrasound*) |
| CNN | Réseau de Neurones Convolutif (*Convolutional Neural Network*) |
| COO | Format de Coordonnées (*Coordinate Format*) |
| CPWC | Composition d'Ondes Planes Cohérentes (*Coherent Plane-Wave Compounding*) |
| CT | Tomodensitométrie (*Computed Tomography*) |
| CVNN | Réseau de Neurones à Valeurs Complexes (*Complex-Valued Neural Network*) |
| DAS | Formation de Voies par Retard et Somme (*Delay-And-Sum*) |
| DETR | Transformateurs de Détection (*DEtection TRansformers*) |
| DL | Apprentissage Profond (*Deep Learning*) |
| DSP | Projection Spécifique au Domaine (*Domain Specific Projection*) |
| DULM | Microscopie de Localisation Ultrasonore Dynamique (*Dynamic Ultrasound Localization Microscopy*) |
| FRC | Corrélation d'Anneau de Fourier (*Fourier Ring Correlation*) |
| fULM | Microscopie de Localisation Ultrasonore Fonctionnelle (*functional Ultrasound Localization Microscopy*) |
| FWHM | Largeur à Mi-Hauteur (*Full Width at Half Maximum*) |
| GAN | Réseaux Génératifs Antagonistes (*Generative Adversarial Networks*) |
| GRU | Unité Récurrente à Porte (*Gated Recurrent Unit*) |
| IQ | En-Phase et Quadrature (*In-Phase / Quadrature*) |
| IRM | Imagerie par Résonance Magnétique (*Magnetic Resonance Imaging*) |
| LOCA-ULM | ULM par Localisation avec Conscience du Contexte (*Localization with Context Awareness ULM*) |
| LSTM | Mémoire à Long et Court Terme (*Long Short-Term Memory*) |
| MB | Microbulle (*Microbubble*) |
| MIP | Projection d'Intensité Maximale (*Maximum Intensity Projection*) |
| MUST | *MATLAB UltraSound Toolbox* |
| NLP | Traitement du Langage Naturel (*Natural Language Processing*) |

| | |
|---|---|
| OOD | Hors Distribution (*Out-Of-Distribution*) |
| PALM | Microscopie de Localisation Photoactivée (*Photoactivated Localization Microscopy*) |
| PSF | Fonction d'Étalement du Point (*Point Spread Function*) |
| RCA | Réseau Ligne-Colonne (*Row-Column Array*) |
| RF | Radiofréquence (*Radiofrequency*) |
| RMSE | Erreur Quadratique Moyenne (*Root Mean Squared Error*) |
| ROI | Région d'Intérêt (*Region Of Interest*) |
| RPCA | Analyse en Composantes Principales Robuste (*Robust Principal Component Analysis*) |
| SMLM | Microscopie de Localisation de Molécule Unique (*Single Molecule Localization Microscopy*) |
| SNR | Rapport Signal sur Bruit (*Signal-to-Noise Ratio*) |
| SOAM | Métrique de Somme des Angles (*Sum Of Angles Metric*) |
| STORM | Microscopie de Reconstruction Optique Stochastique (*Stochastic Optical Reconstruction Microscopy*) |
| SVD | Décomposition en Valeurs Singulières (*Singular Value Decomposition*) |
| TAL | Suivi et Localisation (*Tracking And Localization*) |
| TGC | Compensation de Gain Temporel (*Time-Gain Compensation*) |
| TS-ULM | *Teacher-Student ULM* |
| ULM | Microscopie de Localisation Ultrasonore (*Ultrasound Localization Microscopy*) |

# LISTE DES ANNEXES

Annexe A LISTE DES CONTRIBUTIONS SCIENTIFIQUES . . . . . 139

# CHAPITRE 1 INTRODUCTION

## 1.1 Contexte

### 1.1.1 Importance et limites de l'imagerie vasculaire cérébrale

L'imagerie vasculaire est un outil essentiel pour comprendre la physiologie des organes et détecter leurs dysfonctionnements. De nombreuses techniques sont aujourd'hui disponibles en clinique pour imager l'anatomie vasculaire et caractériser les paramètres hémodynamiques, afin d'appuyer le diagnostic ou de suivre l'évolution des patients. Toutefois, ces techniques se limitent généralement aux vaisseaux dont le diamètre dépasse une centaine de microns. Pourtant, la microcirculation joue un rôle central dans de nombreux processus biologiques, et ses altérations surviennent souvent avant celles affectant les vaisseaux de plus grand calibre [3]. C'est notamment le cas dans l'angiogenèse tumorale [4] ou dans certaines maladies neurodégénératives, telles que la maladie d'Alzheimer ou diverses formes de démence [5]. Dans le cerveau, cet enjeu est encore plus critique : l'activité neuronale dépend étroitement de l'apport en oxygène et en nutriments fourni par le réseau capillaire [6], rendant le tissu nerveux particulièrement sensible aux perturbations microvasculaires [7].

Les techniques optiques, bien qu'offrant une résolution submicronique, sont limitées en pénétration par l'absorption tissulaire et l'opacité du crâne. À l'inverse, l'IRM, la CT ou l'imagerie nucléaire permettent d'explorer l'ensemble du cerveau, mais leur résolution reste typiquement millimétrique [8]. Certaines approches, comme l'imagerie optique avec craniotomie [9] ou les dispositifs endovasculaires [10], permettent d'accéder à des échelles plus fines, mais leur caractère invasif les rend inadaptées à un usage clinique de routine. Enfin, en alliant faible coût, portabilité et sûreté d'utilisation, l'échographie conventionnelle présente plusieurs avantages, mais elle reste contrainte par la diffraction : augmenter la fréquence améliore la résolution au prix d'une réduction de la profondeur de pénétration [11].

À ce jour, aucune modalité clinique ne permet donc d'imager *in vivo* la microcirculation cérébrale en profondeur de manière non invasive [8]. Même en préclinique, les techniques établies imposent un compromis insatisfaisant et n'offrent pas d'accès direct à la microvascularisation capillaire *in vivo* sans recours à des approches invasives.

### 1.1.2 Microscopie de localisation ultrasonore (ULM)

Facilitée par le développement de l'imagerie ultrarapide [12], l'ULM a été conçue pour dépasser la limite de diffraction en échographie en exploitant de faibles concentrations de microbulles, déjà utilisées en clinique comme agents de contraste [13, 14]. À ces concentrations, les microbulles apparaissent comme des sources ponctuelles isolées dont la position peut être localisée avec une précision largement supérieure à la résolution acoustique conventionnelle [15, 16]. En accumulant des milliers de ces localisations, il devient possible de reconstruire une cartographie microvasculaire à l'échelle micrométrique. Cette approche atteint ainsi une résolution d'un ordre de grandeur supérieure à celle de l'échographie classique, tout en conservant la profondeur de pénétration et les autres avantages propres à l'imagerie ultrasonore [11, 17].

En préclinique, l'ULM a déjà démontré sa pertinence pour l'étude de pathologies majeures, permettant par exemple de discriminer les types d'accidents vasculaires cérébraux (ischémique vs hémorragique) [18], de quantifier les déficits de perfusion dans les modèles de la maladie d'Alzheimer [19, 20] ou encore de caractériser l'angiogenèse tumorale [21] et la filtration glomérulaire [22, 23]. Par ailleurs, en synchronisant les trajectoires de bulles avec un signal externe (comme le cycle cardiaque ou une stimulation fonctionnelle), l'ULM dynamique donne accès à des paramètres physiologiques critiques tels que la vitesse du flux sanguin, la pulsatilité vasculaire [24, 25] ou l'activité cérébrale via le couplage neurovasculaire [26]. Le potentiel translationnel de l'ULM est d'ailleurs confirmé par des premières applications cliniques prometteuses, notamment pour la caractérisation de lésions mammaires [27] ou hépatiques [28], l'évaluation de greffons rénaux [29], l'imagerie myocardique [30] et l'imagerie cérébrale chez l'adulte [31], positionnant l'ULM comme une future modalité de référence pour le diagnostic microvasculaire [8].

### 1.1.3 Limitations actuelles et perspectives

Malgré son potentiel, l'ULM conventionnelle est intrinsèquement limitée à l'utilisation de faibles concentrations de microbulles pour s'affranchir de la limite de diffraction. En effet, les méthodes de localisation reposent sur l'hypothèse de microbulles isolées afin d'estimer leur position avec précision. Lorsque les signaux de plusieurs microbulles se superposent spatialement, la modélisation de leurs positions respectives devient extrêmement complexe à résoudre avec les algorithmes classiques [11]. D'une part, la réduction de la concentration ne résout que partiellement ce problème : la présence simultanée de plusieurs microbulles reste inévitable dans les zones hautement vascularisées, ce qui altère la précision de localisation et toutes les mesures subséquentes. D'autre part, cette dilution impose l'accumulation d'images sur des

durées prolongées, parfois jusqu'à plusieurs minutes, pour couvrir l'intégralité du réseau vasculaire [32]. Ce temps d'acquisition long expose l'ULM aux artefacts de mouvement [33,34] et génère des volumes de données considérables, un problème qui s'aggrave exponentiellement avec le passage à l'imagerie 3D, dont la complexité limite encore l'usage généralisé [35–37].

### 1.1.4 Rôle et limitations de l'apprentissage profond en ULM

Face à cette complexité de modélisation et à l'abondance de données, l'apprentissage profond (*Deep Learning*) s'impose comme une approche méthodologique particulièrement adaptée. En effet, ces algorithmes ont démontré leur capacité à modéliser des distributions complexes avec une grande précision en tirant parti de vastes ensembles de données [38]. Portée par l'augmentation conjointe de la puissance de calcul, cette approche a révolutionné l'analyse d'images [38], des premiers succès en classification [39, 40] jusqu'aux récents modèles de fondation capables de généraliser à des nouvelles données médicales [41]. En échographie, elle a déjà permis des avancées notables en formation d'image ou en suppression du signal tissulaire [42, 43].

Dans le contexte spécifique de l'ULM, les réseaux de neurones tirent parti de leur aptitude à modéliser des signaux non-linéaires complexes pour résoudre les superpositions de microbulles à forte concentration, surpassant ainsi les méthodes analytiques conventionnelles [2, 44–46]. Cette capacité ouvre la voie à une réduction significative des temps d'acquisition [45] et à une meilleure robustesse des reconstructions face aux conditions expérimentales difficiles [47, 48].

Toutefois, ces méthodes se heurtent à l'absence de vérité terrain *in vivo* pour l'entraînement supervisé. Si les approches existantes contournent ce problème via des jeux de données simulés, cela engendre un écart de distribution (*domain shift*) vis-à-vis de la réalité expérimentale qui limite souvent la généralisation des modèles *in vivo* [2,49]. Par ailleurs, l'extension de ces architectures à l'imagerie 3D demeure contrainte par la lourde complexité algorithmique de la super-résolution volumique nécessaire en ULM [1].

### Objectifs de la thèse

L'objectif général de cette thèse est d'explorer et de démontrer le potentiel de l'apprentissage profond pour surmonter les limitations actuelles de la microscopie de localisation ultrasonore (ULM), en particulier pour réduire le temps d'acquisition et améliorer la reproductibilité ainsi que la robustesse expérimentale.

Les objectifs spécifiques de ce travail sont de :

1. **Analyser de manière critique et structurée** les approches existantes d'appren-

tissage profond appliquées à l'ULM, afin d'en dégager les fondements, les divergences et les limites, et d'identifier les perspectives d'amélioration de l'ULM.

2. **Proposer une approche d'apprentissage profond pour l'ULM 3D**, permettant d'exploiter des concentrations élevées de microbulles tout en maintenant la précision de localisation et une complexité mémoire maîtrisée, rendant possible l'ajout d'une dimension spatiale.
3. **Développer une méthode de localisation robuste *in vivo***, capable d'apprendre directement à partir de données expérimentales et de généraliser à des conditions d'acquisition variées.
4. **Constituer une base de données publique d'acquisitions *in vivo*** couvrant différentes configurations expérimentales et niveaux de qualité, afin de promouvoir la reproductibilité, la comparaison objective des méthodes et le développement communautaire de nouveaux modèles d'apprentissage [50].

## 1.2 Contributions et impact

Cette thèse apporte quatre contributions majeures qui lèvent plusieurs limitations de l'apprentissage profond appliqué à l'ULM : l'absence de cadre comparant les approches existantes, la portée encore limitée des modèles *in vivo* ou en 3D, ainsi que le manque de données publiques *in vivo*. Elle pose également les fondations conceptuelles et pratiques nécessaires au développement de méthodes d'apprentissage profond capables d'étendre les capacités de l'ULM au-delà de ses limites traditionnelles, qu'il s'agisse de la durée d'acquisition, des contraintes matérielles ou de la sensibilité aux conditions d'imagerie.

1. **Une analyse critique des applications existantes et de leurs limitations structurant l'apprentissage profond en ULM**. Le chapitre 3 offre la première revue exhaustive des approches d'apprentissage profond pour l'ULM, en structurant leurs principes communs, leurs limites, leurs divergences et le potentiel qu'elles démontrent. Il clarifie les défis encore ouverts et établit le cadre méthodologique qui orientera les développements introduits dans cette thèse.
2. **La première démonstration d'ULM 3D basée sur l'apprentissage profond** Le chapitre 4 démontre qu'exploiter la parcimonie spatio-temporelle des trajectoires de microbulles permet d'alléger le coût computationnel de l'apprentissage profond en ULM et permet son application en 3D. Il propose la première méthode d'apprentissage profond pour l'ULM volumique et montre que les réductions de temps d'acquisition obtenues en 2D sont transposables en 3D [1].

3. **Une méthode de localisation *in vivo* robuste et généralisable** Le chapitre 5 introduit une approche de localisation directement entraînée sur des données *in vivo*. En s'affranchissant des simulations pour la génération des données d'apprentissage, cette méthode établit un cadre d'entraînement dans lequel les modèles généralisent de manière robuste *in vivo*, y compris en cas de dégradation des conditions d'imagerie. Par ailleurs, cette approche est également étendue afin de réduire les besoins matériels de l'ULM 2D, ouvrant la voie à des améliorations similaires en 3D susceptibles de diminuer drastiquement le coût de l'ULM volumique.
4. **Une base de données *in vivo* publique de grande ampleur** Le chapitre 6 présente ULMShare, une collection d'acquisitions transcrâniennes *in vivo* de cerveaux de souris. Cette ressource, constituée de 99 acquisitions provenant de 61 animaux, est la première et la seule de cette ampleur, dépassant d'un à deux ordres de grandeur les ensembles de données existants. En mettant à disposition une quantité de données *in vivo* sans précédent, elle ouvre la voie au développement futur de méthodes d'apprentissage profond, notamment celles inspirées du chapitre précédent. Au-delà de l'apprentissage automatique, cette base de données constitue également un support unique pour le développement et l'évaluation quantitative de méthodes conventionnelles, grâce à la diversité et au volume des acquisitions proposées.

### Impact global

En réunissant un cadre structurant, de nouvelles méthodes d'apprentissage pour l'ULM *in vivo* et en 3D, ainsi qu'une base de données *in vivo* de référence, cette thèse démontre que l'apprentissage profond peut dépasser plusieurs limitations fondamentales de l'ULM conventionnel et pose les fondations nécessaires au transfert des avancées récentes de l'apprentissage profond vers l'ULM. Les contributions proposées renforcent la robustesse, la rapidité et la reproductibilité de l'ULM, tout en ouvrant la voie à des applications volumétriques plus accessibles et à une meilleure standardisation du domaine. Ensemble, elles établissent les conditions nécessaires au développement d'une ULM plus viable et plus fiable menant à une utilisation préclinique plus large et, à terme, clinique.

# CHAPITRE 2 MICROSCOPIE PAR LOCALISATION ULTRASONORE : PRINCIPES, APPLICATIONS ET POSITIONNEMENT DANS L'IMAGERIE BIOMÉDICALE

## 2.1 L'imagerie microvasculaire : cruciale mais hors de portée

Pour assurer l'apport en oxygène et en nutriments aux tissus biologiques, le sang transite depuis les artères et artérioles vers le lit capillaire, où a lieu la plupart des échanges métaboliques, avant de repartir vers le cœur par les veinules et les veines [51]. Si l'imagerie des vaisseaux macroscopiques (diamètre $> 100$ $\mu$m) est aujourd'hui routinière, l'observation non-invasive de cette microcirculation (vaisseaux $< 100$ $\mu$m) reste un défi technologique majeur (Voir Figure 2.1). Pourtant, l'imagerie de cette microcirculation est cruciale, tant pour le diagnostic précoce et le suivi thérapeutique en clinique que pour la compréhension des mécanismes fondamentaux en recherche préclinique. Cette limitation actuelle freine notamment l'étude de nombreuses pathologies où les altérations microvasculaires précèdent souvent les dommages tissulaires macroscopiques, telles que les cancers [4], les maladies neurodégénératives associées aux micro-infarctus cérébraux [5, 52, 53] ou certaines maladies cardiaques liées à la dysfonction microvasculaire coronaire [54].

### 2.1.1 Limites des modalités cliniques actuelles

En clinique, les modalités d'imagerie de référence, telles que l'Imagerie par Résonance Magnétique (IRM), la Tomodensitométrie (CT) ou l'échographie Doppler, offrent une couverture anatomique complète et une profondeur de pénétration suffisante pour imager l'organe entier. Cependant, elles restent intrinsèquement limitées à des résolutions bien supérieures à la taille des vaisseaux impliqués dans la microcirculation.

En IRM, la résolution spatiale est intrinsèquement liée au rapport signal-sur-bruit (SNR) [55]. Pour améliorer ce compromis, deux leviers principaux existent. L'augmentation du temps d'acquisition permet théoriquement d'échantillonner de plus hautes fréquences spatiales tout en maintenant un SNR exploitable. Cependant, cette approche est rapidement limitée par des contraintes pratiques ainsi que par la détérioration de l'image due aux mouvements physiologiques (respiration, pulsatilité). Alternativement, l'augmentation de l'intensité du champ magnétique, et notamment le passage à l'ultra-haut champ (7 Tesla), permet théoriquement d'atteindre des résolutions de l'ordre de la centaine de microns. Toutefois, cette technologie se heurte à des obstacles majeurs : coût élevé, disponibilité restreinte et artefacts d'inhomo-

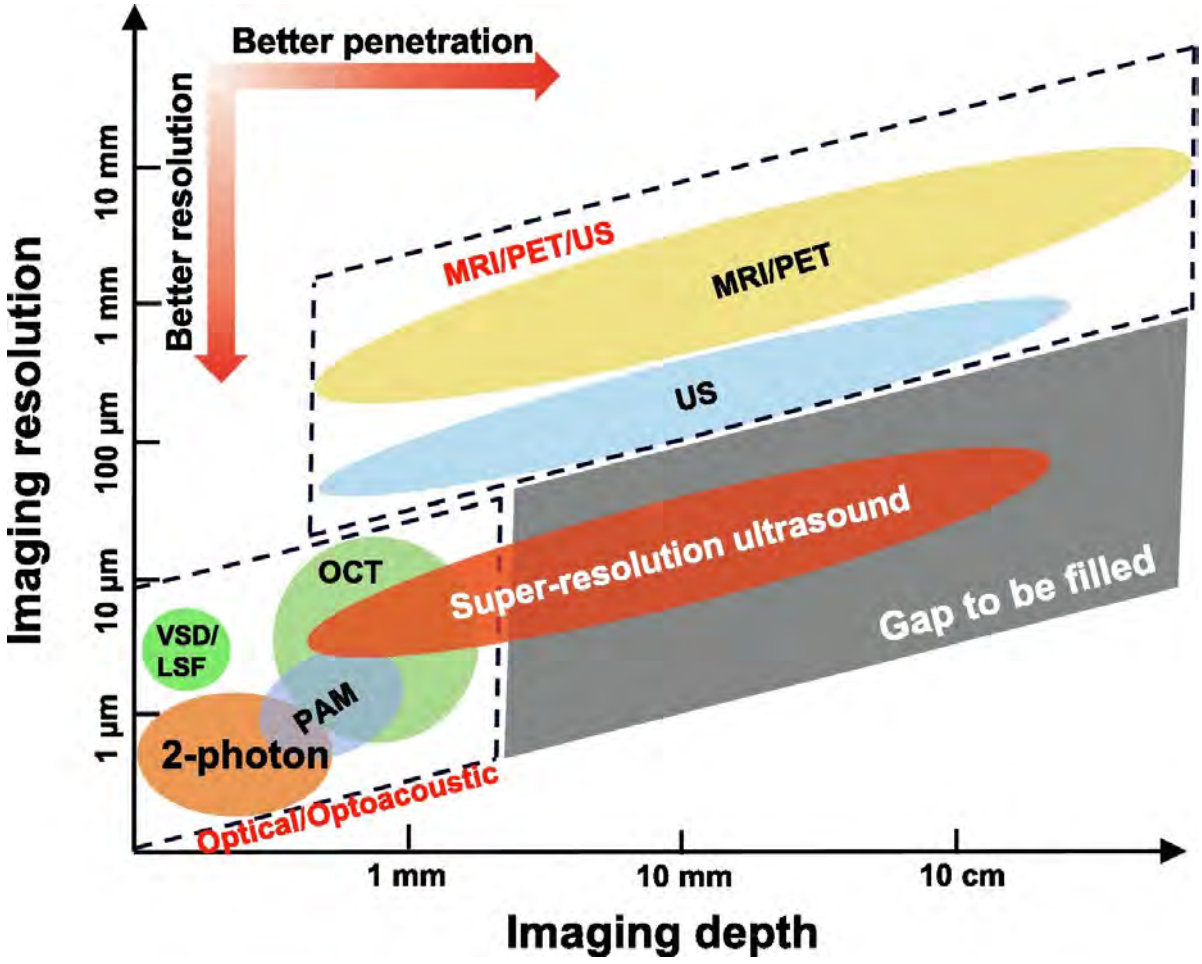



FIGURE 2.1 **Comparaison de la résolution spatiale et de la profondeur de pénétration des différentes modalités d'imagerie biomédicale.** Ce diagramme illustre le rôle à jouer pour l'ULM en offrant une résolution microscopique à des profondeurs cliniques, là où les autres modalités sont soit limitées en résolution (IRM, CT, US), soit en profondeur (Optique). *Adapté de Song et al.* [8].

généité qui freinent encore son adoption en routine clinique [56]

De même, la résolution du scanner à rayons X (CT) est régie par un compromis strict entre la résolution spatiale et la dose délivrée [57]. L'amélioration de la géométrie d'acquisition (finesse des collimateurs, réduction de la taille du foyer) entraîne mécaniquement une augmentation du bruit quantique (bruit de Poisson) en réduisant le flux de photons par voxel. Pour compenser cette perte de signal et maintenir une image exploitable, il est nécessaire d'agir sur le second levier : l'intensité du rayonnement. Cette augmentation de la dose devient rapidement inacceptable pour la sécurité du patient, limitant drastiquement le suivi longitudinal ou l'acquisition continue nécessaire à l'étude dynamique de la microcirculation [58].

Enfin, l'échographie conventionnelle est fondamentalement régie par la limite de diffraction acoustique et le compromis inhérent entre résolution spatiale et pénétration. En effet, si l'augmentation de la fréquence centrale permet d'améliorer la résolution, elle se fait au coût d'une atténuation tissulaire accrue et donc d'une profondeur d'exploration réduite. Pour atteindre une résolution microscopique, l'utilisation de fréquences extrêmement élevées ($> 50$ MHz) serait requise. L'atténuation des ondes ultrasonores dans les tissus biologiques augmentant quasi-linéairement avec la fréquence, la profondeur de pénétration se trouverait alors limitée à quelques millimètres superficiels. Cette barrière physique rend l'imagerie microscopique non invasive impossible pour les organes profonds chez l'homme avec les méthodes conventionnelles [11].

En conséquence, il n'existe à ce jour aucun outil clinique capable d'observer la microcirculation en profondeur.

### 2.1.2 Limitations des approches précliniques

Pour contourner les limites de résolution clinique, la recherche fondamentale s'appuie largement sur des techniques de microscopie optique avancées, telles que la microscopie biphotonique [59] ou la microscopie à feuillets de lumière (*light-sheet*) [60]. Ces méthodes atteignent une résolution submicronique et permettent de visualiser l'architecture cellulaire et le réseau vasculaire à l'échelle capillaire. Toutefois, la diffusion des photons dans les tissus biologiques restreint la profondeur d'imagerie effective à quelques centaines de microns, confinant l'observation aux tissus superficiels [61]. L'accès aux organes profonds impose alors des procédures chirurgicales lourdes. Cette contrainte est particulièrement critique pour le cerveau, où l'implantation de fenêtres optiques, de prismes, ou le sacrifice pour clarification tissulaire (*clearing*) sont nécessaires pour accéder aux structures sous-corticales [9]. Ces méthodes invasives ou terminales compromettent l'intégrité physiologique et interdisent l'étude non invasive de la microcirculation dans l'ensemble du volume d'un organe intact.

Par ailleurs, le micro-scanner à rayons X (micro-CT) offre une résolution spatiale isotrope élevée à l'échelle de l'animal entier. Cependant, cette modalité souffre d'un faible contraste intrinsèque des tissus mous, nécessitant l'injection de forts volumes d'agents de contraste denses. De plus, le rapport signal-sur-bruit étant lié à la dose, atteindre une résolution micrométrique impose d'augmenter considérablement le flux de rayons X. Cela expose l'animal à des doses cumulées souvent létales ou incompatibles avec le maintien de l'homéostasie lors d'études longitudinales [62]. Ainsi, bien que le micro-CT soit l'outil anatomique de référence *ex vivo*, il ne permet pas le suivi dynamique et fonctionnel de la microcirculation fine *in vivo* [8].

En résumé, il n'existe à ce jour aucune modalité d'imagerie capable de concilier simultanément une résolution microscopique, une profondeur de pénétration à l'échelle de l'organe et un caractère non invasif adapté au suivi longitudinal.

### 2.1.3 Dépasser la diffraction : de la microscopie par localisation optique à l'ULM

Des avancées importantes en microscopie optique, récompensées par un prix Nobel de Chimie en 2014, permettent de s'affranchir de la limite de diffraction à condition d'imager des diffuseurs isolés (voir Figure 2.2). Pour ce faire, les méthodes comme le PALM [63] ou le STORM [64] exploitent l'activation stochastique des sources lumineuses pour garantir l'isolement spatial des signaux reçus. Grâce à un clignotement aléatoire des émetteurs, seule une infime fraction d'entre eux est active simultanément et chaque point lumineux apparaît seul dans son voisinage, sans chevauchement avec ses voisins. L'image finale super-résolue est alors reconstruite point par point, en accumulant ces localisations sur une longue séquence d'images.

La Microscopie de Localisation Ultrasonore (ULM) constitue la transposition acoustique directe de ce concept [13, 14]. Par analogie, les agents de contraste ultrasonores (microbulles) jouent ici le rôle des sources lumineuses. Ces microbulles, déjà utilisées couramment en clinique [65,66], ont une taille comparable à celle d'un globule rouge (1–3 $\mu$m) et peuvent donc circuler librement dans le réseau vasculaire.

Cependant, à la différence des méthodes optiques, l'ULM exploite le flux sanguin et l'imagerie ultrarapide pour séparer les sources acoustiques : la haute cadence d'acquisition permet de figer le mouvement des microbulles et de les distinguer temporellement, à condition qu'elles aient été injectées en concentration suffisamment faible. Dans ce cas, chaque microbulle peut être localisée avec une précision de l'ordre du micron, soit plus de dix fois inférieure à la résolution native du système [13, 14, 16] (selon le même principe que dans la Figure 2.2). Grâce à la cohérence temporelle du signal des microbulles, elles peuvent être suivies sur

plusieurs images successives. L'accumulation de ces trajectoires au cours du temps permet de cartographier la microvascularisation en profondeur, brisant ainsi le compromis historique entre résolution et pénétration [17].

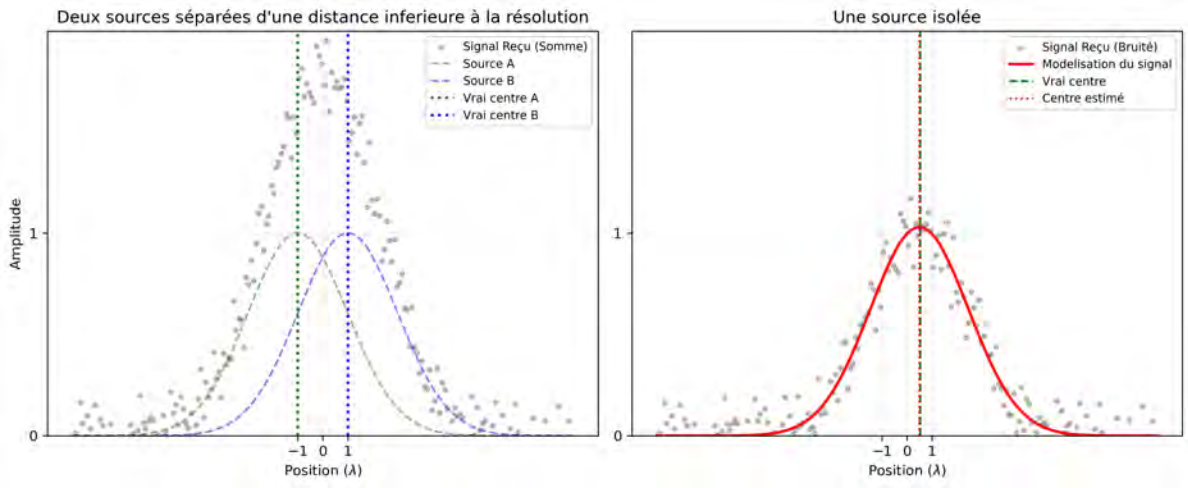



FIGURE 2.2 **Illustration du dépassement de la limite de diffraction par la localisation. (Gauche)** Limite de diffraction classique (Critère de Rayleigh) : lorsque deux sources ponctuelles sont séparées par une distance comparable à la largeur de la réponse impulsionnelle (PSF), leurs signaux se chevauchent, rendant les sources indiscernables. Le signal mesuré (points gris) ne permet pas de séparer les contributions individuelles. **(Droite)** Principe de la localisation (ULM) : lorsqu'une source est isolée spatialement, il est possible d'ajuster un modèle mathématique (courbe rouge) sur le signal bruité pour estimer la position de son centre avec une précision (erreur de localisation) très inférieure à la largeur de la tache de diffraction.

## 2.2 Principes physiques et implémentation pratique de l'ULM

La mise en œuvre de l'ULM repose sur un traitement complexe, composé de plusieurs étapes et nécessitant l'acquisition de milliers d'images sur des durées pouvant atteindre plusieurs minutes [32]. Ce processus peut être divisé en quatre étapes principales : l'acquisition et la formation de l'image, le filtrage du signal, la localisation des microbulles et leur suivi. Cependant, cet ordre n'est pas strict et certaines étapes peuvent être inversées (comme le filtrage et la formation d'image) ou réalisées conjointement (localisation et suivi des microbulles). Cette section décrit les principes fondamentaux régissant ces différents blocs dans ce que l'on peut considérer comme l'approche standard en ULM [49, 67].

### 2.2.1 Acquisition et formation d'image

Rendue possible par le développement des échographes programmables, l'imagerie ultrarapide par ondes planes (*Plane Wave Imaging*) permet d'insonifier l'ensemble du milieu en une seule transmission, par opposition au balayage focalisé ligne par ligne. Ce changement de paradigme permet de faire passer la cadence d'acquisition de 50 Hz à plus de 20 000 Hz [12], ouvrant la voie à une imagerie Doppler haute sensibilité [68].

Le signal radio-fréquence (RF) enregistré par chaque élément de la sonde est de nature temporelle. L'opération de formation de voies, ou *beamforming*, vise donc à convertir cette information temporelle en une cartographie spatiale. Bien que des approches adaptatives plus complexes existent, l'algorithme *Delay-and-Sum* (DAS) demeure la référence en ULM pour sa robustesse, sa simplicité et sa linéarité. Pour chaque pixel de l'image finale, cet algorithme somme les contributions des signaux bruts après avoir compensé le délai correspondant au temps de vol aller-retour théorique. C'est cette opération géométrique qui assure la focalisation dynamique du signal en réception en tout point de l'image (voir Figure 2.3).

Le contenu fréquentiel du signal brut RF est confiné dans la bande passante du transducteur, autour d'une fréquence porteuse $f_0$. En exploitant cette propriété, il est possible de réduire l'échantillonnage sans perte d'information. En effet, le signal est classiquement démodulé en composantes En-Phase et Quadrature (IQ), une opération qui recentre le spectre autour de 0 Hz, ce qui permet d'appliquer un filtrage passe-bas. La fréquence d'échantillonnage minimale pour respecter le critère de Nyquist-Shannon ne dépend alors plus de la porteuse, mais uniquement de la largeur de bande du signal. En pratique, cela permet souvent de diviser le volume de données par un facteur 2 à 4, un gain critique pour la gestion des séquences longues en ULM.

Cette décimation est souvent opérée directement par le système d'imagerie, qui délivre des données sous forme complexe (IQ). Le DAS peut alors s'appliquer sur ces signaux de manière analogue aux données RF. S'il est courant de se contenter de l'amplitude du signal résultant, qui correspond à l'image échographique standard (mode B), pour simplifier l'analyse structurelle, l'information de phase demeure indispensable aux méthodes avancées. Elle permet en effet l'estimation de la vitesse du flux sanguin (Doppler) [68] ou la correction des aberrations induites par le milieu [69].

Toutefois, une onde plane unique souffre d'un contraste et d'une résolution dégradés. Pour pallier ce défaut, la stratégie standard consiste à utiliser le *Coherent Plane-Wave Compounding* (CPWC) : plusieurs ondes planes inclinées sont transmises successivement (typiquement de -5 à +5 degrés) et leurs signaux rétro-diffusés sont sommés de manière cohérente [70]. Pour

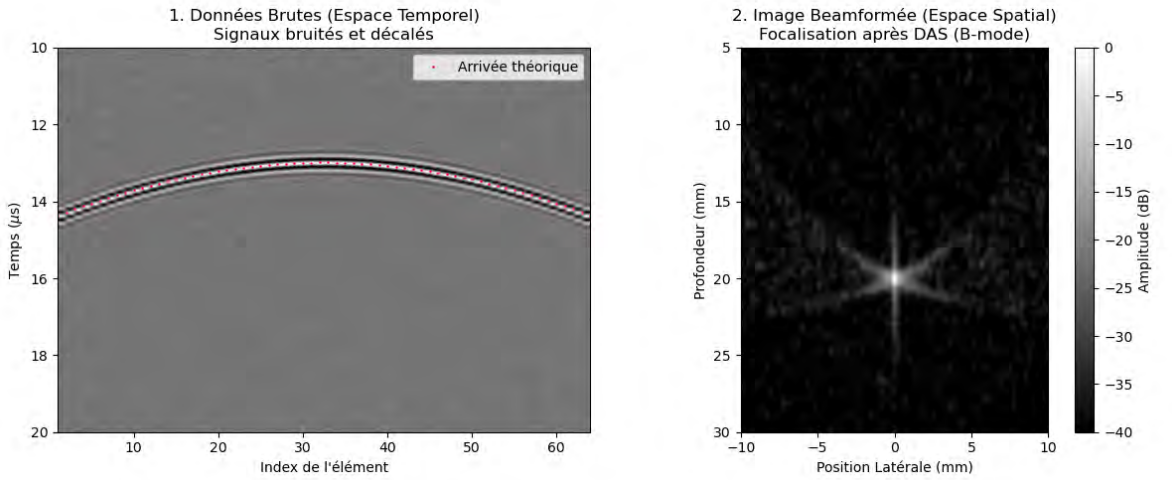



FIGURE 2.3 **Illustration du principe de *Beamforming* Delay-and-Sum (DAS). (1) Espace Temporel (Données Brutes) :** La sonde enregistre les échos reçus par chaque élément en fonction du temps. Une cible ponctuelle unique génère un front d'onde qui arrive à des instants différents sur chaque élément, traçant une signature hyperbolique caractéristique (pointillés rouges). **(2) Espace Spatial (Image Beamformée) :** L'algorithme DAS est appliqué sur une grille 2D de pixels. Pour chaque point, il compense les retards géométriques et somme les contributions. Cette opération focalise l'énergie de l'hyperbole en un point unique à la position réelle de la cible (20 mm).

augmenter davantage la spécificité de détection, il est également possible d'exploiter la réponse non-linéaire des microbulles afin de mieux séparer leur signal des tissus avoisinants. Bien que très efficaces pour rejeter le tissu, ces méthodes requièrent plusieurs transmissions par image, accentuant le compromis sur la résolution temporelle [71]. En conséquence, leur usage reste souvent limité à des cas particuliers d'imagerie ULM où la sensibilité prime sur la cadence.

### 2.2.2 Filtrage du signal (Clutter filtering)

A ce stade, le signal ultrasonore contient la superposition de la réponse des tissus (*clutter*) et celle, beaucoup plus faible, des microbulles. La technique la plus courante pour les séparer est la Décomposition en Valeurs Singulières (SVD) spatio-temporelle [72]. Pour cela, la séquence d'images est réorganisée en une matrice spatio-temporelle (matrice de Casorati) où chaque ligne représente une image vectorisée et chaque colonne la variation temporelle de la valeur d'un pixel. Ensuite, la SVD permet d'obtenir une décomposition en somme pondérée de matrices séparables en un vecteur temporel et un vecteur spatial. Les valeurs singulières sont ordonnées par cohérence spatio-temporelle et il est alors possible de rejeter les premières

composantes, correspondant aux tissus énergétiques et cohérents, pour ne conserver que le signal dynamique des agents de contraste.

Cependant, cette approche a aussi tendance à supprimer le signal des microbulles les plus lentes, seul indicateur de la présence des vaisseaux les plus fins. Pour observer ce signal, d'autres stratégies ont été développées, en parallèle des méthodes d'imagerie non-linéaire déjà mentionnées. Le **Sensing ULM** emploie une stratégie de filtrage multi-étape adaptative pour isoler spécifiquement les populations de microbulles lentes, indispensable à l'imagerie des glomérules [22] mais reste limité. Plus récemment, l'approche **SCaRe** introduit l'utilisation de *Long Ensemble SVD* : en augmentant considérablement la fenêtre temporelle d'analyse, cette méthode améliore la séparation entre les tissus et les microbulles les plus lentes, permettant la détection des capillaires [73].

Enfin, l'efficacité de ce filtrage est conditionnée par la sélection critique du seuil de rejet des première valeurs. Si des méthodes automatiques ont été proposées pour déterminer ce rang optimal, ce paramètre reste en pratique le plus souvent ajusté de manière empirique, introduisant une variabilité inter-opérateur ainsi qu'un risque de biais dans la quantification vasculaire.

### 2.2.3 Localisation et Suivi des microbulles

Après filtrage du tissu, les microbulles sont détectées en tant que maxima locaux puis leur position exacte est estimée à l'aide d'algorithmes de localisation. Les méthodes les plus précises et les plus couramment utilisées sont l'estimation gaussienne (*Gaussian Fitting*) ou la symétrie radiale (*Radial Symmetry*).

Les positions obtenues sont ensuite temporellement associées pour reconstruire des trajectoires (*tracking*). La confiance dans les détections successives d'une même microbulle permet alors de ne conserver que les trajectoires les plus longues. L'algorithme hongrois est communément utilisé pour résoudre le problème d'appariement. En minimisant la somme des distances d'association, cette méthode favorise par construction les solutions de moindre déplacement, ce qui peut induire des erreurs d'appariement dans les zones de flux rapides ou complexes.

Pour contourner ce problème, certaines approches visent à isoler les trajectoires avant l'étape de localisation [74]. Des méthodes d'apprentissage profond ont aussi visé à s'affranchir d'une étape de localisation explicite et à estimer directement la densité ou la vitesse à partir des données spatio-temporelles [45, 75]. Toutefois, les méthodes basées sur la localisation individuelle restent limitées par la densité de microbulles : à forte concentration, la superposition

des signaux altère inévitablement la précision de la localisation et rend l'appariement plus difficile.

### 2.2.4 Mesure de la qualité de reconstruction

Pour obtenir une image vasculaire, il reste à projeter l'ensemble des trajectoires accumulées sur une grille de reconstruction fine. Pour garantir la continuité visuelle des vaisseaux, les trajectoires sont interpolées, spatialement ou temporellement, avant d'être accumulées. En plus de la densité de microbulles, il est alors possible de générer des cartes de vitesse en moyennant la vélocité des trajectoires traversant chaque pixel. Qualitativement, la reconstruction et la confiance de la mesure réside alors dans le grand nombre d'accumulation et leur cohérence spatiale, la longueur des trajectoires obtenus ainsi que l'uniformité spatiale des détections.

Toutefois, l'évaluation quantitative de ces reconstructions est intrinsèquement limitée par l'absence de vérité terrain *in vivo*. En effet, les modalités d'imagerie alternatives manquent de résolution ou de profondeur pour servir de référence fiable. Quant à la validation croisée avec des techniques *ex vivo*, elle est complexifiée par les difficultés de recalage tridimensionnel et les déformations tissulaires inhérentes à la fixation ou aux manipulations *post-mortem*. Par ailleurs, si les simulations offrent un cadre contrôlé pour comparer théoriquement les algorithmes, elles ne reproduisent que partiellement la complexité acoustique et physiologique réelle, limitant de fait la portée de leurs conclusions.

En l'absence de référence absolue, la qualité des images *in vivo* est estimée principalement à l'aide de trois métriques complémentaires. La cohérence spatiale est mesurée par la *Fourier Ring Correlation* (FRC), qui estime la plus haute fréquence spatiale pour laquelle deux sous-images indépendantes conservent une corrélation supérieure au bruit [76]. L'uniformité et la complétude sont évaluées par la courbe de saturation, qui quantifie le ratio de pixels détectés sur le nombre total de pixels en fonction du temps d'accumulation [32]. Enfin, la durée moyenne des trajectoires est fréquemment rapportée pour quantifier la robustesse du suivi et la capacité à maintenir l'identité des microbulles sur de longues durées.

Toutefois, il est important de noter que ces métriques peuvent être artificiellement optimisées par des reconstructions dégénérées, telles que des artefacts de grille ou un fond constant. Par conséquent, elles ne sauraient se substituer à un examen visuel critique de la fidélité de l'image reconstruite.

## 2.3 Applications et extensions

L'ULM s'est établie comme un outil d'investigation biomédicale, au-delà de la preuve de concept. Son développement s'articule autour de deux axes : la translation clinique vers l'exploration physiopathologique et l'innovation méthodologique pour l'imagerie volumique et dynamique. Cette section synthétise l'état de l'art de ces applications et les avancées technologiques étendant les capacités de l'ULM.

### 2.3.1 Applications précliniques et translation clinique

Dans le cerveau, la difficulté majeure réside dans les aberrations induites par la différence de vitesse de propagation du son dans le crâne et celle dans les tissus mous. Au stade préclinique, l'ULM a démontré un fort potentiel d'application. Dans les modèles d'Alzheimer chez la souris, des altérations de la tortuosité et un découplage hémodynamique ont été observés avant l'apparition des plaques amyloïdes [19]. De même, l'imagerie post-AVC chez le rat a démontré sa capacité à discriminer l'ischémie de l'hémorragie et à cartographier la réorganisation vasculaire lors de la récupération [18]. Les applications cliniques de l'ULM chez l'humain ont pu avoir lieu soit en évitant les aberrations soit en les corrigeant. Chez les nouveau-nés, l'imagerie au travers de la fontanelle permet de limiter les aberrations et l'ULM permet alors de visualiser la microvascularisation humaine lors du traitement de malformations vitales [77]. Chez l'adulte, en combinant une imagerie par la fenêtre temporal et des méthodes de correction d'aberrations, il est possible de caractériser la dynamique des flux à une résolution de $25\mu m$ [31].

Au-delà de l'imagerie cérébrale, l'ULM a été étendu à d'autres organes en surmontant les défis liés aux mouvements physiologiques (battements du cœur, respiration etc.). Des stratégies de correction robustes ont permis son application dans le myocarde, d'abord chez le rat [78] puis chez l'humain [30]. Des avancées méthodologiques ont également ouvert l'accès à des structures complexes comme la moelle épinière [79]. Enfin, le potentiel clinique de l'ULM a été démontré pour l'évaluation de la fonction rénale [22] et tumorale [21], ainsi que pour le diagnostic différentiel des lésions hépatiques [80] et mammaires [81].

### 2.3.2 ULM dynamique, fonctionnelle et 3D

Initialement développée en 2D avec un temps d'acquisition de l'ordre de la minute, l'ULM a rapidement été étendue afin d'imager des phénomènes rapides avec une résolution effective de l'ordre de la milliseconde [25, 82, 83]. En faisant l'hypothèse de la périodicité d'un signal temporel, comme le cycle cardiaque ou un stimulus fonctionnel répété, l'accumulation des

positions des microbulles peut être synchronisée a posteriori avec le signal. D'abord introduite avec une synchronisation sur le cycle cardiaque en 2D [24, 84], cette approche a permis l'imagerie de la pulsatilité cérébrale à une échelle microscopique. Elle a ensuite été appliquée à l'imagerie fonctionnelle (fULM) [2,26] afin d'observer la réponse hémodynamique du cerveau à des stimuli.

Parallèlement à ces avancées temporelles, l'extension vers l'ULM 3D a permis de s'affranchir des limitations intrinsèques à l'imagerie en coupe. En effet, l'acquisition 2D souffre d'une forte dépendance à l'opérateur pour la sélection du plan d'imagerie, que ce soit pour le positionnement ou l'orientation. En pratique, cela nuit à la répétabilité des mesures ainsi qu'au recalage précis des volumes pour le suivi longitudinal ou à la validation par d'autres modalités. L'ULM 2D est intrinsèquement limitée dans la fidélité des mesures qu'elle peut extraire en projetant une structure tridimensionnelle sur un plan [85]. Cette projection peut notamment fausser les mesures de tortuosité, de vitesse et de connectivité vasculaire [85] et limiter l'efficacité des méthodes de correction de mouvement hors-plan [78].

L'ULM 3D se base principalement sur des sondes matricielles adressées individuellement ("fully-addressed") [35,37,85], multiplexées [23,25,36,86], ou des sondes lignes-colonnes (Row-Column (RCA)) [82, 87, 88]. Les sondes matricielles *"fully-addressed"* offrent un contrôle optimal de l'onde transmise et une cadence d'imagerie élevée, mais nécessitent de piloter individuellement des milliers d'éléments piézoélectriques, entraînant une complexité matérielle, des coûts élevés et de grands volumes de données. Le multiplexage permet de réduire cette complexité électronique, mais au prix d'une réduction drastique de la cadence d'imagerie volumique, ce qui peut limiter l'observation de phénomènes rapides tels que les flux des plus gros vaisseaux [25] ou dégrader la qualité de l'image. L'utilisation de sondes RCA permet de réduire massivement le câblage via des éléments allongés disposés en lignes et en colonnes. La focalisation repose sur l'émission de multiples angles et une réception orthogonale, ce qui induit une perte de qualité d'image comparée aux matrices pleines. Cependant, l'optimisation des paramètres d'acquisition offre un équilibre suffisant entre cadence et qualité pour capturer la pulsatilité hémodynamique du cerveau de souris [82]. Récemment, l'utilisation d'une sonde multi-lentilles (*Multi-lens*) a permis d'étendre considérablement le champ de vue pour caractériser des organes entiers chez le porc (foie et rein)

## 2.4 Limites inhérentes et directions d'optimisation

### 2.4.1 Compromis entre résolution temporelle, volume de données et concentration

L'obtention d'une reconstruction vasculaire complète repose sur la détection d'un nombre suffisant de microbulles pour échantillonner l'intégralité du réseau vasculaire. Pour saturer la reconstruction capillaire avec des méthodes de localisation conventionnelles, qui requièrent des bulles spatialement isolées, il est nécessaire d'accumuler les images sur des temps longs, de l'ordre de plusieurs minutes [32]. Cette contrainte est exacerbée en dynamique (DULM) ou en imagerie fonctionnelle (fULM), où la répétition de cycles de stimulation s'étend sur une dizaine de minutes pour obtenir une robustesse statistique suffisante.

Cette durée d'acquisition génère des volumes de données massifs, imposant des contraintes matérielles sévères pour le stockage et le traitement. À titre d'exemple, une acquisition volumique haute fréquence peut générer plusieurs centaines de gigaoctets, voire depasser le téraoctet pour des séquence d'imagerie les plus longues.

L'augmentation de la concentration en microbulles permet théoriquement de réduire la durée d'acquisition et le volume de données. Toutefois, cette densification entraîne un chevauchement des signaux (*overlapping*) qui sature l'image et invalide l'hypothèse de sparsité requise par les algorithmes conventionnels. Ce compromis impose une limite stricte : une faible concentration préserve la précision au prix du temps, tandis qu'une forte concentration accélère l'examen mais dégrade la résolution. Au-delà d'un seuil critique, l'incapacité des méthodes standards à discriminer les sources chevauchantes provoque un échec de la détection et une chute drastique du nombre de microbulles localisées

### 2.4.2 Limites de la localisation conventionnelle : formalisme et intractabilité

Afin d'expliciter les limites de l'hypothèse de bulles isolé, le signal ultrasonore peut se formaliser comme une somme $I(\mathbf{r})$ à la position $\mathbf{r} \in \mathbb{R}^3$. En ULM, ce signal est la superposition linéaire des réponses impulsionnelles de $N$ microbulles d'amplitudes $\sigma_i$ et de positions $\mathbf{r}_i$, additionnée d'un bruit $\eta(\mathbf{r})$ :

$$I(\mathbf{r}) = \sum_{i=1}^{N} \sigma_i \cdot \mathrm{PSF}(\mathbf{r}, \mathbf{r}_i) + \eta(\mathbf{r}) \tag{2.1}$$

**Cas 1 : Régime parcimonieux (Bulles isolées)** L'hypothèse de sparsité suppose une distance inter-bulles supérieure à la résolution du système $\Delta_{\mathrm{res}}$. Si $\forall i \neq j, \|\mathbf{r}_i - \mathbf{r}_j\| > \Delta_{\mathrm{res}}$,

les supports des PSFs sont disjoints (i.e. $\mathrm{PSF}(\mathbf{r}, \mathbf{r}_i) = 0$). Au voisinage de la $k$-ème bulle, l'équation (2.1) se simplifie :

$$I(\mathbf{r} \approx \mathbf{r}_k) \approx \sigma_k \cdot \mathrm{PSF}(\mathbf{r}, \mathbf{r}_k) + \eta(\mathbf{r}) \tag{2.2}$$

Le problème inverse se réduit à une estimation locale de paramètres (centroïde ou ajustement Gaussien), robuste au bruit comme illustrée en Figure 2.4, (A).

**Cas 2 : Régime dense (Interférences)** À haute concentration, la condition de séparation n'est plus respectée. Le signal au voisinage d'une bulle $k$ s'ajoute à celui des bulles dans son voisinage $\mathcal{V}_k$ :

$$I(\mathbf{r}) = \underbrace{\sigma_k \cdot \mathrm{PSF}(\mathbf{r}, \mathbf{r}_k)}_{\text{Signal d'intérêt}} + \underbrace{\sum_{j \in \mathcal{V}_k, j \neq k} \sigma_j \cdot \mathrm{PSF}(\mathbf{r}, \mathbf{r}_j)}_{\text{Interférences}} + \eta(\mathbf{r}) \tag{2.3}$$

Ce terme d'interférence agit comme un bruit structuré dépendant de la configuration inconnue des voisins. Les interférences créent des maxima artificiels ou déplacés, comme illustré en Figure 2.4.(B). Les algorithmes conventionnels basés sur la détection de maxima locaux échouent à séparer les sources, conduisant à une sous-estimation du nombre de microbulles et à une augmentation de l'incertitude de localisation.

### Conclusion du chapitre

L'ULM comble un manque crucial de solutions d'imagerie microscopique en profondeur *in vivo*. Son potentiel translationnel et son intérêt préclinique ont été démontrés par les premières applications chez l'humain et de multiples études sur modèles animaux.

Cependant, son adoption à grande échelle reste aujourd'hui contrainte par des limites inhérentes, dictées par les compromis stricts entre temps d'acquisition, volume de données et concentration de microbulles. Pour s'affranchir de ces limitations physiques, l'apprentissage profond (*Deep Learning*) apparaît comme un outil intéressant pour utiliser au mieux les larges quantités de données générées afin de mieux pouvoir modéliser de manière implicite les signaux dans les zones à haute concentration. Par sa capacité à modéliser des interférences complexes là où les méthodes classiques échouent, il offre la perspective d'une imagerie super-résolue rapide et robuste.

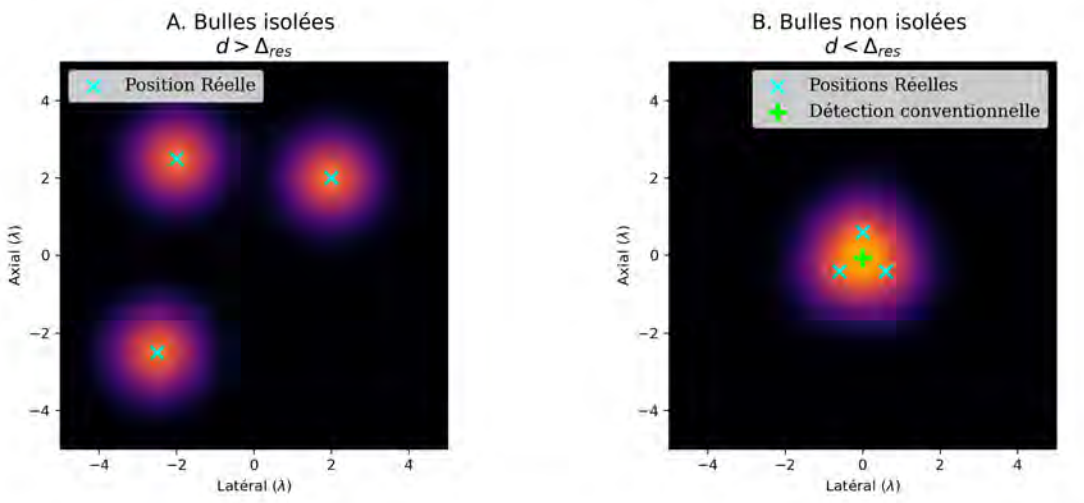



FIGURE 2.4 Impact de la densité de sources sur l'imagerie ULM 2D. (A) Régime Parcimonieux : Trois microbulles séparées par une distance supérieure à la résolution ($\Delta_{res}$). Les taches de diffraction sont distinctes, permettant une localisation précise (croix cyans). (B) Régime Dense :Trois microbulles regroupées dans un volume inférieur à la résolution. Les interférences forment une unique tache large. Un algorithme conventionnel détecte à tort une seule source centrale (croix verte), induisant une erreur de comptage et de positionnement.

# CHAPITRE 3 ARTICLE 1 : DEEP LEARNING IN ULTRASOUND LOCALIZATION MICROSCOPY : APPLICATIONS AND PERSPECTIVES

Le chapitre précédent présentait un panorama général de la microscopie de localisation ultrasonore (ULM), en retraçant ses principes physiques, ses principales applications ainsi que les limitations expérimentales et méthodologiques qui limitent aujourd'hui son déploiement à grande échelle. Cette mise en contexte permet d'identifier clairement les étapes du pipeline ULM qui demeurent sensibles aux conditions d'acquisition, aux paramètres expérimentaux ou aux choix de traitement, et pour lesquelles des améliorations restent nécessaires.

Dans la continuité de cette revue générale, le présent chapitre se concentre sur un axe particulier qui a suscité un intérêt croissant au cours des dernières années : l'utilisation de l'apprentissage profond pour améliorer les différentes étapes de l'ULM. L'objectif est d'offrir une lecture structurée des contributions récentes, de souligner leurs atouts, leurs limites et leurs divergences méthodologiques, et de mettre en lumière les défis encore ouverts pour une utilisation robuste et reproductible de l'ULM, en particulier dans des conditions *in vivo* ou à fortes concentrations de microbulles.

Ce chapitre correspond au premier article de cette thèse, publié en ligne en septembre 2024 dans *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, dans le numéro spécial *Breaking the Resolution Barrier in Ultrasound* (Volume 71, Issue 12, December 2024). Il présente une revue de littérature exhaustive sur les méthodes d'apprentissage profond appliquées à l'ULM et constitue l'état de l'art thématique qui servira de fondement aux développements méthodologiques proposés dans les chapitres suivants.

# Deep Learning in Ultrasound Localization Microscopy: Applications and Perspectives

Brice Rauby[1,2], Paul Xing[1], Maxime Gasse[3,4,2], Jean Provost[1,5]

[1]Department of Engineering Physics, Polytechnique Montréal, QC, Canada
[2]Mila – Quebec AI Institute, Montréal, QC, Canada
[3]ServiceNow, Montréal, QC, Canada
[4]Department of Computer Engineering and Software Engineering, Polytechnique Montréal, QC, Canada
[5]Montreal Heart Institute, Montréal, QC, Canada

## 3.1 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.

## 3.2 Introduction

Inspired from development of super resolution imaging techniques in optical microscopy [63, 64], Ultrasound Localization Microscopy (ULM) leverages the detection and localization of individual microbubbles injected into the bloodstream to overcome the diffraction limit in ultrasound imaging [13, 14]. 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 [11]. ULM has also been extended to 3D imaging either using fully-addressed matrix arrays [35, 37, 85], multiplexed matrix arrays [36] or row-column arrays [82, 87, 88]. ULM proof-of-concepts in pathological animal models have shown, e.g., the characterization of vascular function impairments in Alzheimer's Disease (AD) mice models [19] and between early phases of ischemic and hemorrhagic strokes in mice models [18]. Novel sequences also enable the extraction of dynamic quantities in ULM such as pulsatility imaging in the brain [24, 25], cardiac imaging [30, 78] or functional imaging [2, 26] by using high microbubble detection rates and retrospective gating. Singular microbubble behaviors have also been leveraged to highlight specific structures, such as glomeruli [22, 23]. Applications in humans have also been proposed for aneurysm imaging in the brain [31], breast [21, 27] or pancreas cancer imaging [28], lymph node metastatic cancer [89], kidney [22, 28, 29], prostate [90], lower limb muscle [33], liver imaging [28], vasa vasorum of the carotid wall [91], and testicular microcirculation [81].

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 [32]. Second, ULM is degraded by skull aberration in brain imaging, clutter and cardiac motion in cardiac imaging, and tissue motion in general [8]. 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 [8].

Deep learning algorithms excel at signal processing tasks, driven by the increasing availability of computational power and large-scale datasets [92]. Since AlexNet [39] reduced the top-5 error rate on the ImageNet Large Scale Visual Recognition Challenge from 26.1% to 15.3% in 2012 [93], subsequent deep learning models have further decreased this error to only 3.6% within three years [94], 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 [95],

multi-instance segmentation [96] or image generation [97, 98]. Advances in one specific domain often translate into other domains as well, with several key components often reusable across different tasks, like optimization algorithms [99, 100], normalization layers [101, 102], activation functions [103, 104] or backbone building blocks [105, 106]. For example, transformers originally developed for Natural Language Processing (NLP) [106] have later been applied to image processing tasks with great success [107]. Foundation models, which are also originating from the field of NLP, [108], 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 [109], and can be applied to segment new images even from unrelated distributions [110]. 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 [41]. 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 [111–115], and clutter suppression in Contrast-Enhanced Ultrasound (CEUS) [43]. For recent reviews of deep learning methods in a general ultrasound settings, see [42, 116]. In ULM, deep learning methods have served several purposes such as reducing processing [46, 117, 118] or acquisition times [45, 46, 117], enhancing image quality [2, 46, 118, 119], improving blood velocity estimation [2, 47], and increasing robustness to challenging experimental settings such as increased microbubble concentrations [2, 45, 120] or phase aberrations [69]. 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 3.2):

— 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 3.3, we discuss the different existing approaches related to generating labeled datasets used either in training or evaluation of deep learning-based ULM methods. In 3.4, we review the different application stages of deep learning in the ULM pipeline. In 3.5, we focus on the localization stage of the ULM pipeline, where most deep learning approaches have been applied. Finally, in 3.6, we summarize this review with an overview of the successes, limitations, and open challenges for deep learning-based ULM methods.

### 3.3 Generation of labeled 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 [121]. 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 [50], and PALA [67], 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 [93, 122, 123], 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.

### 3.3.1 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 $p_{\text{train}}$ and $p_{\text{test}}$ to differ, which limits the validity of the i.i.d. hypothesis. Some level of realism in the generation of $p_{\text{train}}$ 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.

### 3.3.2 Prior probability $p(y)$: label generation

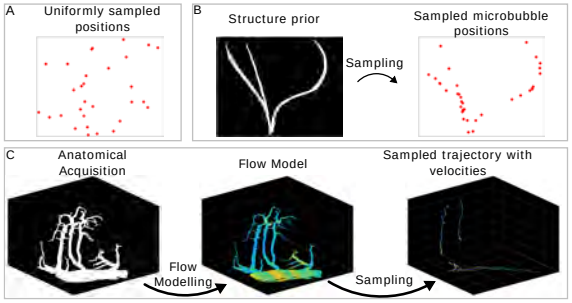



Figure 3.1 Representation of the different types of label generation strategies illustrating sampling from different prior distribution. A: Uniform sampling of independent frames. B: Structure based sampling of independent frames. C: Trajectory sampling based on anatomical acquisition and simulated flow from [124]

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 3.1 with the corresponding prior distribution and temporal context, and different categories of prior distribution are illustrated in Fig. 3.1. 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 [44, 46] or trajectories across multiple frames [45, 75, 134].

#### Single-frame simulations

In single-frame simulations, some approaches have used spatially uniform sampling [46, 47, 119, 126, 130, 133, 135], while others have employed a spatial distribution of scatterers conditioned by a given structure, often mimicking vasculature [44, 46, 118, 126]. The structure used to condition the prior have been handcrafted [67, 118], simulated [44], based on ULM acquisitions [46, 126], or derived from acquisitions from other modalities [45, 126, 134]. 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 model microbubble positions constrained to blood vessels, potentially reducing the validity of the i.i.d. hypothesis

Table 3.1 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. [125] | Independent frames | Uniform | Uniform | N.A. | Non-linear propagation and response |
| Chen et al. [126] | Independent frames | Uniform and Structure | Structure | CAM | Field II [127] |
| Chen et al. [75] | Simulated flow | Structure | Structure | CAM | Field II [127] and convolution with Gaussian PSF |
| Gu et al. [128] | Independent frames | Structure | Structure | Artificial Vasculature | Convolution with Gaussian PSF |
| Hahne et al. [118] | Independent frames from simulated flow | PALA [35] | PALA [35] | Artificial structure | Verasonics Research Ultrasound Simulator |
| Liu et al. [129] | Independent frames | Uniform | Structure | Artificial vasculature | Convolution with Gaussian PSF |
| Liu et al. [46] | Independent frame | Uniform | Structure | Vasculature from *in vivo* acquisition | Convolution with Gaussian PSF |
| Luan et al. [130] | Independent Frames | Uniform | Structure | Artificial Vasculature | High-resolution convolution with Gaussian PSF |
| Milecki et al. [45] | Simulated flow | Structure | Structure | Vasculature from *ex vivo* acquisition | SIMUS [131] |
| Pustovalov et al. [132] | Displacement map | Structure | Structure | CAM and Artificial Structure | Convolution with experimental PSF |
| Shin et al. [2] | Simulated flow | Uniform | Uniform | N.A. | Convolution with learned distribution of PSF |
| van Sloun et al. [42] | Independent frames | Uniform | Structure | Artificial Vasculature | Convolution with Gaussian PSF |
| Youn et al. [47] | Independent frames | Uniform | Uniform | N.A. | Field II [127] |
| Yu et al. [133] | Independent frames | Uniform | Uniform | N.A. | Convolution with Gaussian PSF |
| Zhang et al. [119] | Independent frames | Uniform | Uniform | N.A. | Convolution with Gaussian PSF |
| Zhang et al. [120] | Independent frames | Uniform | Structure | Artificial Vasculature | Convolution with Gaussian PSF |

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 [67,76]. 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.

**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) [124]. The 2PM-acquired vascular structure was segmented and converted into a graph model using an existing framework [136]. 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. [75] 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. [132] 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 [137] using a recursively-generated simulation framework. Flow and pressure were then simulated based on Navier-stokes equations for an incompressible Newtonian 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 [50].

Alternatively to the low-entropy prior based on anatomical structures, Shin et al. proposed a multi-frame simulation approach with a high-entropy prior [2]. 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.

**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 [126]. 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 [22], 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 [138, 139], 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.

### 3.3.3 Conditional probability $p(x|y)$: 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 [46, 128, 130, 133, 135]. 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 [2].

Several ultrasound simulators have been proposed in the literature to model acoustic wave propagation. Some are mesh-based, like k-wave [140], which allows for simulation of nonlinear acoustics with multiple scattering and heterogeneous media. To allow for fine po-

sitioning of scatterers and short computation time, most of the surveyed studies have used particle-based simulators such as Field II [127] and SIMUS [131], 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 [141] would allow the application of deep learning methods to process ULM data relying on the non-linear response of microbubbles [142] 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 [141]. Inspired by the combination of k-wave and the Marmottant model of microbubble response by Brown et al. [71], Lerendegui et al. [137] proposed integrating the non-linear microbubble response into a linear simulator, Field II, using a two-step process. First, BuFF [137] 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 [141]. 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 [141], Blanken et al. [125] proposed to also account for the non-linear propagation of ultrasound [143]. 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) [144, 145], 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 [46, 126, 130, 135], or Rice distribution on B-mode, which is more specific to ultrasound data [2]. Aiming to produce realistic noise, Xing et al. [69] 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. 3.3).

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.

### 3.3.4 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 [2, 119, 133]. This ensures that the training and test sets are i.i.d., leading to improved *in vivo* performance [2].

PSF deconvolution is an image processing technique, which, when applied to conventional ULM, jointly estimates the PSF and the position of isolated scatterers [146]. PSF 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 [146]. Improving on this method, Zhang et al. proposed training two networks concurrently to estimate the PSF and the scatterer distribution [119]. 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. [147] proposed a similar approach, leveraging self-supervised learning on *ex vivo* CAM dataset. Shin et al. [2] 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 [2]. Lok et al. [134] 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. [133]. They proposed a method aiming to accelerate an existing block matching algorithm for denoising of ULM data before localization [148]. With a few *in vivo*

samples labeled 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.

## 3.4 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 3.2, 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 3.2. Additionally, we discuss steps such as aberration correction and denoising, which may enhance image quality when incorporated into the ULM pipeline.

### 3.4.1 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 [150]. Existing approaches often utilize speckle brightness [151], measurements cross-correlation from neighboring transducer elements [152], or iterative time reversal [153, 154] 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 [155–157]. Deep learning-based approaches have also been developed [158–161], 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 [13,34]. In small animals, such as young rodents, direct imaging through the skull is possible because its impact on imaging quality is minimal [31,36,162].

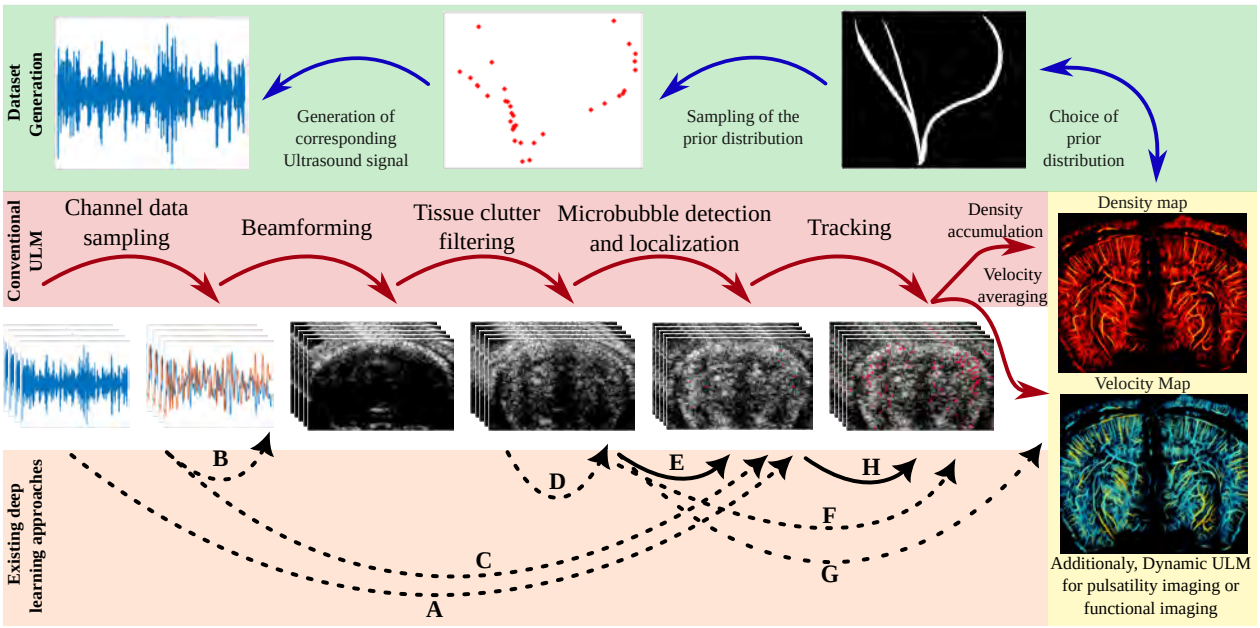



Figure 3.2 Overview of ULM processing and simulation pipeline, and the existing deep learning approaches. Dashed lines represent imperfect correspondence with pipeline steps, full lines represent perfect replacement of one or more steps. We note that some pipeline stages can be performed in different orders: such as the beamforming and the clutter filtering [118] or the tracking and localization [74]. 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 [125]. B: Xing et al. correct for aberration based on In-Phase/Quadrature (IQ) data [69]. 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 [47, 118]. D: Yu et al. enhance the Signal-to-Noise Ratio (SNR) denoising post SVD filtering [133]. E: The localization step has been investigated by several studies [2,44,46,119,120,126,129,130,132]. F: Milecki et al. [45] and Lok et al. [134] 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 [75]. H: Zhang et al. [149] introduced a Gated Recurrent Unit based Multitasking Temporal Neural Network (GRU-MT) to solve the assignment problem to form microbubbles trajectories from detected positions.

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 [86, 163, 164]. By leveraging the theoretical RF echoes of isolated microbubbles, these methods estimate a phase aberration function using either iterative estimation and virtual focusing [163] or by solving the inverse problem of the imaging process [164].

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 [69]. 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. 3.3, this approach demonstrates convincing *in vivo* results in older mice (6 months), outperforming coherence-based 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 [45, 126].

### 3.4.2 Beamforming

While initial ULM proof-of-concepts detected microbubble echoes in channel/fast-time space prior to beamforming [15, 16, 166], more recent ULM studies performed beamforming, typically using the delay-and-sum (DAS) algorithm [70], before microbubble localization [67]. 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 [45, 47, 118, 125, 126]. To do so, some studies used the signal in channel/fast-time space and performed localization either in this space [125] or directly in the image space with specific projection layers [47, 118]. Since only the signal corresponding to microbubble positions needs to be projected into spatial coordinates, the beamforming operation can be simplified [118], which can save processing time and limit beamforming issues like grid artifacts.

Youn et al. [47] argued that overlapping PSFs caused by high scatterer density induce a loss of information. To alleviate this issue, they used a Convolutional Neural Network (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 [167], 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

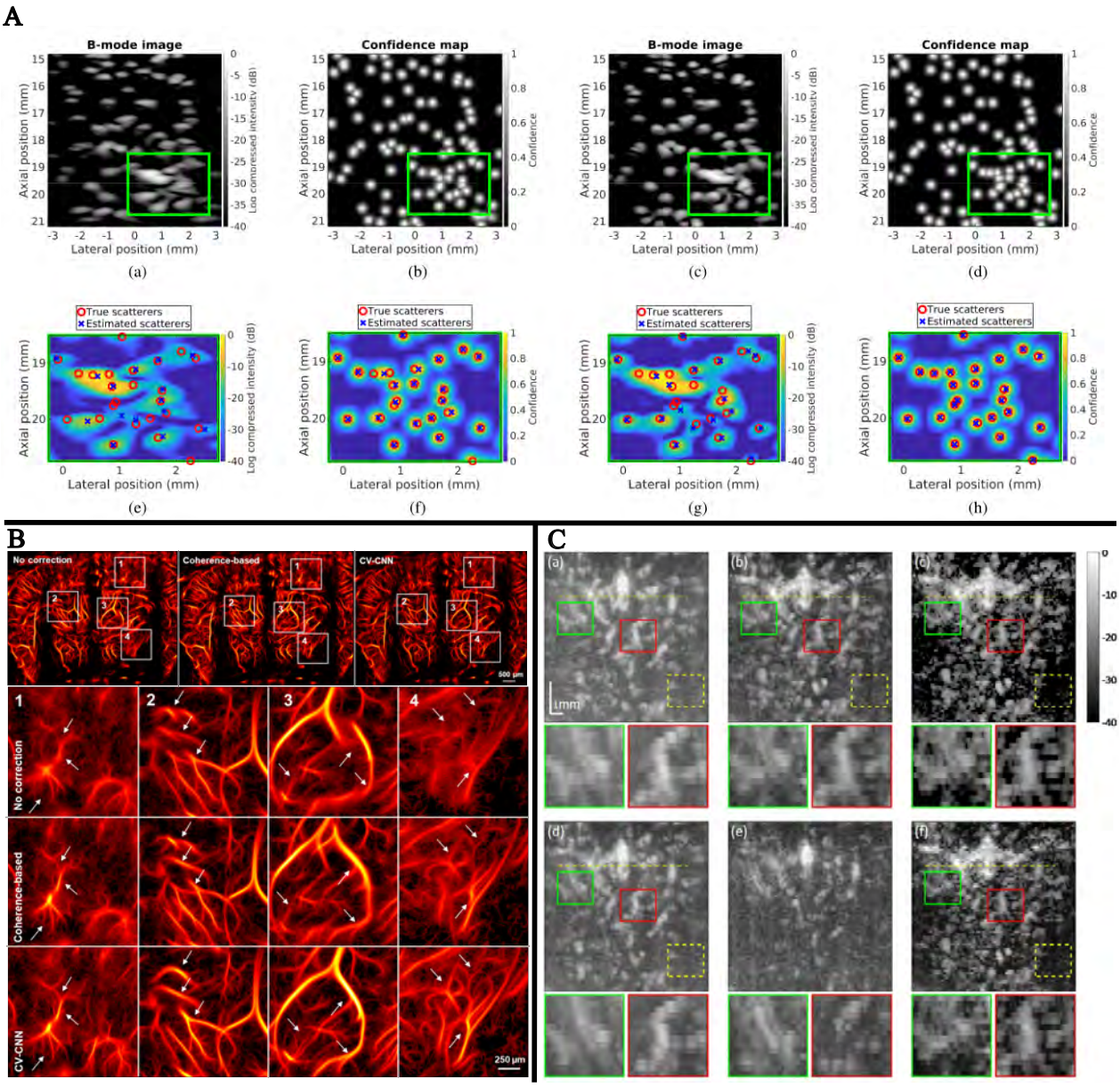



Figure 3.3 A: In silico results from [47] ©2020, IEEE. B-mode image and local maxima detection compared with the confidence map from the deep learning model [47]. B : Illustration of results obtained with deep learning based aberration correction from [69]. *In vivo* imaging on 6 months old mouse through 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 [152]. 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 [43]. The Deep unfolded model-based method, CORONA (c) from [43] is compared with SVD clutter filter [72] (a), the model based approach, FISTA [165], (b),6th order Butterworth filter (d and e), and a ResNet [40] trained to perform clutter removal. Color bar is in dB. Reused with the authorization of the authors ©2020, IEEE

single channel [125]. 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 [125], applying a 2D end-to-end approach (i.e., using all RF channels and producing an 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. [118] 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 [118].

### 3.4.3 Clutter filtering and denoising

To detect microbubbles, various strategies have been explored to remove tissue clutter, including non-linear imaging [30,142,168,169] and post-processing filtering [72]. In ULM, SVD clutter filtering [72] is widely used to remove tissue clutter due to its simplicity and efficiency. Alternative filtering methods such as high-pass filtering [22,76] and mean removal [170] 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 [43, 48, 171]. Among these, Solomon et al. [43] 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 [172] or block matching [148] 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. [133] proposed a data-driven approach to learn the denoising step performed by block matching [148] from *in silico* and *in vivo* data. This approach uses a domain adaptation method, namely Domain Specific Projection from [173], 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.

### 3.4.4 Localization

After beamforming and filtering, conventional ULM often performs a simple detection step based on local maxima of intensity [35], SNR [36], or local correlation with the PSF [24, 45, 172]. 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 [67]. 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 [32]. 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* [2,45]. Numerous approaches have been proposed, and a more focused and exhaustive review is provided in 3.5.

### 3.4.5 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 [25, 28, 174, 175]. Deep learning improved Kalman filter [176] has been applied jointly with the Hungarian algorithm [50,177], 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. [149] have proposed an approach that solves the assignment problem and improves the position estimation based on a Gated Recurrent Unit [178] 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. [179] 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. [45] 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 [134]. These approaches were able to reconstruct high quality density maps *in vivo* using high microbubble concentration [45, 134]. To extend these applications to velocity estimation, Chen et al. [75] 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.

## 3.5 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. 3.4. 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.

### 3.5.1 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 [50] 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 [137]. Heiles et al. also proposed a benchmark to evaluate localization methods [67] 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 [183], allowing the comparisons of methods that utilize the full RF information in channel/fast-time space [47, 118, 125] or in image space [45, 126]. This benchmark has been used in recent studies [74, 118, 184].

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.

#### 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.

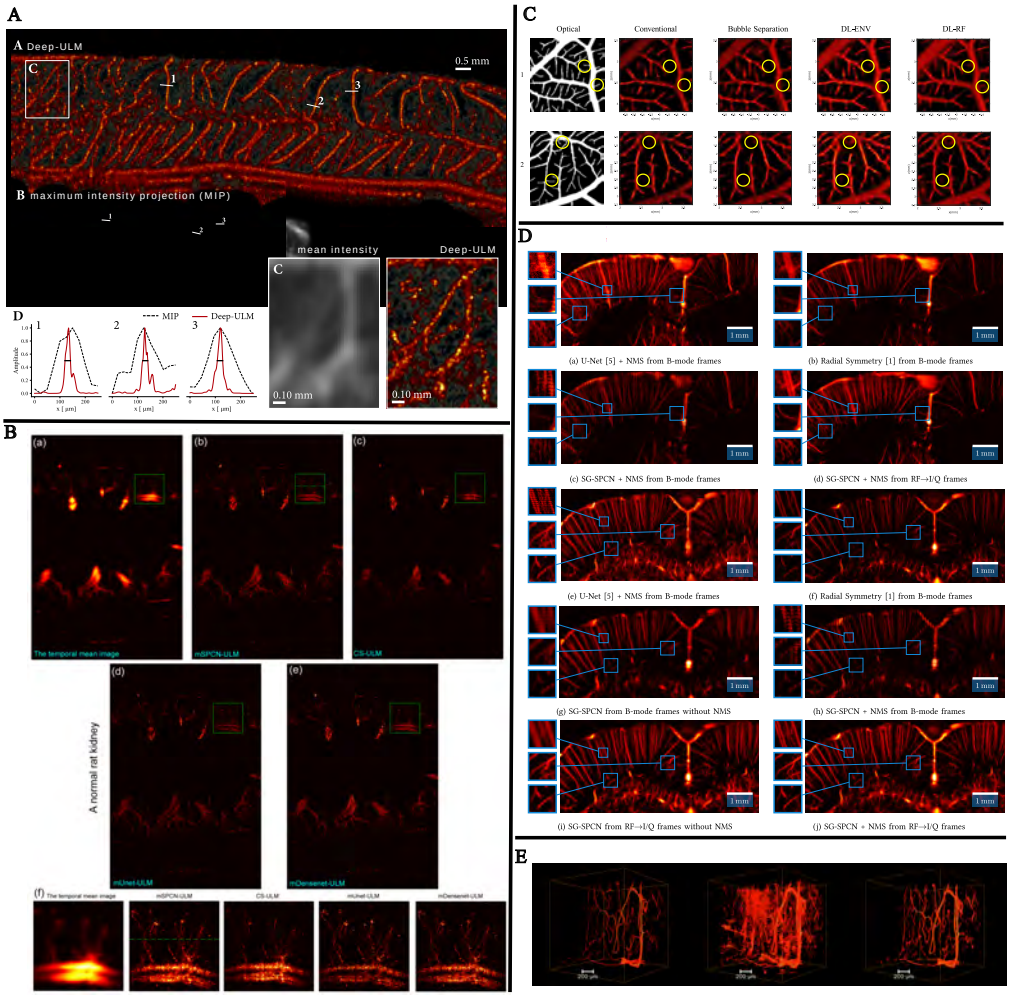



Figure 3.4 Illustration of *in vivo* results in 2D from different deep learning localization approach in various organ and animal models (A, B, C, and D) and *in silico* results in 3D from [1] (E). A: Rat spinal cord processed with Deep ULM [44] compared to Maximum Intensity Persitence (MIP) image from [44] ©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 $21\mu$ m, $19\mu$ m, and $20\mu$ m for profiles 1, 2, and 3. B: Rat Kidney from [46] ©2020, IEEE. Acquisition with a frame rate of 400Hz and a transmit frequency of 15.625MHz processed mSPCN-ULM [46], mUnet-ULM adapted from [42], mDensenet-ULM adapted from [180], and CS-ULM [181]. C: Chicken embryo CAM image from [126]. 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 [126] compared with Fourier based microbubble separation from [182] and conventional ULM. D: Rat brain image from [118] ©2024, IEEE. Data from PALA [67] processed with SG-SPCN with and without Non-Maximum Suppression (NMS) on B-mode or IQ in channel/fast-time space [118] and compared with U-Net from [44] and Radial Symmetry [67] on B-mode. E: In silico comparison in 3D for conventional ULM (center) and sparse tensor neural network (right) with ground truth (left) from [1].

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 [129] or to quantify the impact of using Sparse Tensor Neural Networks in ULM [1].

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 [119,125], aberrations [2], and high microbubble concentration [45,119,185]. 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 [45, 119, 185].

*In silico* evaluations often employ pre-existing metrics to measure the precision and recall of the models [46]. To measure localization precision, distance-based metrics such as Root Mean Squared Error (RMSE) can be used for detected localizations [46]. These are standard evaluation metrics also found in PALA and ULTRA-SR benchmarks [50,67]. 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 [45].

### *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 [2, 75, 126], qualitative image assessments [45], or evaluation metrics that do not require ground truth, such as Fourier Ring Correlation (FRC) [76] 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 [44–46, 75, 130]. 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 [76] to evaluate the resolution of deep learning methods *in vivo* [2]. 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 [2].

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 [2, 75, 126]. 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 aberrations. Validation using Magnetic Resonance Imaging (MRI) or computed Tomography (CT) in sheep brain, human brain, and heart has been performed by conventional ULM studies [30,31,170]. 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 [23].

**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 [67], with reproducible metrics [76], 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.

### 3.5.2 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.

#### Ultrasound representation

Many approaches directly use B-mode images [42, 46, 75, 126, 129, 186]. RF data can also be used either before [125] or after beamforming [126] 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 [126].

To benefit from lossless IQ compression and reduce computation cost, some approaches [45, 118] 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) [187], leveraging complex arithmetic and complex model parameters, can also be used to process the complex-valued IQ data in ultrasound image reconstruction [188] or for aberration correction in ULM data [69].

#### 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 [44, 46]. Relying on temporal context to enhance detection, some methods consider multiple frames simultaneously, raising challenges in input handling and output formulation. Deep-stULM [45], for instance, used a 3D spatio-temporal CNN to process 512 frames, outputting a temporal projection of microbubble positions. Lok et al. [134] 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 [128]. To provide further information, Deep-SMV [75] used 16-frame temporal context and sequence modelling, specifically Long Short Term Memory (LSTM) [189] to predict velocity maps. Lee et al. proposed to use optical flow estimation

to represent temporal context [190]. Shin et al. proposed LOCA-ULM inspired by Single Molecule Localization Microscopy (SMLM) architecture [191] to incorporate temporal context from adjacent frames for improved localization in single frame [2]. 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. [132] also proposed an architecture inspired by DECODE [191], which uses 3D convolutions to incorporate the temporal context.

**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 [44, 45].

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 [44]. Similarly, Liu et al. applied the convolution kernel to both the network output and the labels [46]. Reducing the Gaussian kernel's standard deviation during training can also promote sparser predictions [118], allowing for the use of simple $L_2$-based loss functions, such as Mean Squared Error. Incorporating more complex loss functions based on SSIM [126] or focal loss [130] can smooth the loss function and improve training stability. Inspired by medical image segmentation [192], some works have used the dice coefficient either alone [45] or in combination with $L_1$ applied to soft labels [125]. 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. [193] leveraged Wasserstein GANs [194] to improve their loss formulation and enhance localization performance. Shin et al. [195] 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 [191], 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 multi-

channel 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 [75] or non-overlapping Gaussian confidence maps [47], also reduces training instability and permits the use of simpler loss functions. Alternative approaches using object detection formulations have also been proposed [196, 197].

### 3.5.3 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 [44–46, 126, 134], 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.

#### 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 [198], resulting in gridding in the ULM image. Upsampling has been done using methods such as nearest neighbor upsampling [45] or transposed convolution [42]. A more efficient approach is to use sub-pixel convolution [199], which has been applied for localization in ULM [46]. 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 [2, 191], which allows for sub-pixel resolution with limited dependence on the grid. However, grid artifacts can still occur in high-concentration areas [191] but can be mitigated by interpolating the model's input [2]. Performing the localization in channel/fast-time space can also be key in eliminating grid artifacts [118].

### 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 [125]. 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 [47]. Alternatively, Blanken et al. utilized dilated convolutions [125]. Hahne et al. proposed adding a semi-global bottleneck block to ensure a sufficiently large receptive field [118]. These methods can also handle the beamforming operation, which can be done implicitly through position embedding [47] or with a learned projection of the detection [118].

### 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 [117, 120, 129, 135, 197, 200, 200–203]. For example, Liu et al. proposed SR-MT (Super-Resolution Modified Transformer) [135], based on the Swin transformer block [204], 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. [130] 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. [197] modified the problem formulation to an end-to-end object detection framework and applied DEtection TRansformers (DETR) [205]. To address inherent limitations of DETR in small object detection, Gharamaleki et al. [201] also proposed using Deformable DETR [206] within a similar framework. Transformer-based architectures have also been used to improve ULM reconstruction from averaged B-mode images [117], showing improvements both *in silico* and *in vivo* compared to previously introduced GAN-based methods [128]. Zhang et al. [120] 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*.

### Scaling to 3D imaging

ULM and Dynamic ULM (DULM) have been extended to 3D imaging [25,35,36,207] 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 [1, 208]. Piepenbrock et al. proposed using a 3D CNN to perform localization and process each volume independently [208]. Rauby et al. [1] used Sparse Tensor Neural Networks [209] 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 3.4.

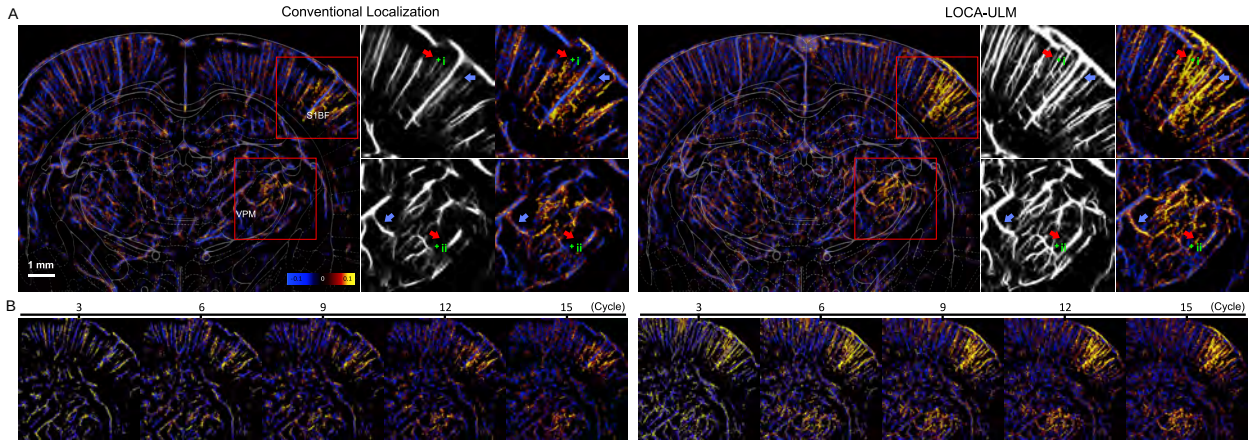



Figure 3.5 Functional ULM (fULM) comparison between conventional localization and LOCA-ULM from [2] showing the improved microbubble detection ability of LOCA-ULM, allowing to reduce the required number of stimulation cycles. Reused with the authorization of the authors

## 3.6 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.

### 3.6.1 Limitations and future challenges

#### 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 [20]. Sensing ULM (sULM) utilized trajectories metrics such as normalized distance, remanence and dispersity [22, 23] to identify microbubbles travelling in glomeruli. Trajectories can also be used to estimate blood flow pulsatility [210] which has been link to AD and cognitive decline [211–214]. 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 [74]. 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 [45, 134], 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 [1], but they have not yet been applied *in vivo*.

**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 [50, 67], 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 [67] and ULTRA-SR challenge [50] 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 [67, 76]. Works in biomedical image analysis [215] 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 [215, 216] should be adapted and implemented to ensure that progress enabled could translate to 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 [46] 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 [75] or ULM at high concentration [45, 134].

### 3.6.2 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 [2, 44, 46, 47, 120, 126, 130, 200] 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 [2, 44, 46, 47, 126, 130, 200] and in 3D [1]. *In vitro* studies have also shown similar results [47]. *In vivo* comparison studies with conventional ULM have also reported improvements when using deep learning methods for ULM at high concentrations [2, 45, 120, 126]. 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. 3.5).

Recent developments in model architectures have been demonstrated that performance in high concentration of microbubbles could be further improved [120].

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 [43, 148]. Using these algorithms to generate labeled datasets *in vivo* and training deep learning models allows for either direct approximation of the operation [133] with faster computation times or acceleration of the convergence of an iterative process [43]. Efforts to accelerate the processing of ULM have also taken the form of directly forming the ULM image with averaged B-mode images [117, 128, 186]. In addition to drastically reducing acquisition and processing time [117, 186], 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. [217] 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 [218] using supervised [219–222], semi-supervised [223, 224], or unsupervised learning [225, 226], 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. [46] 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. [2] 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 [170, 227], 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.

# CHAPITRE 4 ARTICLE 2 : PRUNING SPARSE TENSOR NEURAL NETWORKS ENABLES DEEP LEARNING FOR 3D ULTRASOUND LOCALIZATION MICROSCOPY

Le chapitre précédent présentait une revue thématique des méthodes d'apprentissage profond appliquées à la microscopie de localisation ultrasonore (ULM), en mettant en évidence les défis méthodologiques actuels ainsi que les pistes ouvertes par les approches récentes. Cette synthèse a notamment souligné la difficulté d'étendre l'ULM au volume complet, en raison des contraintes computationnelles liées au traitement tridimensionnel et de la forte augmentation des besoins en mémoire.

Dans la continuité de cette revue et des limitations identifiées, le présent chapitre se concentre sur le développement d'une approche d'apprentissage profond spécifiquement conçue pour l'ULM 3D. Nous proposons une architecture basée sur des tenseurs parcimonieux (« sparse tensor neural networks ») permettant de réduire drastiquement la complexité mémoire tout en préservant la précision de localisation, rendant ainsi possible l'apprentissage direct à partir de volumes tridimensionnels à forte densité de microbulles.

Ce chapitre correspond au deuxième article de cette thèse, publié en ligne en mars 2025 dans *IEEE Transactions on Image Processing* (DOI 10.1109/TIP.2025.3552198). Il constitue la première contribution méthodologique du manuscrit, marquant la transition entre l'analyse critique de l'état de l'art et les développements techniques proposés dans les chapitres suivants.

# Pruning Sparse Tensor Neural Networks Enables Deep Learning for 3D Ultrasound Localization Microscopy

Brice Rauby[1,2], Paul Xing[1], Jonathan Porée[1], Maxime Gasse[3,4,2], Jean Provost[1,5]

[1]Department of Engineering Physics, Polytechnique Montréal, QC, Canada
[2]Mila – Quebec AI Institute, Montréal, QC, Canada
[3]ServiceNow, Montréal, QC, Canada
[4]Department of Computer Engineering and Software Engineering, Polytechnique Montréal, QC, Canada
[5]Montreal Heart Institute, Montréal, QC, Canada

## 4.1 Abstract

Ultrasound Localization Microscopy (ULM) is a non-invasive technique that allows for the imaging of micro-vessels *in vivo*, at depth and with a resolution on the order of ten microns. ULM is based on the sub-resolution localization of individual microbubbles injected in the bloodstream. Mapping the whole angioarchitecture requires the accumulation of microbubbles trajectories from thousands of frames, typically acquired over a few minutes. ULM acquisition times can be reduced by increasing the microbubble concentration, but requires more advanced algorithms to detect them individually. Several deep learning approaches have been proposed for this task, but they remain limited to 2D imaging, in part due to the associated large memory requirements. Herein, we propose the use of sparse tensor neural networks to enable deep learning-based 3D ULM by improving memory scalability with increased dimensionality. We study several approaches to efficiently convert ultrasound data into a sparse format and study the impact of the associated loss of information. When applied in 2D, the sparse formulation reduces the memory requirements by a factor 2 at the cost of a small reduction of performance when compared against dense networks. In 3D, the proposed approach reduces memory requirements by two order of magnitude while largely outperforming conventional ULM in high concentration settings. We show that Sparse Tensor Neural Networks in 3D ULM allow for the same benefits as dense deep learning based method in 2D ULM i.e. the use of higher concentration *in silico* and reduced acquisition time.

## 4.2 Introduction

Ultrasound Localization Microscopy (ULM) is an imaging method that non-invasively maps the vascular tree and blood velocities at depth *in vivo*. By localizing and tracking individual microbubbles injected into the blood flow [13, 14], ULM achieves an imaging resolution approximately equal to one tenth of the diffraction-limited resolution. More recently, Dynamic Ultrasound Localization Microscopy (DULM) [24, 78] has extended the capabilities of ULM by enabling the generation of retrospectively-gated, super-resolved movies of the blood flow dynamics, with applications in pulsatility mapping [24], functional imaging of the brain [26], and cardiac imaging [78]. ULM has been applied in pathological mouse models, such as to classify ischemic and hemorrhagic stroke [18] and to observe vascular impairments in Alzheimer's Disease [19]. Studies on patients have demonstrated the feasibility of ULM in various human organs, including the brain [31], breast [21, 27], kidney [22, 28, 29], prostate [90], lower limb muscle [33], liver [28], pancreas [28], vasa vasorum of the carotid wall [91], testicular micro-circulation [81], and lymph node metastatic cancer [89]. Existing works have used fully addressed [35, 37, 85], multiplexed array [25, 36, 207], and row-column array probes [82, 87, 88] to extend ULM and DULM to 3D imaging. In addition to the inherent advantages from 3D imaging, such as reduced user dependence in imaging plane selection and robustness to out-of-plane motion, 3D ULM is less sensitive to detection errors caused by out-of-plane trajectories and velocity estimation bias from elevation direction projection, which likely enhances clinical relevance and reliability. Both localization and velocity estimation can be improved by rejecting microbubbles that do not appear across several frames [13]. Such tracking can be performed, e.g., using the Nearest Neighbor algorithm [13, 24] or the Hungarian method [67, 228]. Some approaches have also incorporated Kalman filtering to refine the position estimations of a track [25, 174, 175, 207]. The acquisition time necessary to construct a complete vascular map is mainly dependent on the required time to perfuse all vessels and, thus, on microbubble concentration [32]. However, a trade-off exists between the microbubble concentration and the localization precision and accuracy [124], which can be partially lifted using, e.g., methods based on the compressed sensing theory [229], on the division of the k-space in several subregions [182], or tracking the microbubble signals prior to sub-pixel localization [74]. Deep learning-based methods have also investigated frame-by-frame, spatial-only approaches [44, 46, 126], and, more recently, the spatio-temporal context through convolution [45, 195] or sequential modeling [75].

Despite promising results with increased microbubble concentrations both *in silico* [45] and *in vivo* [45, 195], deep-learning based approaches have been limited to 2D imaging. Indeed, the addition of a third spatial dimension considerably increases the size of intermediate

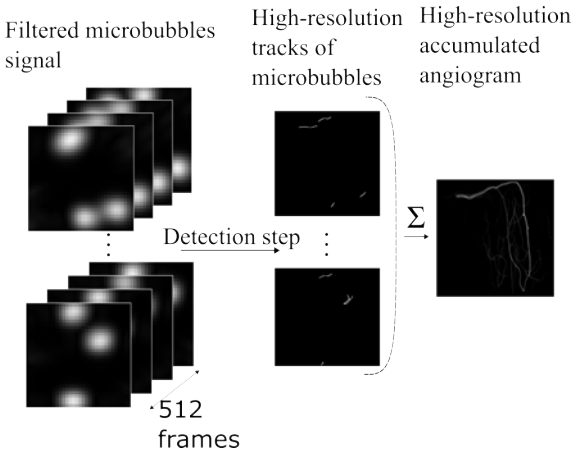



Figure 4.1 The left column represents the filtered microbubble signal (i.e, the input of the network), the center column represents the corresponding microbubble tracks (i.e., the desired output of the network) and the right column represent the final result after summation of all the predictions from a dataset (i.e., the vascular structure imaged)

feature maps and highly-resolved outputs. For example, a straightforward implementation of Deep-stULM [45] in 3D would require at least two orders of magnitude more memory than its 2D counterpart. Thus, the development of deep learning based models for 3D ULM is conditioned on successfully addressing their memory complexity.

To improve the scaling of memory complexity of deep learning approaches in ULM and enable deep learning application for 3D ULM, we propose to leverage the recently introduced Sparse Tensor Neural Networks [209]. Sparse Tensor Neural Networks store only non-zero pixels (referred to as active sites), along with their spatial coordinates and corresponding pixel values (often described as COO format in the literature). This enables the handling of arbitrarily shaped tensors. During training and inference, convolutions and subsequent operations are only applied to these active sites. While ultrasound images are typically dense data that cannot be stored directly as sparse tensors efficiently, filtered microbubbles responses are sparsely distributed in space and time as seen in Figure 4.1. To leverage the sparsity of microbubbles response, one must thus design a filter that extracts microbubble responses from ultrasound images and is sufficiently restrictive that it minimizes memory requirements without discarding information enables the neural network to outperform conventional approaches. Hereafter, filtering out most of the input signal from dense tensors before conversion to sparse format while keeping the signal of interest is designated as the *dense-to-sparse* operation (represented in dashed green in Figure 4.2).

Our contributions can be summarized as follows:

— A sparse formulation of Deep-stULM enabling deep learning-based 3D ULM.

— A comparative study *in silico* between ULM and the proposed approach under varying concentrations in 3D.
— A 2-D *in silico* comparison of performance and memory usage with deep learning baselines and conventional ULM

We show that Sparse Tensor Neural Networks reduce memory cost and scale better with added input dimensions, which allows for the first deep learning application to 3D ULM and outperforms existing 3D ULM in high concentration. We also provide the code and the dataset needed to reproduce the results at https://github.com/provostultrasoundlab/SparseTensorULM.

## 4.3 Theory

After image formation, ULM data is typically filtered to remove the signal from the tissue while retaining microbubble responses. In this study, we focus on the subsequent step, which is to find the sub-resolution positions of the microbubbles based on their response. As illustrated in Figure 4.1, this detection step is performed on a group of frames and is followed by an accumulation across several sets of frames, where the detected positions are later added to form the image or the volume of the vascular network. Several deep learning approaches [44, 46] use convolutional architectures where the microbubble position is projected onto a grid with a finer resolution than the input signal. The upscaling factor $r$ between the input dimension and the output dimension typically ranges from 4 to 8. Therefore, $c_{\text{dense}}$, the memory complexity for storing the output grid of such architectures in dense format scales with:

$$c_{dense}(r, d, D) = (r \times D)^d \tag{4.1}$$

where $d$ represents the dimensionality of the output grid ($d = 2$ for 2D ULM, and $d = 3$ for 3D ULM), and $D$ is a typical dimension of the input in pixels. ULM assumes a sparse distribution of microbubbles, thus an upper bound on the number of microbubbles that can be detected in a certain volume is given by:

$$N < (\frac{\rho D}{\alpha})^d \tag{4.2}$$

where $N$ is the number of microbubbles, $\rho$ is the size of a pixel in wavelength, and $\alpha$ approximately describes the size of the point spread function in wavelengths. In sparse format, $c_{\text{sparse}}$, the memory complexity of storing the microbubble positions scales with $d \times N$ and can be written as:

$$c_{\text{sparse}}(N, d) \simeq \eta \times d \times N_d \tag{4.3}$$

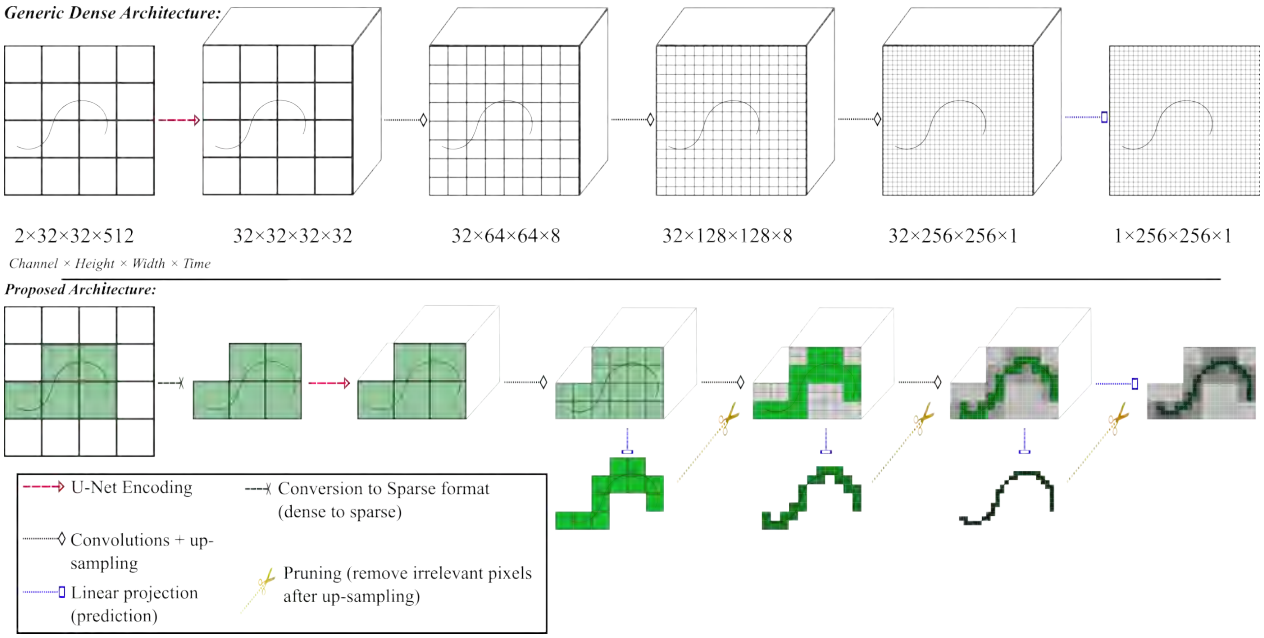



Figure 4.2 The top row shows a dense representation of a trajectory in Deep-stULM as well as the intermediate map dimension. The bottom row illustrates how sparse formulations could reduce the memory cost: the green pixels represent the pixels of interest at each resolution, and the gray pixels represent the pixels removed through pruning based on intermediate prediction

with $\eta$ being a constant depending on the implementation of the sparse tensors. Consequently, $\gamma_d$, the ratio between the sparse and the dense representation of the outputs of the networks in dimension $d$ scales with:

$$\gamma_d \simeq \frac{\eta \times d \times N_d}{(r \times D)^d} \tag{4.4}$$

When extending from 2D to 3D, the ratio between the sparse and the dense representation of an ideal outputs from such network is expected to be multiplied by a factor $\delta$:

$$\delta = \frac{\gamma_{3D}}{\gamma_{2D}} \tag{4.5}$$

Using eq. 4.4, we obtain

$$\delta = \frac{N_{3D}}{N_{2D}} \times \frac{3}{2(r \times D)} \tag{4.6}$$

This scaling law makes the sparse representation very attractive for the extension in 3D. Indeed, assuming that N reaches its upper bound from eq. 4.2, the ratio between sparse and dense representation becomes:

$$\delta = \frac{3}{2}\frac{\rho}{\alpha r} \tag{4.7}$$

Using typical values, $\alpha = 3$, $\rho = \frac{1}{2}$, and $r = 8$, we obtain:

$$\delta = \frac{1}{32} \tag{4.8}$$

This numerical application illustrates the fact that in 3D the use of sparse formulation is from one to two orders of magnitude more efficient than in 2D when compared to the dense representation of ULM output images. However, practical factors such as the variability of the upscaling factor $r$ within the network architecture, the reduced sparsity of intermediate representations, the non-uniform distribution of microbubbles in space, or temporal context considerations, can alter this scaling law.

Time complexity follows a similar scaling law to memory complexity, as convolutions are only applied to active sites (i.e., non-zero pixels) in sparse representation. In contrast, convolutions are applied to every site in dense representation. In practice, convolution implementations on dense tensors are heavily optimized in deep learning libraries and efficiently leverage GPUs and parallel computations [230]. Sparse convolutions require storing active sites, which can be sped up using unordered maps and caching [209]. Different optimization levels of convolution implementations may outweigh the theoretical gains in computational costs of the proposed architectures. For these reasons, we proposed inference time comparisons reported in Table 4.1.

## 4.4 Method

First, we detail the approach used to simulate the 2D and 3D datasets used for training and evaluation of the different methods. Then we describe the model architecture, training parameters, and evaluation metrics. Additional studies on the impact of the *dense-to-sparse* operations and on further architecture modifications such as pruning [231] (represented in gold dotted line and scissors in Figure 4.2) and deep-supervision [232] are presented.

### 4.4.1 Simulations

#### 2D dataset

To compare Sparse Tensor Neural Networks with their dense counterpart, we based our study on a previously introduced dense method [45] and used the same 2D dataset based on the previously published simulation pipeline. Microbubble flow was simulated using a realistic model [124] based on *ex vivo* mice brains obtained with two-photon imaging. Four portions of different mice brains were used to generate the training set, one other portion

was used for model selection and validation and the last one was kept as a test set to assess the performance. This approach ensures that the test set comprises microbubble trajectories from a completely unseen mouse brain, allowing us to evaluate the model inter-individual generalization. Since each region covers a volume of only $500 \times 500 \times 500 \mu\text{m}^3$, we dilated the vascular network by a factor of 2 to fill a $1000 \times 1000 \times 1000 \mu\text{m}^3$ area, as done previously [45]. The ultrasound signal corresponding to the microbubble position was simulated using an in-house GPU implementation of SIMUS [131] with parameters corresponding to an L22-14 probe (Vermon, Tours). Three 15.625-MHz, tilted plane waves with angles of $-1°$, $0°$, and $1°$ were simulated with a frame rate of 1 kHz. The simulated signal was subsampled to match the 100% bandwidth IQ signal, mimicking the Verasonics Vantage system. Finally, the IQ data were beamformed with a GPU implementation of the delay and sum algorithm on a grid of $32 \times 32$ pixels with a resolution of $\frac{\lambda}{4}$ (i.e.,$25\mu$m), in groups of 512 frames. The point spread function (PSF) of the system was simulated at the center of the grid and used to compute the local correlation between the beamformed IQ data with the PSF. The obtained correlation maps were used as the input for the different deep-learning models. In total, 2250 movies were generated for training, 250 for validation, and 500 for testing. The concentration of microbubbles simulated for the training and validation sets was set to 5 microbubbles per field of view (FOV), as done previously [45]. Several test sets were simulated based on the trajectories from the test angiogram with varying concentrations (1, 5, 10, and 20 microbubbles per FOV) also matching the test set from the previous study [45].

**3D Dataset**

The 3D dataset was obtained similarly but since 3D convolution filters have more than their 2D counterpart, additional microbubble trajectories were included to prevent 3D models overfitting. We divided the generated spatio-temporal samples in three groups: 3500 samples for training, 500 for the validation, and 2000 for testing. We dilated the vascular network by a factor of 8 to account for the larger wavelength and the coarser beamforming grid ($\frac{\lambda}{2}$). We simulated a 750-fps imaging sequence containing 5 angles ($\{-2°, 0°\}$, $\{2°, 0°\}$, $\{-1°, 0°\}$,$\{1°, 0°\}$, $\{0°, 0°\}$) emitted with a 7.8125-MHz center frequency using a matrix array with parameters matching a commercially available 8-MHz 2D matrix probe (Verasonics, WA, USA). The concentration of microbubbles simulated for the training, validation, and test sets was increased to 30 microbubbles (compared to 5 for the 2D-case) per field of view (FOV) given the additional dimension.

#### Additional test sets

To complement the evaluation on anatomically realistic dataset, we generated random trajectory datasets and simulated corresponding ultrasound signal based on similar imaging sequences. 500 movies were generated for each of the following concentrations 1, 5, 10, and 20 microbubbles per FOV in 2D. As there is no vascular network to reconstruct, the Dice was computed for each frame and averaged across the whole dataset. Since this metric measures the overlap between few microbubble trajectories, it is hereafter referred to as *trajectory Dice* and is expected to yield lower value than typical Dice between vascular networks.

We assessed the out-of-distribution generalization and robustness of the proposed method and baselines to additive Gaussian noise. At test time, random noise was added to the real and imaginary parts of the correlation map of the angiogram-based test set at 5 MB/FOV. The noise had a mean of 0, and the standard deviation was increased from 0.1 to 0.25 in steps of 0.05.

### 4.4.2 Model training and evaluation

#### Sparse Tensor Neural Network and 4D convolutions

After the *dense-to-sparse* operation, the sparse tensor containing the low-resolution signal was given as input to a Sparse Tensor Neural Network implemented using the Python library MinkowskiEngine [209]. For each intermediate layer, Sparse Tensor Neural Networks only apply their convolutions and activations on non-zero values, yielding another sparse tensor. Conventional operations used in CNNs are implemented in MinkowskiEngine, leading to a relatively straightforward translation of the model from dense to sparse format. To assess the benefits of sparse formulation, we converted the dense Deep-stULM architecture to a sparse formulation without additional change, this approach is designated as *Sparse Deep-stULM* hereafter. However, such dense architecture might not take most of the sparse tensor implementation. Pruning or cascaded learning could further improve the memory efficiency of the sparse formulation. Both of these additional modifications require intermediate supervision and are detailed in the following sections. The resulting models are also extended directly to 3D imaging with 4D convolutions to handle 3D+T tensors. 4D convolutions are directly implemented in the Python library MinkowskiEngine [209].

### Training procedure

For the 2D models based on Deep-stULM, the hyperparameters were set to the same value as in the original study [45]: the optimizer used was Adam [233] and the training was divided into two parts. During the first 150 epochs, the ground truth microbubble trajectories were dilated with a radius of 2 to stabilize the training. The initial learning rate was set to 0.1 and then decayed by a factor of 10 at the epochs 15, 45, 75, and 100. During the last 150, the ground truths were no longer dilated, and the learning rate was set to 0.001 at epoch 150 and then decayed by a factor of 10 at epochs 160, 200 and 250. We did not optimize the hyperparameters for the sparse formulation of Deep-stULM and used the same as the original study. The batch size was set to 8 for all the runs in 2D. For the 3D models, we also used the Adam optimizer with an initial learning set of 0.1. We trained the 3D networks for 20 epochs in total, and the learning rate was decayed by a factor of 10 at epochs 15 and 17. For the first epoch, the batch size was set to 2 to allow every configuration to fit in memory, then for the remaining epochs, the batch size was increased to 4. Deep-stULM was trained using Dice loss in both is sparse and dense formulation.

### Performance comparison with ULM and Deep-stULM

**Evaluation metric** To compare our results with the previously established method [45] and conventional ULM, we measure the overlap between the network prediction and ground truth using the Dice coefficient :

$$\text{Dice} = \frac{2 \times |\text{GT} \cup \text{Pred}|}{|\text{GT}| + |\text{Pred}|} \tag{4.9}$$

with GT being the projection of all the trajectories from the ground truth to the super-resolved grid and Pred the prediction of the network. Similarly to the previous study [45], the Dice values displayed use a Dice computed between the binarized angiograms (i.e., between the logical summation of all the microfilms from the test set).

**Conventional ULM** We also provide the results of a standard, non-deep-learning ULM method, described in [24]. Briefly, Gaussian fitting was used to localize microbubbles and the Hungarian method [234] was used for the tracking step. The number of detections in each frame was set to the optimal value based on the number of microbubbles simulated in the FOV (i.e., for the 5MB/FOV concentration, the number of detection would be set to 5). Note that this setting is ideal and may favor the conventional ULM. Indeed, in real applications, the exact number of microbubbles in the FOV is unknown.

**Deep learning baseline, mSPCN-ULM** We retrained the mSPCN-ULM architecture [46] on our datasets using original training parameters. We ensured that the loss had converged at the end of the training (60 epochs), and the learning rate was decayed by ten after 30 epochs. The output of the mSCPN-ULM was interpolated to match Deep-stULM output resolution. Since training with a Dice loss was unstable, mSCPN-ULM was trained using the original training loss [46] based on L1 distance between the target and the prediction, both convolved with a Gaussian kernel.

**Memory monitoring** We monitored the memory usage of the training using CometML and took the maximum value reached during the training of each method. As there is some stochasticity involved both in training and in the measurement of the memory, we used the average across 3 different runs and provided the standard deviations between each run for deep learning approaches. For the 3D dense formulation of Deep-stULM, it was not possible to train the model due to practical memory limitation. Therefore, we provide only an estimate of the memory usage. This estimate was based on scaling the 2D memory usage based on the increase of memory for the intermediate maps due to the addition of a spatial dimension. For mSPCN-ULM, the 3D memory usage was linearly extrapolated from memory usage with smaller batch size.

**Computation time measurements** Processing times for one group of 512 frames were measured on NVIDIA V100 Volta (32G HBM2 memory) GPUs. To account for I/O latency on computation servers, processing times were measured and averaged across 5 redundant forward paths. Measured times for multiple samples were averaged to provide robust estimates. For 3D inference times, mSPCN-ULM could be implemented by simply translating PyTorch 2D operations to 3D operations. However, the whole group of 512 could not fit in memory. We measured and linearly extrapolated the processing times for smaller subsets of 16, 32, and 64 frames to estimate the processing times of 512 frames. 3D Dense formulation of DeepST-ULM was not implemented as it would have required 4D convolutions not supported in PyTorch.

### 4.4.3 Additional studies

#### *Dense-to-sparse* strategies

To assess the loss of information and its impact on the performance caused by the *dense-to-sparse* filtering, we compared the performance of the sparse model for two simple *dense-to-sparse* strategy referred as Top-k and thresholding strategy with varying value for their

respective parameters. To provide better intuition on the performance that one can expect with more sophisticated filtering, we developed a deep learning based solution. To compare between each method, we computed the average number of non-zero pixels in the test set movies to compare across methods and plotted it in Figure 4.6. In addition, to differentiate between the performance loss induced by the *dense-to-sparse* strategy from the effect of the sparse implementation, we also evaluated a dense model with inputs filtered according to the *dense-to-sparse* strategy (each model was retrained on the filtered data).

**Top-k strategy** Typical ULM approaches use regional maxima for microbubble detection before localizing them with high precision [67]. Based on the same underlying assumption that the microbubble signals are located near local maxima, it is reasonable to consider only the k-largest pixel of each input tensor. This operation is designated as top-k operation later on. In practice, due to the smoothness of the input, this approach is very similar to the use of local maxima value while providing better control of the memory usage of the input tensor. The explored values used for the top-k approaches range between 5000 to 50000 pixels.

**Thresholding strategy** Previously published deep-learning methods [45] used a threshold based on the value of the local correlation between the signal and the point spread function of the imaging system to remove residual after clutter filtering *in vivo*. We applied this same approach directly on our simulated datasets. For the thresholding approaches, the threshold values were set to $\{0.01, 0.05, 0.10, 0.25\}$. As the thresholding strategy with a threshold set at 0.10 in 2D offered a good trade-off between performance and sparsity, we used it for all the experiments where the *dense-to-sparse* strategy was not specified. The threshold was heuristically set to 0.25 for the 3D experiments.

**Deep learning based strategy** We trained a dense CNN to localize microbubbles at low resolution. To do so, it is trained to predict the presence of microbubbles in every pixel of the beamforming grid (low resolution). The dense network used to localize microbubbles at low resolution is fully convolutional both in space and time direction and takes as input a tensor of shape $2 \times H \times W \times T$ in $2D$. The inputs channels encode the real and imaginary parts of the input signal. The input signal is composed of the local correlation of the beamformed IQ data with the simulated PSF of the imaging system. The spatial resolution was kept unchanged throughout the whole network. However, the temporal dimension was reduced by a factor of 2. The output of this network was then interpolated to match the correlation map dimension. The resulting mask is used to convert the correlation map to a sparse format, where only the pixel values with microbubbles are stored along with their coordinates. This

filtering network was trained using the Dice loss between its predictions and the ideal mask at low resolution obtained from the simulated microbubble position.

### Architecture modifications

Herein, we describe further experiments to refine the sparsity using pruning on the intermediate representation of the network along with deep-supervision and cascaded learning.

**Deep Supervision and pruning** Similarly to other architectures [44, 46, 126], DeepstULM [45] uses upsampling layers that preserve the sparsity of the upsampled tensor. This is sub-optimal as at finer resolution the sparsity of the trajectory is increased as depicted in Figure 4.2. To mitigate these issues, we used the previously introduced pruning operations [231], that aim to gradually remove the pixels where no microbubbles are detected (green pixels versus gray pixels in Figure 4.2). The feature maps are masked based on the output of an intermediate classifier. Consequently, the removed pixels are no longer considered in the following operation and their coordinates are not stored in memory. As depicted in Figure 4.2, we implemented pruning at every resolution level based on intermediate prediction. Since pruning requires the training of intermediate classifiers at each level, we trained them in a supervised fashion. These intermediate classifiers are trained using the same loss as the final loss and consist of pointwise convolution directly applied to the intermediate representation of the network. This form of supervision is similar to deep supervision [232] and can also serve as regularization and improve the performance of the network. These intermediate classifiers are required to perform pruning, as they provide the mask used to remove the less relevant pixels from the following operations.

**Cascaded learning** In the case of super-resolution, deep supervision also makes possible a certain form of cascaded learning inspired by [235]. Indeed, the intermediate classifiers can be trained sequentially: during the first phase of the training, only the first classifier is trained to predict the presence of microbubbles on a grid at the input resolution. Then, during the following phase, the intermediate classifiers corresponding to higher resolution levels are sequentially added to the global loss. When applied, the cascaded learning strategy used one epoch for each intermediate level, and the number of epochs for the last resolution level was the same as in standard training.

Table 4.1 Comparison of the memory usage, inference time and angiogram reconstruction performance. * Value estimated.

| | 2D (5MB/FOV) | | | 3D (30MB/FOV) | | |
|---|---|---|---|---|---|---|
| | GPU Inference time (ms) | Memory requirements (GB) | Dice (%) | GPU Inference time (ms) | Memory requirements (GB) | Dice (%) |
| Sparse Deep-stULM | $\mathbf{19 \pm 1.3}$ | $\mathbf{6.8 \pm 0.2}$ | $73.93 \pm 1.96$ | $\mathbf{70 \pm 13.8}$ | $\mathbf{9.9 \pm 0.1}$ | $\mathbf{49.97 \pm 1.79}$ |
| Deep-stULM | $34 \pm 0.8$ | 12.6 | $\mathbf{80.01 \pm 1.76}$ | N.A | 694* | N.A |
| mSPCN-ULM | $77 \pm 0.9$ | 22.7 | $62.54 \pm 0.30$ | $8016 \pm 533.2$* | 366* | N.A |
| ULM | N.A | N.A | 60.80 | N.A | N.A | 12.34 |

## 4.5 Results

### 4.5.1 Processing time, memory reduction and performance comparison in 2D

The memory usage results and performance at 5 MB/FOV are reported in Table 4.1 for Conventional ULM, Deep-stULM, mSPCN-ULM and Sparse Deep-stULM. The threshold strategy was used for the sparse method to convert the input to sparse formulation. Sparse formulation reduced the memory usage of Deep-stULM from 12.6 GB to 6.8 GB during training, the processing time from 34 ms to 19 ms, while the Dice decreased from 80.1% to 73.9%. In comparison, mSPCN-ULM, with 4 times more processing time and with nearly four-fold higher memory usage in training, obtained a lower Dice of 62.54%, yet slightly outperforming conventional ULM (Dice of 60.8%). The results of the different methods in 2D under varying concentrations are shown in Figure 4.3, and the evolution of the Dice value computed is reported in Figure 4.4a along with the results on random trajectory datasets with varying concentrations in Figure 4.4b. Qualitatively, the standard ULM performances degraded as the concentration increased with degradation in resolution starting from 5 microbubbles per field of view (MB/FOV). Quantitatively, this diminution of performance was highlighted by a drop in Dice coefficient from 71.4% with 1 MB/FOV to 60.8% for 5 MB/FOV on angiogram reconstruction. The performance continued to decrease with the concentration until it reached a Dice of 55.9% at 20 MB/FOV, with only the biggest vessels being visible. In contrast, the dense Deep-stULM approach exhibited a smaller performance degradation from 83.9% for 1 MB/FOV to 73.7% for 20 MB/FOV. The sparse formulation of Deep-stULM showed robustness to increased concentration and reached performance levels at high concentration (with a Dice coefficient of 70.6% (resp. 68.0%) for 10 (resp. 20) MB per FOV that were very close to its performance level at low concentration (74.4% for 1MB per FOV). However, the performances at low concentration (1MB per FOV) were lower than Deep-stULM (83.9%) but were comparable to conventional ULM (71.4%). In opposition, mSPCN-ULM performance increased with the concentration from 54.9 at 1MB/FOV to 67.4

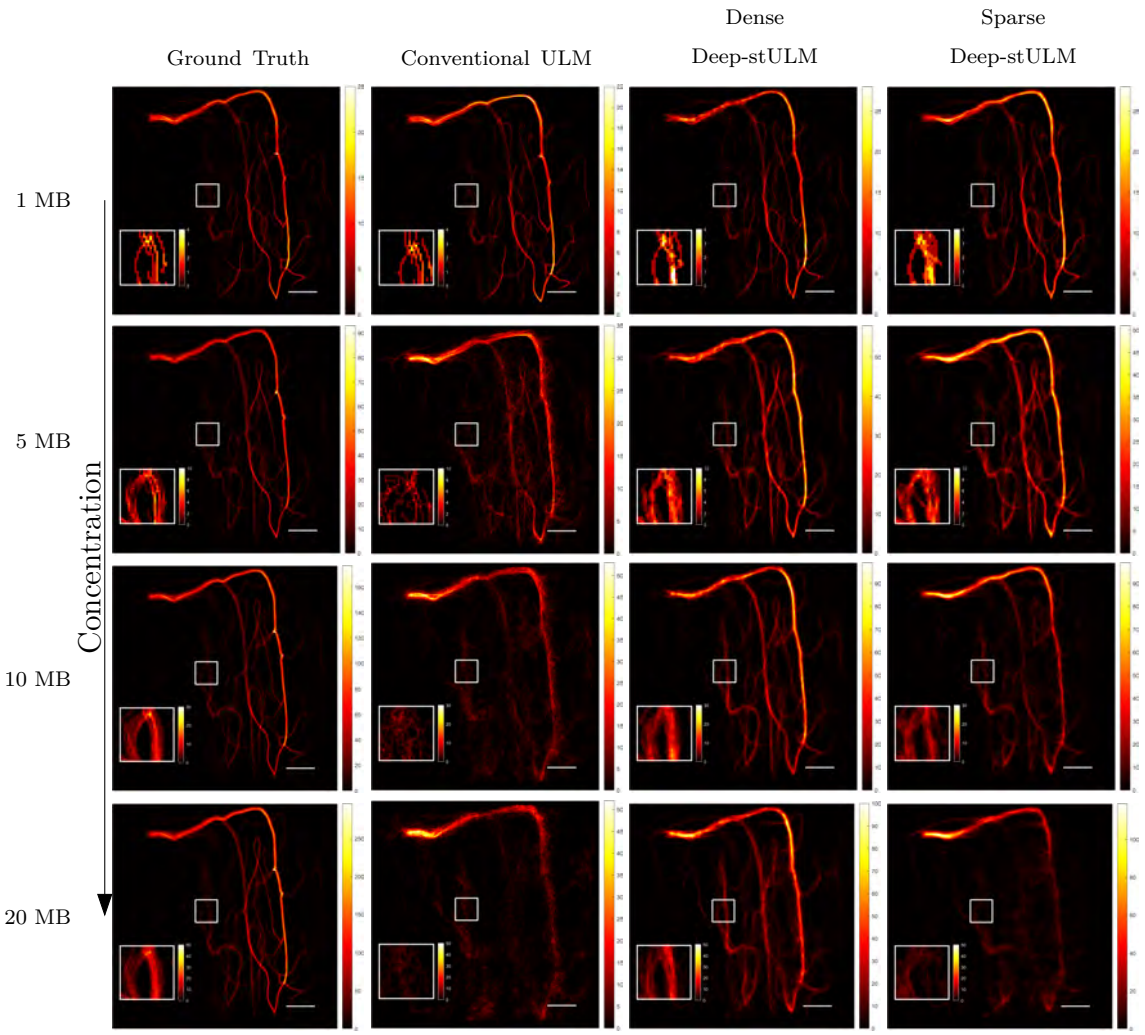


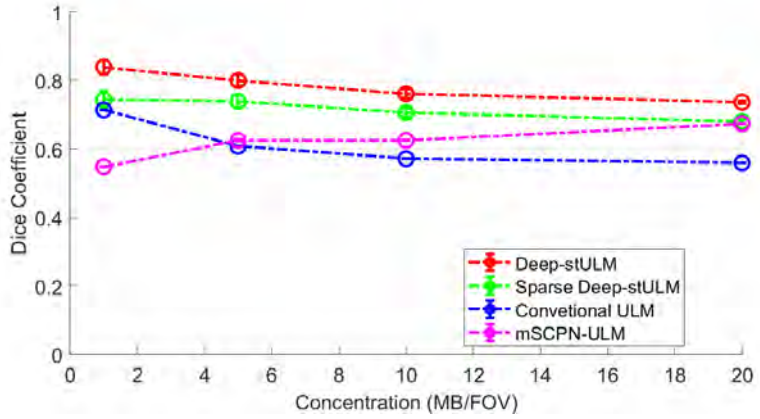

Figure 4.3 Comparison of performance under increasing concentration for conventional ULM (center left column), Deep-stULM dense formulation (center right column) and its sparse formulation (right column). Ground truth is given for comparison (left column). The scale bar is 98µ and corresponds to the wavelength of the simulated pulse. Concentration increases from 1 (top row), 5, 10, and 20 (bottom row) microbubbles per field of view.

at 20MB/FOV, reaching a level of performance similar to sparse and dense Deep-stULM and outperforming conventional ULM. Evaluation on random trajectory datasets in Figure. 4.4b showed a similar trend for Conventional ULM and sparse and dense Deep-stULM. However, for mSPCN-ULM the *trajectory Dice* slightly decreased from 35.5% at 1MB/FOV to 27.7% at 20MB/FOV outperforming the other methods.

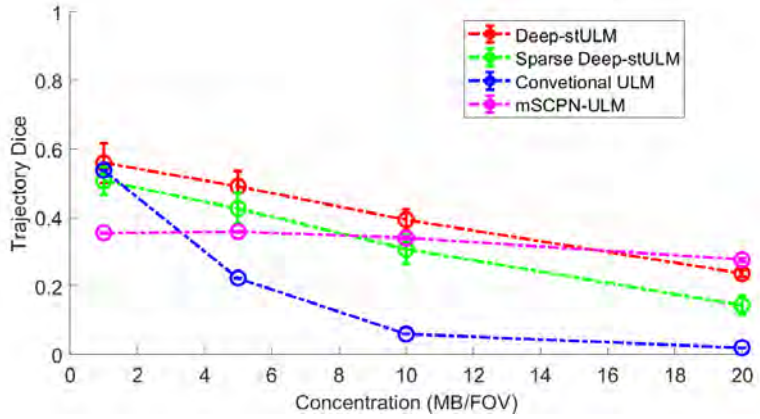
As seen in Figure 4.4c, evaluation under additive noise at test time showed a decrease in performance for all the methods but the sparse formulation of Deep-stULM, which performed similarly at all noise levels (from 73.9 for $\sigma = 0.1$ to 73.1 for $\sigma = 0.25$).

### 4.5.2 3D feasibility study

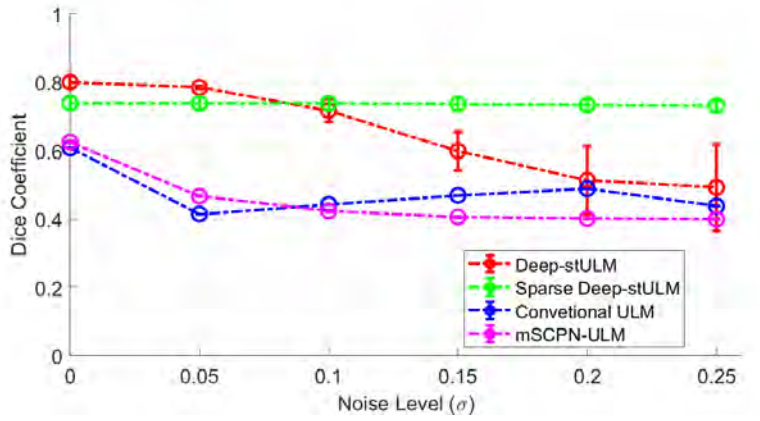
In Table 4.1, we display the results and the memory usage for Sparse Deep-stULM and conventional ULM as well as estimated memory usage in 3D for Deep-stULM, mSPCN-ULM. We also provided inference times for Sparse Deep-stULM and mSPCN-ULM, showing that Sparse Deep-stULM was more than two orders of magnitude faster than mSPCN-ULM. The reported Dice are computed in 3D, which means that the angiograms are sparser than in 2D and leads to values for the Dice in 3D that are typically lower. To provide comparative insights on the 3D Dice values, we projected to 2D the ground truth and reconstructed 3D angiograms and measured their Dice in 2D. The 2D projected Dice values were typically higher and comparable to the 2D results. Indeed, conventional ULM reached a projected Dice of 67.56% for 1MB/FOV (close to the 71.4% obtained in 2D with the same concentration) and 54.58 for 30MB/FOV (close to the 55.9 for 20MB/FOV in 2D). Projected Dice values for Sparse Deep-stULM were 77.71% for 1MB/FOV and 77.81% for 30MB/FOV. In addition, since the simulation parameters between the 2D and 3D datasets are different to match realistic imaging sequences, the 3D PSF is larger and has important side lobes, which makes the localization process more challenging. It is important to note that just using the sparse formulation allowed us to train the network with less than 11GB of GPU memory while outperforming conventional ULM (50.0% versus 12.3%). For qualitative analysis, the reconstructed angiograms from the test set are displayed in Figure 4.5 with concentration increasing from 1 MB/FOV to 30 MB/FOV. At high concentration (10 MB/FOV and 30 MB/FOV), the sparse model accurately reconstructed the angiogram when conventional ULM failed to do so. Indeed, the conventional ULM produced many false detections that were not present in the sparse model reconstruction. At low concentration (1 MB/FOV), both conventional ULM and sparse model reconstructed the angiogram with fidelity.




(a) Angiogram reconstruction Dice across increasing concentrations.




(b) Trajectory detection Dice across increasing concentrations.




(c) Performance as a function of additive noise standard deviation.

Figure 4.4 Performance study under varying conditions for the Sparse Deep-stULM model.

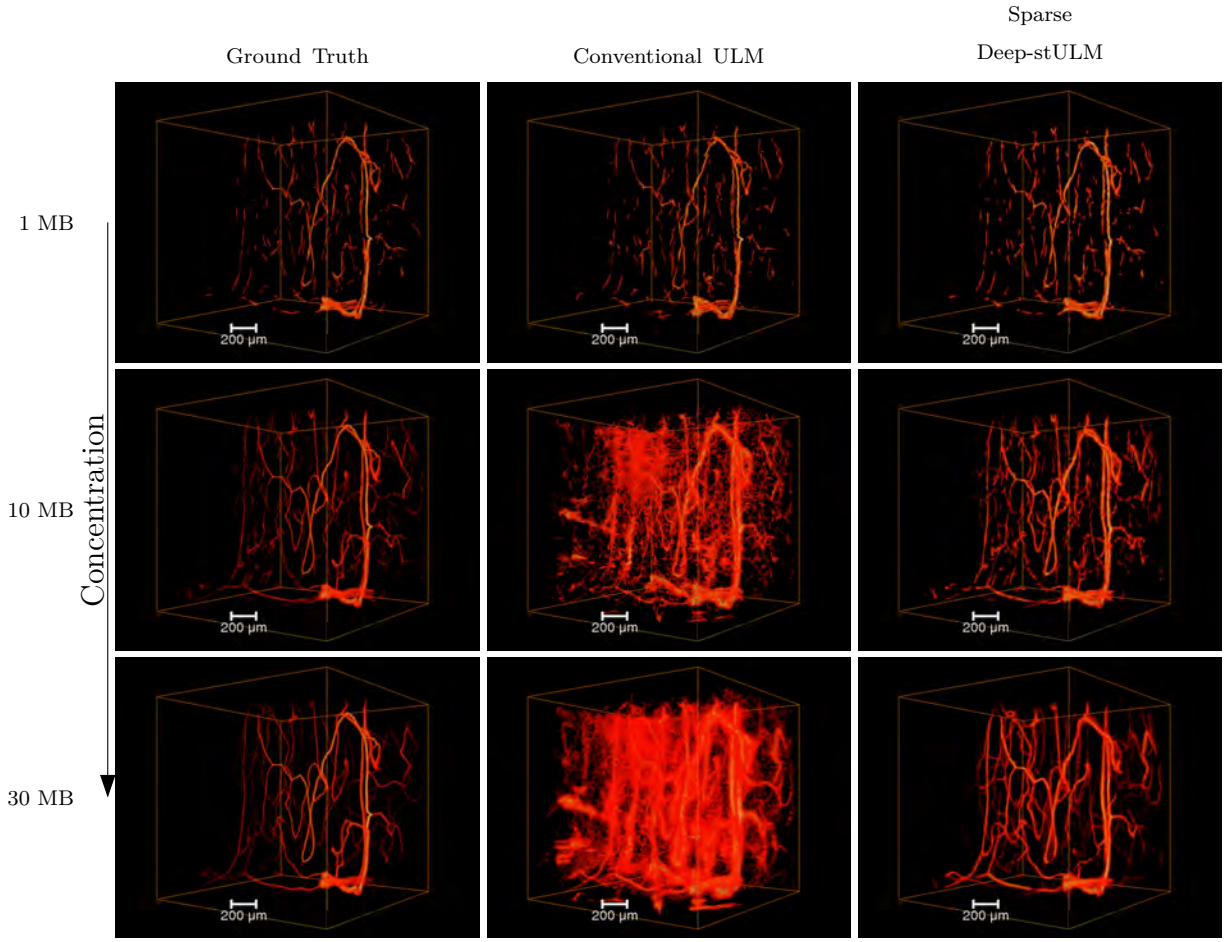



Figure 4.5 3D Comparison of performance under increasing concentration for conventional ULM (center column) and Deep-stULM sparse formulation (right column). Ground truth is given for comparison (left column). Scale bar is 200 $\mu$m and corresponds to the wavelength of the simulated pulse. Concentration increases from 1 (top row), 10 and 30 (bottom row) microbubbles per field of view.

### 4.5.3 Additional studies

#### Evaluation of strategies to convert to sparse formulation

In figure 4.6, it is observed that the thresholding and top-k *dense-to-sparse* approaches reached a very similar trade-off between the sparsity obtained and the level of performance independently of the formulation of the model (dense or sparse). Indeed, for the sparse model with the thresholding *dense-to-sparse* strategy the Dice varies from 67.5% with around 21000 pixels to 75.3% with 90000 pixels while it varies from 65.4% with around 5000 pixels to 73.9% with 100000 pixels for the top-k strategy. Similarly, for the dense model with the thresholding *dense-to-sparse* strategy the Dice varies from 72.3% with around 11000 pixels to 79.9% with 64000 pixels while it varies from 69.6% with around 5000 pixels to 79.5% with 100000 pixels for the top-k strategy. Finally, the CNN *dense-to-sparse* operation yielded a better trade-off than all the other approaches. Indeed, the CNN *dense-to-sparse* operation reached a Dice of 67.8% with only 1400 pixels.

*Dense-to-sparse* strategy failure cases are displayed in Figure 4.7. These failure cases were selected when the threshold strategy, with the threshold set to 0.1, removed at least one pixel containing a microbubble and accounted for between 5% and 10% of the total test set frames. For the threshold and top-k strategies, some low-intensity microbubbles in the vicinity of other microbubbles were filtered out, especially when the peak of the correlation map did not match the microbubble's position. These microbubbles were not detected by the CNN masking strategies, which also made localization errors, even for clearer microbubbles.

#### Impact of architecture modifications

In table 4.2, we observed that using pruning jointly with sparse formulation led to a decrease in memory requirements in 2D (6.9 GB to 5.6 GB) but also led to an important degradation of the performance (9.6% of Dice). It appeared that combining pruning and cascaded learning has a small impact on the memory (5.6 GB versus 5.7 GB) while degrading the performance (more than 3% of Dice). The addition of intermediate loss for deep supervision did not benefit the training performance and had a small memory cost. In 3D, every variation of the model trained significantly outperformed the conventional ULM (between 38% and 50% for the sparse models and 12.3% for the conventional ULM). In contrast to 2D, the use of pruning reduced memory usage by a factor of approximately 4. However, similarly to the 2D case, pruning also degraded the performance (50% to 38.6.%).

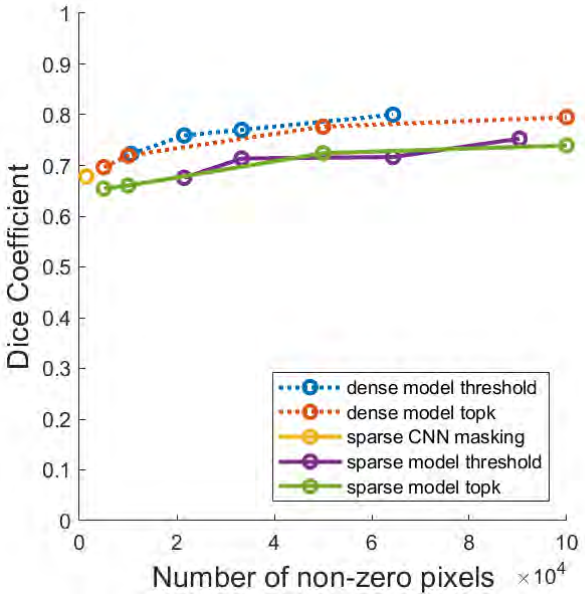



Figure 4.6 Evolution of performance as a function of the sparsity achieved for different *dense-to-sparse* operations and comparison with dense network with masked input as well as CNN-based dense to sparse operations.

## 4.6 Discussion

In this work, we studied the impact of using a sparse formulation on the memory usage and performance of Deep-stULM. We also investigated more multiple methods to further increase sparsity and reduce the memory usage. Our results suggest that solely using the sparse formulation allows for the extension of existing deep learning architectures to 3D ULM, while preserving the performance robustness at high concentration. However, more complex modifications to the architecture had a less important impact on the memory usage while degrading the performance.

### 4.6.1 Reducing memory usage of existing methods in 2D

The sparse formulation offers a relatively simple approach to divide by nearly two-fold the memory requirement of an existing architecture in 2D, while outperforming conventional ULM. Indeed, the sparse formulation reconstructed quantitatively more precise angiograms for high concentration (5MB/FOV and 10 MB/FOV). Dense deep learning approaches consistently outperformed the sparse formulation. Leveraging larger and more complex datasets could help mitigate the performance degradation, while still benefiting from the computation time and memory improvements. Combining the shorter acquisition time with reduced

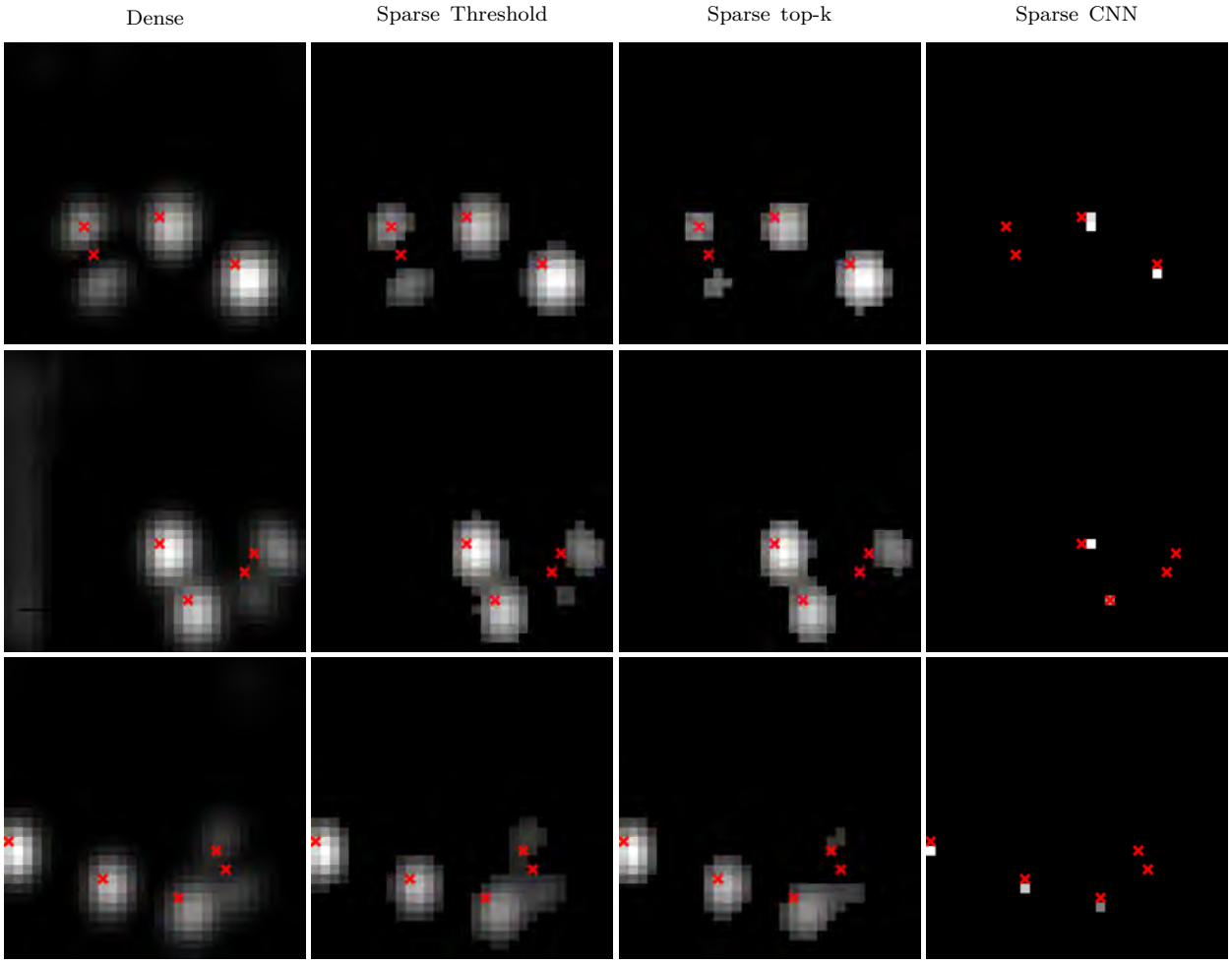



Figure 4.7 Failure cases of the *dense-to-sparse* thresholding strategy with a 0.1 threshold (center left column) along with the dense correlation map (left column), the results of the top-k with 50000 values (center right column), the results of the CNN based strategy (right column)

Table 4.2 Comparison of the memory usage and performance of the different additions to Sparse DeepST-ULM architecture.

| Ablation settings | | | | 2D (5 MB/FOV) | | 3D (30 MB/FOV) | |
|---|---|---|---|---|---|---|---|
| Sparse input | Intermediate loss | Pruning | Cascaded learning | Memory (GB) | Dice (%) | Memory (GB) | Dice (%) |
| ✓ | ✓ | ✓ | ✓ | $\mathbf{5.6 \pm 0.2}$ | $59.22 \pm 0.49$ | $3.7 \pm 0.4$ | $38.53 \pm 0.27$ |
| ✓ | ✓ | ✓ | | $5.7 \pm 0.1$ | $62.36 \pm 1.57$ | $\mathbf{3.6 \pm 0.2}$ | $38.65 \pm 0.58$ |
| ✓ | ✓ | | | $6.9 \pm 0.1$ | $71.98 \pm 0.16$ | $10.6 \pm 0.3$ | $43.51 \pm 0.99$ |
| ✓ | | | | $6.8 \pm 0.2$ | $\mathbf{73.93 \pm 1.96}$ | $9.9 \pm 0.1$ | $\mathbf{49.97 \pm 1.79}$ |

constraints on computation, the proposed sparse formulation could be relevant in near real-time processing of ULM acquisition. The robustness of the sparse formulation to additive noise suggests that the information loss introduced by the *dense-to-sparse* operation and the resulting drop in performance are primarily due to features that are sensitive to noise.

### 4.6.2 Scaling to 3D imaging

In 3D, the sparse formulation allowed for training, with less than 11 GB of memory, an architecture that would require close to 700GB of memory in dense formulation. As expected, the sparse formulation impact on memory was greater in 3D than 2D. Experimentally, the ratio between the sparse and the dense formulation was multiplied by $\delta_{exp} = 2.6 \times 10^{-2}$ when extending from 2D to 3D whereas the theoretical value from Eq. 4.8 was $\delta = 3.1 \times 10^{-2}$. The smaller ratio could be explained by the aforementioned experimental factors such as temporal context considerations, as well as the higher threshold for the *dense-to-sparse* strategy in 3D. The performance gap between the sparse formulation of Deep-stULM and the conventional ULM was larger in 3D, suggesting an even higher potential for deep learning based approach in 3D ULM. On the one hand, the drop of performance of conventional ULM might be caused by the important side lobes of the simulated PSF in 3D. Indeed, these sidelobes create false detection which are supposed to be filtered during the tracking step based on the track length. However, in high concentration, candidate detections in successive frame are multiple causing the filtering on track length to be less effective and contribute to 3D conventional ULM map with a high number of false positives. On the other hand, it is also possible that the additional dimension increases the capability of the deep learning approach to distinguish crossing trajectories, leading to better performance relatively to conventional ULM. The results in 3D show that akin to the dense approach in 2D, sparse models can accurately reconstruct angiograms at concentration where conventional ULM is failing. When translated in vivo, such performance in high concentration would allow for

reduced acquisition time. In 3D, where dataset tends to be larger, reducing acquisition time is even more crucial as it also reduces the storage needed and the associated transfer time.

### 4.6.3 Further reductions of memory and performance trade-off

#### *Dense-to-sparse* strategies

The parameter studies showed that top-k and thresholding *dense-to-sparse* strategies performed similarly, and a better trade-off could be achieved with deep-learning based *dense-to-sparse* operations. However, this approach lacked flexibility in the trade-off between sparsity achieved and performance, which translated to very restrictive filtering and poor overall performance. Top-k and threshold *dense-to-sparse* strategies still filtered out some microbubbles signal, presumably altering the localization performance. Local misalignment between the correlation map peak intensity and the localization of the microbubble caused all the proposed *dense-to-sparse* strategies to fail. Such mismatches are expected to occur when microbubbles are closer than the size of the PSF used for correlation. Removing the correlation step and directly using beamformed IQ could improve the performance of *dense-to-sparse* strategies. Additionally, increasing the number of pixels considered in the sparse formulation reduced the performance gap between sparse and dense formulation of Deep-stULM, which also suggested suboptimal *dense-to-sparse* operations. The sparse formulation might benefit from more sophisticated *dense-to-sparse* strategies to further improve the proposed trade-off between performance and memory requirements in training.

#### Architecture modifications

It appears that architecture modifications such as deep-supervision, cascaded learning and pruning had an important impact on the performance with smaller gain in memory than that then sole sparse formulation. Cascaded learning and deep-supervision negative impact on the performance could be explained by the fact that microbubble detection at coarse resolution is different from localization at super-resolution and therefore intermediate losses enforce the learning of less relevant latent representations. Considering the additional complexity in training and architecture constraint going with this modifications, it is less clear that they present an interesting trade-off between memory usage and performance.

### 4.6.4 Limitations and perspectives

The proposed method demonstrates that sparse tensor neural networks can extend the benefits of deep learning based approach in ULM from 2D to 3D imaging by improving the scaling

law of memory costs with dimensionality and resolution. However, some limitations should be mentioned and could be addressed by future studies.

### Problem formulation

This study is mostly focused on measuring the impact of using a sparse formulation in deep learning for ULM. For this reason, it does not tackle some key challenges in framing the learning problem. On the one hand, it is important to mention that even though the Dice loss has been proposed and successfully applied in the previous study [45], it is unstable and lacks smoothness when working with temporal projection of trajectories. In addition, the final representation of the prediction, being a projection of the microbubble trajectories on a grid, does not offer the same liberty for downstream analysis as the individual detections provided by conventional ULM. A formulation tackling these issues has been recently proposed [195] based on the DECODE method [191] and could be worth investigating. On the other hand, the encoding of the real and imaginary parts of the input signal as channels lacks the proper arithmetic of complex numbers. Using a complex value neural network [187] could allow a better representation of the signal with less overfitting and better overall performance. Such networks have shown interesting potential when dealing with ultrasound data [69, 188].

### Validity of the *in silico* model

The proposed approach reaches the level of performance of the state-of-the-art dense model *in silico* under varying concentrations of microbubbles. It would be interesting to evaluate the validity of the conclusions on real data as it has not been tested *in vivo*. As current training datasets for ULM are limited in their diversity and realism, they did not allow for direct domain transfer to *in vivo* of both dense and proposed methods. Consequently, *in vivo* applications required carefully tuned pre-processing steps, which were not consistently successful. Due to the preprocessing inconsistency between training and testing, it is not clear that better learning ability on the simulated dataset would lead to improved performance *in vivo*. Furthermore, as the main benefit of the proposed method is an improvement of the scaling law of the memory complexity, it is reasonable to assume that it is still valid *in vivo*.

## 4.7 Conclusion

In this study, we studied the potential of sparse formulation when applying deep learning in 3D ULM. We also proposed further optimization of memory efficiency through pruning of the 3D ULM deep learning model. The proposed Sparse Deep-stULM method successfully

improve the scaling of memory requirements of deep learning based approaches, addressing the challenge of their extension in 3D. While it comes at a small cost of performance in 2D, the use of deep-learning in 3D ULM seems to be even more beneficial than in 2D. To the best of our knowledge, it is the first application of deep learning for 3D ULM, and it could pave the way for further approaches both *in silico* and *in vivo*. Further applications of such models could allow for improved architectures capable of fitting to more diverse datasets, yielding better results both in 2D and 3D ULM or DynULM.

# CHAPITRE 5 ARTICLE 3 : TEACHER-STUDENT MODELS FOR ROBUST IN VIVO DEEP-LEARNING IN ULTRASOUND LOCALIZATION MICROSCOPY

Le chapitre précédent a permis de lever une limitation technologique majeure concernant le passage à l'échelle de l'apprentissage profond pour l'ULM 3D. En exploitant la parcimonie des données, nous avons démontré qu'il est possible de traiter des volumes importants de données avec une complexité mémoire maîtrisée.

Cependant, au-delà de l'architecture des réseaux, une limitation fondamentale persiste dans la littérature actuelle : la dépendance aux simulations pour l'entraînement des modèles. Comme souligné dans l'introduction générale de cette thèse, l'écart de distribution (*domain shift*) entre les données synthétiques et la réalité expérimentale — marquée par le bruit, les aberrations et la complexité des tissus — contraint souvent la généralisation des modèles lorsqu'ils sont appliqués *in vivo*.

Dans ce troisième chapitre, nous abordons directement cette problématique en proposant un changement de paradigme dans la stratégie d'apprentissage. Nous présentons une méthode fondée sur une architecture *Teacher-Student*, capable d'apprendre exclusivement à partir de données *in vivo*, sans aucun recours à la simulation. En utilisant des pseudo-labels générés par des méthodes conventionnelles et en combinant des stratégies de distillation et de perturbations du signal d'entrée, cette approche vise à surpasser les méthodes classiques, notamment en conditions d'imagerie dégradées.

Ce chapitre correspond au troisième article de cette thèse, soumis le 13 décembre 2025 à *IEEE Transactions on Medical Imaging*. Il établit le cadre méthodologique nécessaire pour rendre l'apprentissage profond robuste et applicable aux conditions cliniques et précliniques réelles.

# Teacher-Student models for robust in vivo deep-learning in Ultrasound Localization Microscopy

Brice Rauby[1,2], Alexis Leconte[1], Alice Wu[1], Gerardo Ramos-Palacios[3], Stephen A. Lee[1], Jonathan Porée[1], Abbas F. Sadikot[3], Maxime Gasse[4,1,2], Jean Provost[1,5]

[1]Department of Engineering Physics, Polytechnique Montréal, Montréal, QC, Canada
[2]Mila – Quebec Artificial Intelligence Institute, Montréal, QC, Canada
[3]Montreal Neurological Institute, McGill University, Montreal, QC, Canada
[4]Microsoft AI, Montréal, QC, Canada
[5]Montreal Heart Institute, Montréal, QC, Canada

## 5.1 Abstract

Ultrasound Localization Microscopy (ULM) enables microvascular imaging beyond the acoustic diffraction limit by localizing and tracking clinically-approved individual microbubbles injected in the bloodstream. However, conventional ULM remains constrained by long acquisition times, sensitivity to aberrations, and high hardware and computational costs. Deep-learning approaches have shown promise for faster and more robust reconstruction, but most rely on simulated data for training, introducing domain gaps that hinder *in vivo* performance. In this work, we present a framework for training and adapting deep-learning models directly from *in vivo* data using pseudo-labels obtained from conventional ULM. We construct and release a large-scale dataset of labeled training patches from transcranial mouse acquisitions (39 mice, 68 acquisitions) and introduce an input-perturbation and self-distillation strategy to adapt our proposed approach Teacher-Student-ULM (TS-ULM) to challenging conditions. In ideal condition, TS-ULM performs similarly to conventional ULM baseline with more complete reconstruction than gaussian fitting (saturation $42 \pm 10\%$ v.s. $38 \pm 10\%$) and better spatial coherence than Tracking-and-Localization (FRC $26.5 \pm 6\mu m$ vs $28 \pm 7\mu m$) In strong noise conditions, TS-ULM shows improved robustness and markedly outperforms conventional ULM (FRC $29 \pm 6\mu m$ vs $32 \pm 9\mu m$, saturation $30 \pm 10\%$ vs $33 \pm 8\%$). When the number of receive channels is reduced from 128 to 32, preserving similar resolution and improving saturation. These results demonstrate the feasibility of fully *in vivo* learning and that

TS-ULM enables robust imaging with reduced hardware requirements. Future extensions of this work could help overcome current hardware constraints in 3D ULM, where the number of active channels remains a major limitation.

## 5.2 Introduction

Ultrasound Localization Microscopy (ULM) enables in vivo microvascular imaging at depth with a spatial resolution well below the acoustic diffraction limit [13, 14]. By localizing individual microbubbles injected into the bloodstream, ULM provides super-resolved maps of the vascular network that remains otherwise inaccessible with other non-invasive imaging modalities [11]. Novel acquisition sequences and retrospective gating strategies have further extended ULM to the mapping of dynamic quantities, enabling applications such as pulsatility imaging in the brain [24, 25, 84], cardiac imaging [30, 78], and functional imaging [2, 26]. Despite these advances, conventional ULM remains hindered by its sensitivity to tissue motion, the presence of aberrations in challenging imaging conditions [8], and the long acquisition times required to observe microbubbles throughout the vascular network [32]. In addition, the complexity of the experimental protocol and the computational burden of processing large data volumes limit robustness, reproducibility, and broader translation [8, 49]. Deep learning approaches have been proposed to mitigate several of these challenges [49]. Data-driven denoising and clutter filtering have been introduced to improve the signal-to-noise ratio [133,148], while complex-valued neural networks have been explored for aberration correction, particularly in brain imaging [69]. Simplified learning pipelines have also been applied to reduce the computational burden and simplify the reconstruction process [46].

Deep learning strategies for ULM have mainly focused on improving localization performance at high microbubble concentrations [1, 44–46, 49, 125, 126]. These approaches generally rely on supervised learning using the known ground-truth positions of microbubbles provided by simulated datasets. However, training exclusively on simulations introduces a domain shift that can degrade performance in vivo [1, 49]. Early works relied on uniformly distributed scatterers [44, 46], whereas more recent studies introduced structured vascular priors [126] and realistic flow simulations to better mimic *in vivo* conditions [45, 49, 137]. While these models improve realism, they do not eliminate the domain gap with experimental data. Increasing simulation specificity can even reduce generalization, as networks trained on particular vascular geometries or flow regimes may not transfer to acquisitions with different anatomy or hemodynamics [49]. Simulation frameworks such as Field II [127] and SIMUS [131] assume linear propagation, offering computational efficiency but neglecting nonlinear effects. Extensions combining nonlinear microbubble dynamics [141] or full-wave

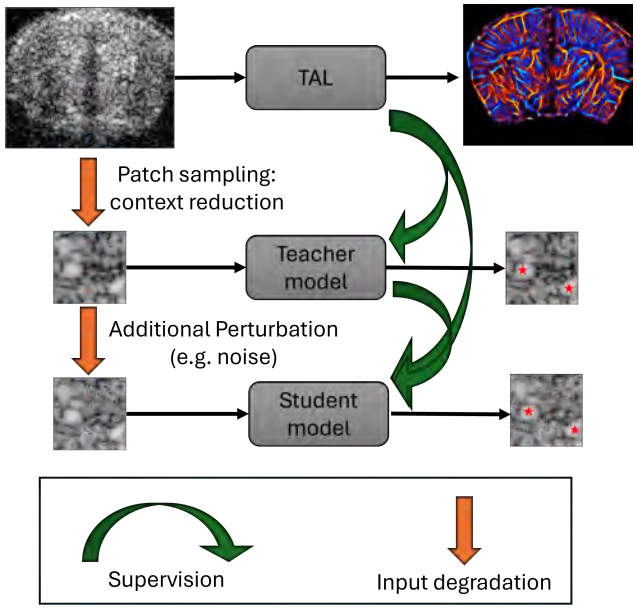



Figure 5.1 Representation of the *Teacher-Student-ULM* (TS-ULM) framework. a) First, a dataset is generated from 6.8 million patches that were extracted based on detections from the previously published TAL method [74]. b) The TS-ULM model is then trained based on the patches and their label. c) The TS-ULM model can be further enhanced by finetuning with knowledge distillation under input perturbation.

solvers like k-Wave [140] improve physical fidelity at higher computational cost [125, 236]. Additional noise models—Gaussian, Rician, or speckle-based—are often applied to bridge the gap with experimental data [2, 69, 126].

Despite these refinements, domain gaps between synthetic and experimental data remain a key challenge, and deep learning methods in ULM have shown limited applications *in vivo* despite their strong performance in simulation. Recent efforts have sought to bridge this gap, either through generative or self-supervised learning strategies that exploit *in vivo* signals [2, 119], or by combining simulated microbubble echoes with experimentally acquired data *in vitro* [134]. However, the scarcity of accessible *in vivo* data and the need for extensive parameter tuning in conventional ULM pipelines have long hindered the creation of large, well-labeled datasets. As a result, no approach to date has relied exclusively on *in vivo* data for supervised training. Recently, robust ULM processing methods—most notably the Tracking-and-Localization (TAL) framework [74], combined with large-scale data collection initiatives [237], have begun to close this gap, enabling the acquisition and processing of extensive *in vivo* datasets suitable for deep learning.

In parallel, advances in knowledge distillation and teacher–student learning have offered new

ways to enhance model robustness. Knowledge distillation is a learning paradigm where a "student" model is trained to reproduce the behavior, outputs, or internal representations of a "teacher" model, effectively transferring knowledge from one network to another [238]. While often used for model compression, recent studies have shown that student models trained under input [239] or label perturbations [240] can generalize better than the teachers that generated their supervision. Inspired by these findings, we investigate whether a similar teacher–student framework can be leveraged to train deep learning models for ULM directly *in vivo*. In our formulation, the initial teacher corresponds to the conventional TAL method, while the student learns from pseudo-labels on localized spatiotemporal patches with an implicit perturbation arising from the training formulation and its limited spatial and temporal context. We then extend this concept through a self-distillation stage, where the pretrained model serves as an additional teacher: the student learns simultaneously from TAL pseudo-labels and from the teacher's internal feature maps, hereafter referred to as latent representations, while being exposed to perturbed inputs.

Our method combines a large-scale *in vivo* dataset with pseudo-labels obtained from robust ULM processing [74] and incorporates input perturbation and self-distillation to improve robustness under challenging imaging conditions. To the best of our knowledge, this work introduces both the largest publicly available training dataset for deep learning in ULM and the first deep learning approach trained exclusively on *in vivo* data. Our codes and training dataset are made available through TS-ULM Github repository

## 5.3 Methods

In this section, we describe our complete framework developed to train and adapt deep learning models for ULM using exclusively *in vivo* data. As illustrated in Fig. 5.1, our approach proceeds in three stages: we first detail the acquisition and processing steps required to construct a large-scale dataset of labeled spatiotemporal patches. Next, we describe the neural network architecture and the supervised training formulation used to learn the localization task. Finally, we present the evaluation protocols and the input-perturbation and self-distillation strategies designed to enhance model robustness and generalization across different experimental conditions.

### 5.3.1 Dataset constitution

To generate our labeled training dataset, we processed *in vivo* acquisitions acquired as part of different projects in our group. We repeated the following patch extraction process for

each ULM dataset. First, we processed the raw data with our standard ULM pipeline (see 5.3.1 for details) to identify microbubbles tracks. We also stored the beamformed In-phase/Quadrature (IQ) data during processing. We then sequentially sampled tracks with their corresponding frame of interest and their relative position inside a patch, with a patch size of $32 \times 32$ pixels corresponding to $8\lambda$. The sampling process also guaranteed that the relative position of the microbubble in the patch did not lie within $1\lambda$ of the patch borders and that each sampled patch contained at least one MB position. We also ensured to store all additional microbubbles position present in the patch. Overall, we extracted approximately 6.8 million patches from 68 different acquisitions on 39 different mice. The dataset was then randomly split between training (80%), validation (10%), and test (10%) sets, ensuring that patches from the same mouse were not shared between different sets. Finally, we saved the dataset in HDF5 format for efficient loading during training.

**Data acquisition**

The ULM acquisitions were performed transcranially on young wild type mice (aged $8 \pm$ 5weeks). The mice were anesthetized with isoflurane before injection of Definity microbubbles. The training set only included acquisitions made with a 128-element linear array transducer (L22-14v, Vermon, France), transmitting 3 cycles at its center frequency of 15.625 MHz. Plane wave acquisition was made with Vantage-256 imaging system (Verasonics, WA, USA). 11 tilted plane waves (from $-5°$ to $5°$, centered at $0°$) were compounded yielding a frame rate of 1000Hz. A more detailed description of the experimental procedure can be found in a previous study [84]. We also evaluated the proposed approach on ULM performed with a 128-element linear array transducer (L8-18iD, GE, IL, USA) transmitting at a center frequency of 10.4667 MHz, more details about the procedure have been published in another study [73].

**ULM processing**

Tracks used for patches generation and ULM vascular maps were computed with the same pipeline and parameters. IQ signal was filtered using a Singular Value Decomposition (SVD) clutter filter to recover MB signals before image formation (i.e., beamforming) by removing the 20 eigenvectors with the highest energy. Similarly to previous study [86], we used spatially dependent time-gain compensation (TGC) and lag-1 autocorrelation. MB tracks were extracted using Tracking and Localization (TAL) approach (with parameters $\tau = 0.6$, and $\sigma = 4$) and kept only the track longer than 20 frames, as it provided a robust approach to obtain high quality angiograms. Processing parameters were equally set for all the acquisi-

tions. For better comparison, we also computed ULM map using Gaussian fitting on local maxima of the correlation with the imaging system Point-Spread-Function (PSF) (with a maximum of 256 microbubbles per frame) before tracking microbubbles with the Hungarian Algorithm (with max linking distance of 5 and allowing a gap of 2 frames), and keeping only the microbubbles tracked for longer than 15 frames.

### 5.3.2 Deep Learning training

#### Model Architecture

We adopted the DECODE architecture introduced for Single Molecule Localization Microscopy [191] and later adapted to ULM [2]. DECODE consists of two cascaded U-Net modules [241]. The first stage independently encodes the frame of interest and its adjacent frames to capture short temporal context. Their encoded representations are concatenated and processed by a second U-Net that predicts the final localization map. The network outputs five channels of the same spatial dimensions as the input: one channel encodes the probability that a pixel contains a microbubble, while the remaining four channels represent sub-pixel offsets parameterized as the mean and variance of a 2D Gaussian relative to the pixel center. This probabilistic encoding enables sub-wavelength localization accuracy. Unlike previous implementations that interpolated the input grid to avoid grid artifacts [2], we preserved the native beamforming grid for both input and output, thus reducing computation cost. During inference, localizations with confidence above 0.5 were retained, and detected MBs were temporally linked using the Hungarian algorithm with identical parameters as the Gaussian-fitting baseline.

#### Training parameters

We trained the model during 50 epochs, with a batch size of 2048 samples, with Adam optimizer. We used the same objective loss as previous applications of DECODE in ULM [2]. Patches of beamformed IQ were used with a dimension of $32 \times 32$ and the real and imaginary part were stored as channels. The training was monitored using CometML, implemented using Pytorch Lightning library in Python, and performed on compute nodes with GPUs (NVidia H100 SXM5) The training was performed in 16bit-mixed precision.

### 5.3.3 Evaluation

In this section, we describe the different experiments implemented to evaluate the performance and the limitations of the proposed approaches. For each experiment, we measured

the Fourier Ring Correlation (FRC) [76] and saturation [67] as complementary metrics for the quantitative analysis, and also visually inspected the reconstructed angiograms for qualitative reporting. First, we focused on evaluating the model performance under independent and identically distributed (i.i.d.) condition, which is when the training and testing distribution are the same. We also studied the generalization of the model on *out-of-distribution* (OOD) settings, using synthetic noise addition, sparse sampling of the receive channels, and a different probe, imaging sequence and parameters. Finally, we evaluated the impact of finetuning with input perturbation and distillation for transferring the proposed approach to a specific setting.

**Evaluation under i.i.d. condition**

As mentioned in subsection 5.3.1, all the acquisitions from 6 mice (12 acquisitions) were left out during training and model selection to perform evaluation on unseen data. These acquisitions were obtained in similar experimental conditions as the ones used in training and therefore provided a reasonable i.i.d evaluation. We also used this configuration to evaluate the limitations of our approach with respect to the impact of the detection errors of the method used for labeling and the impact of the dataset size. To do so, we conducted three independent noise sensitivity analyses, each exploring a specific type of label corruption. Each noise type was evaluated independently, with all other noise sources set to zero. All experiments used identical training configurations and were evaluated on clean, noise-free left-out test data.

*Localization Error Analysis*: We introduced Gaussian noise to microbubble position labels with standard deviations of 0.0, 0.25, 0.5, 0.75 and 1.0 pixels to simulate annotation inaccuracies in ground truth positions.

*False Positive Detection Analysis*: We added microbubble detections at rates of 0.0, 0.1, 0.5, 1.0, and 2.0 false detections per frame, with the number of positions following a Poisson distribution, and their position being uniformly distributed.

*False Negative Detection Analysis*: We removed true microbubble detections with probabilities of 0.0, 0.1, 0.2, 0.3, and 0.5 to simulate missed detections in the ground truth annotations.

To provide insights into optimal data requirements and model data efficiency, we characterized the relationship between training dataset size and model performance. We conducted a systematic ablation study using subsets of the full training dataset at fractions of 1%, 2%, 5%, 10%, 20%, 30%, 50%, 70%, 90%, and 100% of the original size. To ensure fair comparison across different dataset sizes, All models were trained for the same number of optimization

steps (100,000), with the number of epochs automatically adjusted based on dataset size to maintain consistent gradient update counts. Model performance was evaluated on the complete test set, regardless of training subset size.

### *Out-Of-Distribution* Generalization

To evaluate the model sensitivity to domain shifts, we tested its performance without retraining under three OOD settings: noise perturbation, sparse sampling of the receive channels, and cross-probe evaluation.

**Noise perturbation** To simulate different acquisition quality, we added spatially correlated noise to the IQ data after beamforming and clutter filtering. Let $s \in \mathbb{C}^{H\times W\times T}$ beamformed IQ signal. White complex noise $n = (n_{\mathrm{r}} + jn_{\mathrm{i}})/\sqrt{2}$, with $n_{\mathrm{r}}, n_{\mathrm{i}} \sim \mathcal{N}(0,1)$, was convolved with a Gaussian PSF kernel $h$ to obtain noise $n_c = n * h$. The noisy signal was computed as

$$s' = s + \alpha\, n_c,$$

where the scaling factor $\alpha$ was adjusted to achieve a target signal-to-noise ratio (SNR) of either 0.5 or 20 dB. The PSF kernel $h$ was Gaussian with a standard deviation of $1\lambda$ along the lateral and axial directions. All processing parameters were kept identical to those used in the baseline (noiseless) evaluation.

**Sparse sampling of receive channels** To emulate a probe with a reduced number of active elements, a binary mask was applied to the channel dimension of the raw RF data. Only 32 channels were preserved out of the 128 available, selected at random using a fixed random seed. The same mask was applied consistently to all frames within each acquisition to reproduce a hardware configuration with only 32 addressed receive elements. All beamforming, clutter filtering, and localization–tracking parameters were kept identical to the baseline configuration.

**Cross-probe evaluation** To assess generalization across hardware and acquisition protocols, the models were evaluated on a dataset acquired with a different 128-element linear array transducer (L8-18iD, GE, IL, USA) using a distinct imaging sequence acquired by different operators from our group. For this dataset, the clutter filter and beamforming parameters were adapted to account for the acquisition differences. All other parameters were identical to those of the baseline configuration for all the methods to ensure fair comparison.

### 5.3.4 Finetuning with input perturbation and self-distillation

Herein, we first describe a novel and general finetuning strategy, which we applied in two specific settings, but that is not limited to these applications. More precisely we explored how this approach could improve the performance of the model under the noise addition model used for the OOD evaluation (see sec. 5.3.3). We also studied its potential to mitigate image degradations when using only a fraction of the receive elements.

#### Formalism

To improve robustness to acquisition variability without requiring additional manual annotations, we designed a finetuning strategy based on input perturbation and self-distillation. Given an input patch $x$ and its pseudo-label $y$ obtained from the original ULM processing pipeline, we applied a perturbation operator $\nu(\cdot)$ to the input, such as additive noise or channel masking, while keeping the label $y$ unchanged. It is important that the perturbed input $\nu(x)$ remains associated with the same target $y$. This enables the model to learn invariance to small acquisition perturbations without requiring reprocessing or manual relabeling.

Finetuning was then performed using both the fixed pseudo-label $y$ and the pretrained model as a teacher. Specifically, the teacher network processed the clean input $x$, producing latent feature representations $f(x)$, while the student network processed the perturbed input $\nu(x)$ to produce $f_\theta(\nu(x))$. The student was trained to minimize a composite objective combining (i) the original localization loss using $y$, and (ii) a latent-consistency loss encouraging the representation of $\nu(x)$ to align with that of the clean input $x$. This formulation effectively learns a mapping from $\nu(x)$ to the same point in the latent space as $x$, enforcing stability of the learned representation under perturbations.

#### Finetuning with noise perturbation

To improve robustness against spatially correlated acquisition noise, we applied the proposed input-perturbation finetuning strategy using a realistic noise model. At each optimization step, a complex circular Gaussian noise field was generated with the same dimensions as the beamformed IQ data. This white noise was convolved with a two-dimensional Gaussian point-spread function (PSF) kernel ($\sigma = 1\lambda \approx 4$ pixels, kernel size $7 \times 7$) to introduce spatial correlation, producing a noise component $n_c$. The perturbed input was obtained as $\nu(x) = x + \alpha n_c$, where the scaling factor $\alpha$ was randomly adjusted per sample to achieve a target signal-to-noise ratio (SNR) uniformly sampled between 5 dB and 15 dB. This range was selected to produce noticeable image degradation while maintaining training stability.

The teacher network processed the clean signal $x$ to produce a latent representation $f(x)$, while the student network was applied to the perturbed input $\nu(x)$ and produced both a latent representation $f_\theta(\nu(x))$ and a localization prediction $g_\theta(\nu(x))$. The student was trained to minimize

$$\mathcal{L} \;=\; \mathcal{L}_{\text{loc}} + \lambda_{\ell_2}\,\mathcal{L}_{\ell_2} + \lambda_{\cos}\,\mathcal{L}_{\cos}. \tag{5.1}$$

Here, $\mathcal{L}_{\text{loc}} = \mathcal{L}_{\text{loc}}\Big(g_\theta(\nu(x)), y\Big)$ is the loss function, originally used in training, between the student output $g_\theta(\nu(x))$ and the fixed pseudo-label $y$; $\mathcal{L}_{\ell_2} = \|f_\theta(\nu(x)) - f(x)\|_2^2$ is an $\ell_2$ penalty enforcing agreement between the student and teacher latent representations; and $\mathcal{L}_{\cos} = 1 - \cos\Big(f_\theta(\nu(x)), f(x)\Big)$ encourages alignment in feature direction, with $\cos(a,b) = \frac{a\cdot b}{\|a\|_2\|b\|_2}$. The weights were set to $\lambda_{\ell_2} = 0.1$, $\lambda_{\cos} = 100$, and $\lambda_{\text{loc}} = 1.0$, based on observations that increased weighted of the cosine loss improved all training losses. Optimization was performed using the Adam optimizer, for 50 epochs with a batch size of 512. Noise was regenerated at each iteration.

### Finetuning with channel masking perturbation

To evaluate robustness to channel undersampling, we applied the same self-distillation framework using a perturbation operator $\nu(\cdot)$ that simulates missing receive elements. In this configuration, the teacher network processed beamformed IQ data reconstructed from all 128 receive channels ($x_{\text{full}}$), while the student network received a perturbed version reconstructed from only 32 channels ($x_{\text{pert}} = \nu(x_{\text{full}})$). The perturbed inputs were generated offline following the same masking procedure used in the out-of-distribution evaluation (see sec. 5.3.3), and both inputs shared the same pseudo-label $y$. As in the noise experiment, the teacher model remained frozen and the student network, initialized from the teacher weights, was optimized using the same loss in (5.1), with $x$ corresponding to the full-channel input and $\nu(x)$ corresponding to the masked 32-channel input. All training parameters, including optimizer, learning rate schedule, batch size, and number of epochs, were identical to those used in the noise finetuning.

## 5.4 Results

In this section we first present the result obtained on i.i.d. acquisitions, then we evaluate the impact of different shifts in distribution and their impact on the performance. Finally, we display the results obtained by finetuning strategy to mitigate such domain shift for specific use.

Table 5.1 FRC and saturation comparison across methods and datasets.
Method yielding insufficient saturation (< 15%) are strikethrough as FRC showed to be unreliable in these conditions (see Fig. 5.7 for representative examples of low saturation and low FRC reconstruction)

| Method | 128 Channels | | 32 Channels | | 0.5dB Noise | | 20dB Noise | |
|---|---|---|---|---|---|---|---|---|
| | **FRC ($\mu$m)** | **Sat. (%)** | **FRC ($\mu$m)** | **Sat. (%)** | **FRC ($\mu$m)** | **Sat. (%)** | **FRC ($\mu$m)** | **Sat. (%)** |
| TAL | $28.0 \pm 6.6$ | $\mathbf{44.0 \pm 8.4}$ | $30.6 \pm 8.0$ | $40.7 \pm 7.9$ | $31.9 \pm 9.0$ | $\mathbf{33.0 \pm 7.6}$ | $28.1 \pm 7.2$ | $41.5 \pm 8.6$ |
| Gaussian Fitting | $\mathbf{24.0 \pm 5.5}$ | $38.2 \pm 10.3$ | $\mathbf{25.5 \pm 5.8}$ | $30.5 \pm 9.0$ | ~~$26.2 \pm 6.8$~~ | ~~$13.4 \pm 5.6$~~ | $\mathbf{23.4 \pm 5.7}$ | $26.0 \pm 8.5$ |
| TS-ULM | $26.5 \pm 5.7$ | $41.6 \pm 9.5$ | $27.9 \pm 6.1$ | $38.6 \pm 9.7$ | ~~$25.6 \pm 6.4$~~ | ~~$8.3 \pm 4.1$~~ | $26.2 \pm 5.8$ | $38.8 \pm 9.9$ |
| TS-ULM (Noise) | N.A. | N.A. | N.A. | N.A. | $\mathbf{29.0 \pm 6.4}$ | $30.1 \pm 10.3$ | $26.7 \pm 5.8$ | $\mathbf{42.8 \pm 9.2}$ |
| TS-ULM (32 Channels) | N.A. | N.A. | $27.6 \pm 5.9$ | $\mathbf{42.8 \pm 9.1}$ | N.A. | N.A. | N.A. | N.A. |

### 5.4.1 Performance in i.i.d. setting

#### Comparison on similar setting

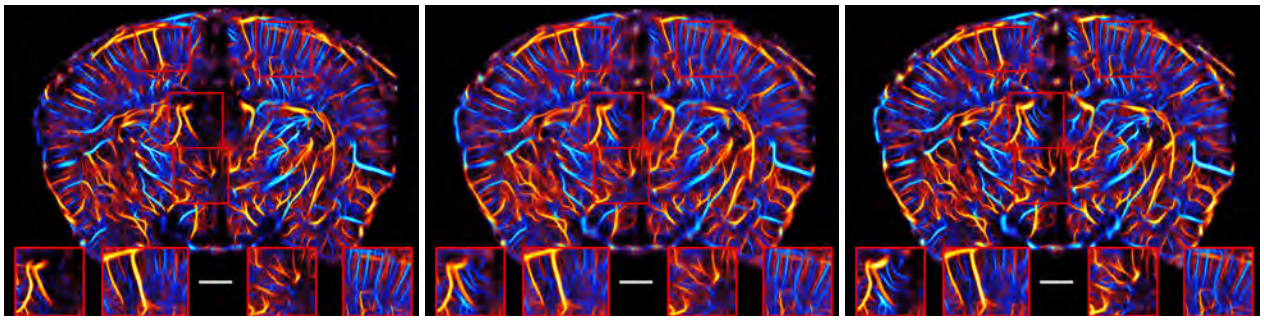


Figure 5.2 Transcranial ULM of a mouse brain processed using Gaussian Fitting and Hungarian tracking (left), Tracking-and-Localize (center), and our proposed approach Teacher-Student ULM (TS-ULM) (right), the scale bar represents 5 wavelengths ( 0.5mm).

When evaluated on mouse datasets acquired under conditions similar to the training set, the proposed TS-ULM method demonstrated qualitative improvements in image quality, showing higher vascular saturation and finer spatial resolution (Fig. 5.2). As illustrated in the left insets of Fig. 5.2, TS-ULM reconstructed vessels in low-signal areas better than TAL (where Gaussian fitting failed to detect vessels). As shown in the center-right inset, TS-ULM displayed higher detection intensity than both TAL and Gaussian fitting in specific regions. These observations were consistent across other visually inspected ULM acquisitions from the test set.

Quantitatively, TS-ULM achieved spatial coherence superior to the TAL method and better saturation than Gaussian fitting; however, it yielded overall performance (FRC : $27 \pm 6 \mu m$,Sat: $42 \pm 10\%$) very similar to the two baseline methods in terms of saturation vs. FRC trade-offs ($28 \pm 7 \mu m$, $44 \pm 8\%$ for TAL and $24 \pm 6 \mu m$, $38 \pm 10\%$ for the Gaussian fitting).

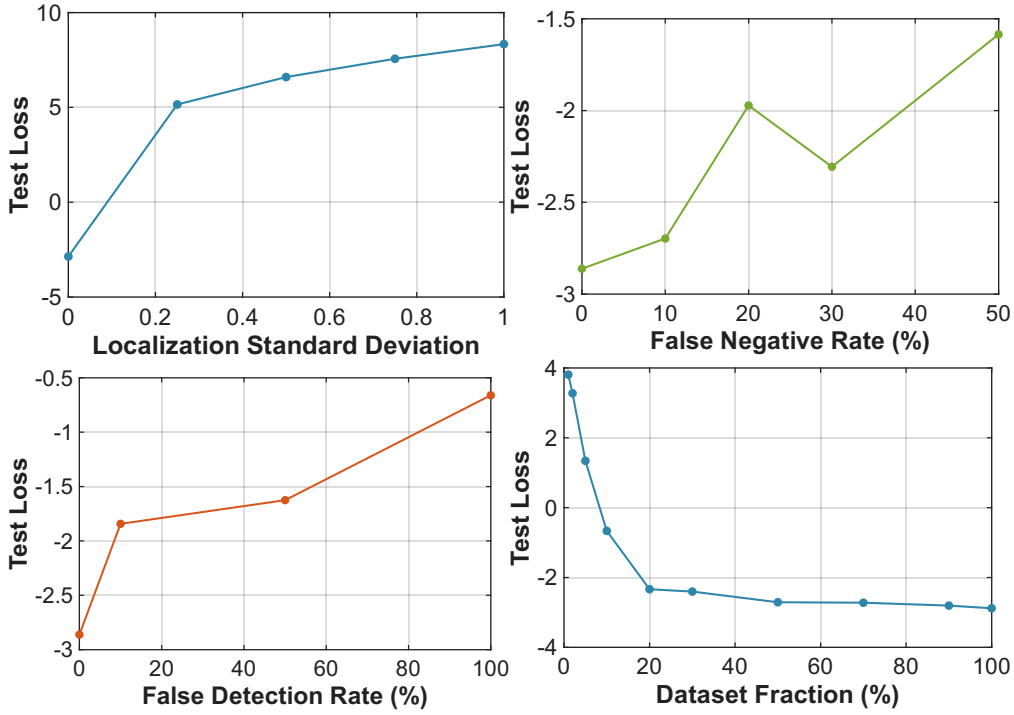



Figure 5.3 Evolution of the loss function value on the test set under varying training perturbation. Top left: localization errors are added to the detection (in pixels $\sim \lambda/4$). Top right: Detections are removed from the labels during training. Bottom Left: Uniformly sampled detection are randomly added to the labels. Bottom right: The dataset size is increased until the whole training set is used.

### Impact of localization errors

When trained on the same input data but with labels intentionally degraded to mimic reduced localization accuracy, the model performance failed to stabilize even at the smallest levels of noise (Fig. 5.3). Introducing positional noise to the localization labels strongly impaired training, with the test loss exceeding 5, corresponding to an almost complete training collapse, even for a standard deviation as small as $\lambda/16$. In contrast, experiments simulating false-positive or false-negative detections produced a more linear increase in test loss and limited training degradation (loss values remained above $-0.5$ in all runs). False detections, however, were found to be more detrimental than missed detections: removing 50% of the true positions resulted in a lower loss than doubling the detections by adding spurious ones.

### Importance of the dataset size

In contrast, model performance appeared to reach a plateau with increasing dataset size, with most of the improvement achieved using only the first half of the training data. With just

20% of the training dataset, the test loss was already below $-2$, and it gradually decreased to $-2.5$ when using 50% of the data and to $-2.6$ with the full dataset. It is worth noting that experiments conducted with smaller dataset fractions exhibited strong overfitting, with training losses dropping to $-14$ while test losses exceeded 5.

### 5.4.2 Out-Of-Distribution (OOD) generalization

To further evaluate the generalization capability of the proposed model, we tested its performance under several OOD conditions that differ from the training setup. These experiments included the addition of correlated acquisition noise, a reduction in the number of active receive channels, and an evaluation on data acquired with a different ultrasound probe. The following subsections detail the results obtained for each of these scenarios.

#### Noise perturbation

Figure 5.4 illustrates the qualitative impact of two levels of noise (20 dB SNR and 0.5 dB SNR). At 20 dB SNR, TS-ULM visually outperformed both Gaussian fitting, which missed many vessels, and TAL, which introduced background noise from false detections. This trend can be observed globally and in the selected insets of Fig. 5.2, and it was consistent across other acquisitions in the test set. At high noise levels (0.5 dB SNR), neither Gaussian fitting nor TS-ULM reconstructed the complete vasculature, while TAL processing showed even greater background noise than at moderate noise levels.

As summarized in Table 5.1, quantitative analysis showed that Gaussian fitting saturation ($26 \pm 9\%$) dropped significantly under moderate noise (20 dB SNR), in contrast to TAL ($42 \pm 9\%$) and TS-ULM ($39 \pm 10\%$), while the FRC remained similar to the IID setting. At 0.5 dB SNR, the FRC increased for TAL, whereas both Gaussian fitting and TS-ULM did not achieve sufficient saturation for the FRC to be meaningful.

#### Sparse sampling of receive channels

Figure 5.5 illustrates the impact of synthetic channel reduction on transcranial ULM reconstructions. When only 32 out of 128 receive elements were used, all methods exhibited visible degradation compared with the full-aperture reconstructions with some shadowing artifact on the left and center of the brain. Nevertheless, the proposed TS-ULM approach maintained superior vessel continuity and overall image coherence compared with the TAL baseline. Background noise similar to the noise experiments appeared for the TAL method.

Quantitatively, all methods had their average FRC increased (between 1.4 for TS-ULM and

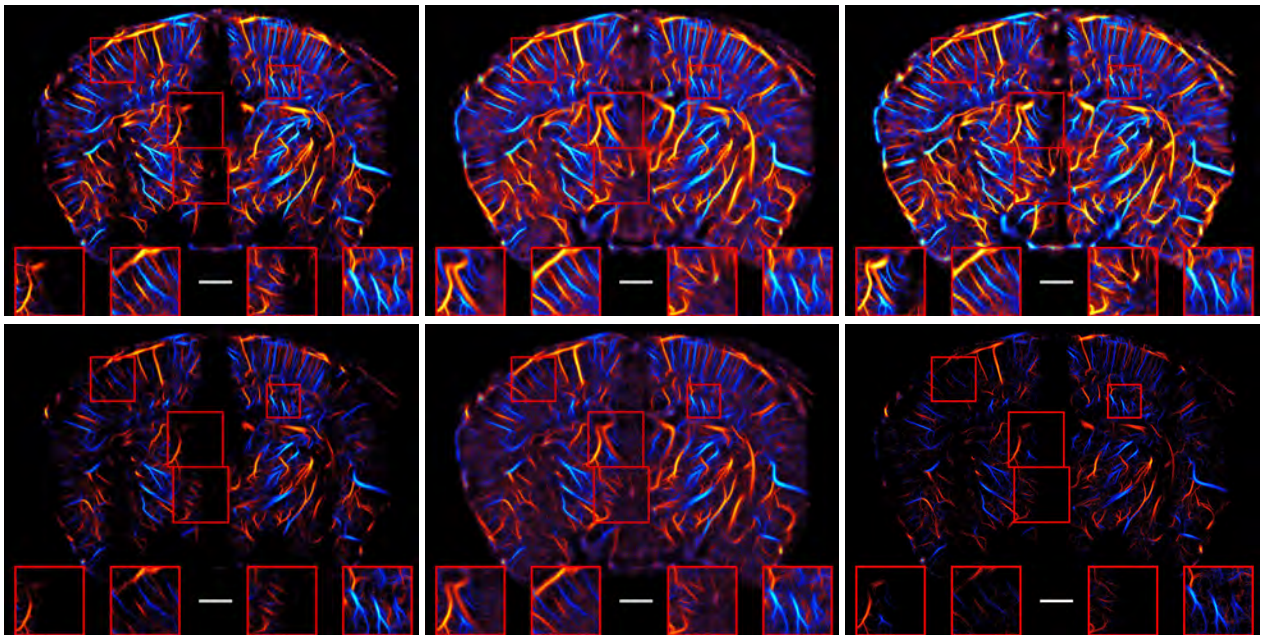


Figure 5.4 Transcranial ULM acquisition with synthetic noise addition (top row: 20db, bottom row:0.5db). From left to right: Gaussian fitting, TAL, and TS-ULM. TS-ULM showed better robustness at moderate level of noise, but its performance drastically worsened under severe noise. Gaussian fitting displayed the best FRC across ($23\mu m$ at 20dB SNR and $26\mu m$ at 0.5dB SNR) despite incomplete vascular reconstruction. The scale bar represents 5 wavelengths ($\sim$ 0.5mm)

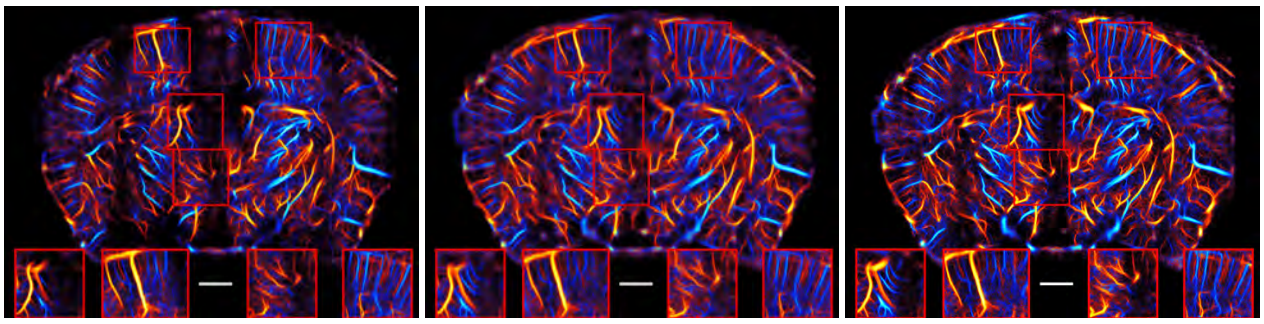


Figure 5.5 Transcranial ULM acquisition of a mouse brain with synthetic reduction of the number of channels in receive (32 out of 128). From left to right: Gaussian fitting, TAL, and TS-ULM. the scale bar represents 5 wavelengths ($\sim 0.5mm$)

2.6 $\mu m$ for TAL) and a vascular saturation dropped twice more abruptly for Gaussian Fitting(Table 5.1) than for TAL and TS-ULM. By contrast, attempts to reconstruct images using only 16 channels failed for all methods, with strong grating lobes.

#### Cross-probe evaluation

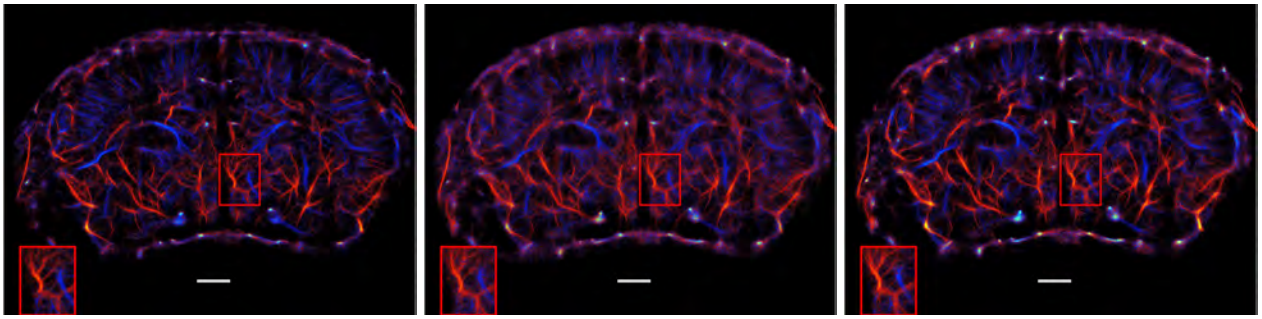


Figure 5.6 Transcranial ULM acquisition with a probe (GE L8-18iD) not used in the training set. The scale bar represents $5mm$ wavelengths.

To evaluate cross-probe generalization, the models were tested on a dataset acquired with a different linear probe (GE L8–18iD) that was not used during training.

As shown in Figure 5.6, Gaussian Fitting, TAL and TS-ULM successfully reconstructed the main vascular architecture, with more refined vasculature for Gaussian Fitting and TS-ULM and more complete imaging for TAL and TS-ULM. This trend was consistent with the quantitative results, where Gaussian-Fitting (FRC: $35\mu m$, Sat: 16%) and TS-ULM (FRC: $36\mu m$, Sat: 19%) achieved a higher spatial coherence and TAL (FRC: $43\mu m$, Sat: 22%) higher saturation.

### 5.4.3 Impact of input perturbation and distillation

Figure 5.7 illustrates the effect of applying task-specific finetuning under two challenging acquisition perturbations: strong additive noise (top row) and channel decimation (bottom row).

#### Noise robustness

Under stronger noise addition (5 dB SNR), the baseline TS-ULM model and Gaussian Fitting were not able to reach sufficient vascular saturation, whereas the TAL method added background noise due to false detections. In opposition, the fine-tuned TS-ULM successfully mapped the complete vasculature without additional background noise (see Fig. 5.7).

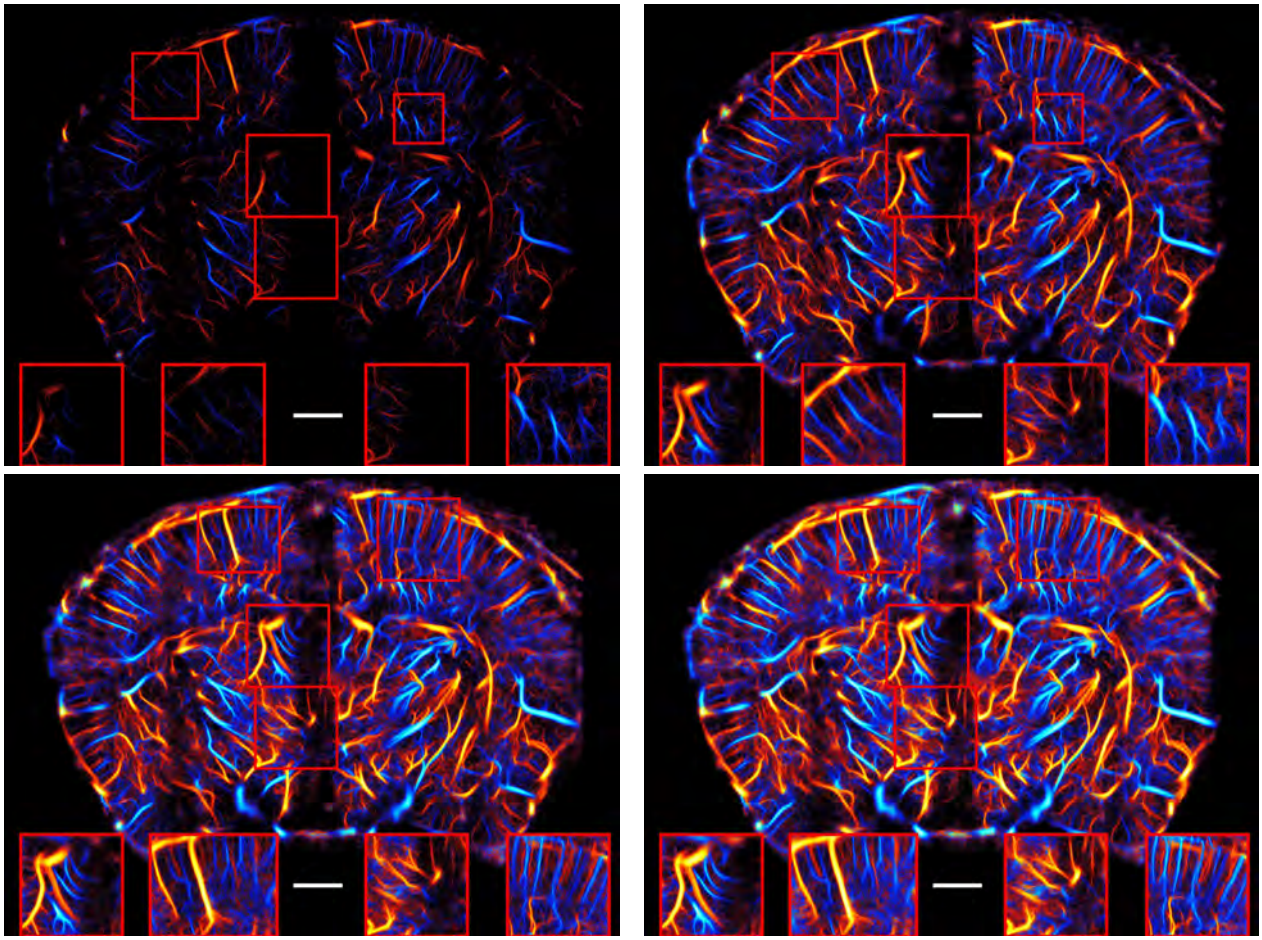


Figure 5.7 Transcranial ULM acquisition with synthetic perturbations. Top row : strong noise addition (0.5 dB SNR). Bottom row : reduction of the number of channels in receive (32 out of 128). TS-ULM :left, TS-ULM with specific finetuning: right. The scale bar represents 5 wavelengths ( 5mm).

Quantitatively, fine-tuned TS-ULM demonstrated better spatial coherence than TAL ($29 \pm 6$ $\mu$m for TS-ULM vs. $32 \pm 9$ $\mu$m for TAL). While vascular saturation was slightly higher for TAL, this metric is likely overestimated due to the background noise.

**Channel decimation**

The bottom row of Fig. 5.7 illustrates the impact of finetuning under a reduction of the number of receive channels from 128 to 32. Visually, the finetuning offered higher intensity level (particularly visible in the insets) with little change to the resolution. Quantitatively, this improvement translated to an increase in vascular saturation from $39 \pm 10\%$ to $43 \pm 9\%$, while maintaining a comparable spatial resolution ($28 \pm 6$ $\mu$m versus $28 \pm 6$ $\mu$m). Overall, the finetuned TS-ULM offered the best compromise between resolution and saturation across all the methods. When the number of receive elements was further reduced to 16, finetuning no longer improved reconstruction quality, as the strong grating-lobe artifacts and limited spatial sampling prevented stable localization.

## 5.5 Discussion

### 5.5.1 Improved image quality in i.i.d. settings

In the *i.i.d.* evaluation setting, where training and test acquisitions were obtained under similar experimental conditions, the proposed TS-ULM approach consistently matched or slightly outperformed the two baseline pipelines (Gaussian fitting with Hungarian tracking and TAL). These results suggest that a deep learning model trained directly on *in vivo* pseudo-labels can recover vascular architecture at the same or higher level of performance than the conventional method used to generate the pseudo labels.

This seemingly surprising outcome can be explained by the training formulation. By repeatedly observing microbubbles at different relative positions and local backgrounds, the network learns to localize them independently of their absolute location within the field of view. As a result, TS-ULM appears to capture spatial features that are less sensitive to local amplitude variations caused by attenuation or skull-induced shadowing, potentially explaining its improved recovery of vessels in shadowed regions where TAL and Gaussian fitting often lose signal. In addition, the aggregation effect of training over large numbers of pseudo-labels may help the model mitigate isolated annotation errors, a phenomenon consistent with observations in large-scale pseudo-label and noisy-label learning frameworks [239].

The dataset size ablation experiment (Fig. 5.3) indicates that the proposed dataset is suffi-

ciently large and diverse to train compact convolutional models such as DECODE. Performance saturated after using roughly half of the available data, suggesting that it already contains enough variability for stable learning. Such an *i.i.d.* evaluation setting based on *in vivo* data is crucial for assessing the representational power of different architectures, offering a more realistic alternative to existing simulation-based dataset.

Finally, the label corruption experiments reveal that training stability depends critically on the spatial accuracy of the pseudo-labels. While learning remained effective under moderate levels of missed or spurious detections, even small subpixel localization errors led to a marked degradation in performance. This sensitivity suggests that improving label reliability—through consensus aggregation of conventional ULM methods or teacher–student refinement—could further enhance robustness and generalization. Such strategies would help fully exploit the scale of the proposed dataset and learning framework while maintaining the spatial precision required for reliable *in vivo* supervision.

### 5.5.2 Robustness to limited change in distribution

Across moderate distribution shifts, TS-ULM remained robust and generally outperformed conventional baselines. These observations indicate a limited performance drop under realistic shifts: in practice, TS-ULM can be applied under various conditions without retraining or parameter tuning.

With spatially correlated noise at 20 dB SNR (Fig. 5.4), TS-ULM preserved finer vessels than TAL and more continuous vasculature than Gaussian Fitting. Similarly, under channel subsampling to 32/128 receive elements (Fig. 5.5), TS-ULM maintained clearer vessel continuity and higher apparent coverage.

In the cross-probe evaluation (Fig. 5.6), TS-ULM offered a better trade-off between spatial coherence and saturation than TAL and Gaussian Fitting. A minor fine-tuning on similar datasets would likely allow TS-ULM to outperform TAL.

By contrast, extreme shifts led to failure modes for all approaches. At very low SNR (0.5 dB), reconstructions degraded substantially; while TAL retained more structure than TS-ULM, the overall quality remained poor. The next section describes how our proposed perturbation-specific fine-tuning strategy mitigates these effects.

### 5.5.3 Finetuning for enhanced robustness under severe perturbations

Perturbation-specific fine-tuning substantially improved model stability under strong perturbations. By combining input perturbation with self-distillation, this approach can, in

principle, accommodate a wide range of experimental variations, such as noise, channel configuration, sequence differences, or suboptimal parameters in the processing pipeline. While we explored only two representative perturbations—additive noise and channel subsampling—this strategy establishes a foundation for extending model robustness to other domains of variability.

Under strong noise (0.5 dB SNR), the fine-tuned TS-ULM model successfully mapped the entire vascular network, whereas no other approach achieved this. At moderate noise levels (20 dB SNR), the pre-trained and fine-tuned models performed similarly, as the baseline model was already robust to moderate degradation; fine-tuning primarily proved beneficial in more severe noise regimes.

A similar trend was observed under channel subsampling. With only 32 of 128 receive elements, fine-tuning improved vascular completeness and image coherence beyond both the baselines and the base TS-ULM, indicating successful adaptation to limited-aperture acquisition. However, when the number of channels was reduced to 16, all methods failed due to strong grating-lobe artifacts and significant signal loss. This highlights the robustness limits of our current approach. Future improvements might be achievable with larger input patches—allowing the model to account for grating lobes—or more expressive models (e.g., deeper backbones or Transformer-based architectures) to compensate for extreme signal degradation.

A similar fine-tuning strategy could also be extended to volumetric ULM, enabling reconstruction from matrix arrays with reduced channel counts (e.g., 256 instead of 1024), thereby lowering hardware and data throughput requirements for 3D imaging.

## 5.6 Conclusion

In summary, this work demonstrates that deep learning models can be trained directly from *in vivo* ULM data using pseudo-labels, achieving performance comparable or superior to conventional pipelines while remaining robust across acquisition conditions. The combination of input perturbation and self-distillation provides a general framework for improving adaptability without additional annotation. Together with the proposed dataset, these findings lay the groundwork for reproducible benchmarking and future extensions toward 3D imaging and broader applications.

# CHAPITRE 6 ARTICLE 4 : ULMSHARE : A LARGE-SCALE IN VIVO ULTRASOUND LOCALIZATION MICROSCOPY DATASET FOR MICROVASCULAR IMAGING

Les chapitres précédents ont illustré le potentiel de l'apprentissage profond pour repousser les limites de la microscopie de localisation ultrasonore (ULM), que ce soit pour accélérer l'imagerie 3D ou pour améliorer la robustesse des reconstructions face aux aléas expérimentaux. En particulier, le chapitre 5 a démontré que l'entraînement sur des données *in vivo* permettait de franchir un cap en termes de performance et de généralisation par rapport aux modèles entraînés sur simulation.

Cependant, la démocratisation de telles approches se heurte à un obstacle pratique majeur : la rareté des données expérimentales accessibles. La grande majorité des jeux de données publics actuels sont soit simulés, soit limités à un très faible nombre d'acquisitions *in vivo*, ce qui restreint considérablement la capacité de la communauté à entraîner des modèles performants et à comparer objectivement les algorithmes de reconstruction.

Dans ce dernier chapitre de contributions, nous présentons **ULMShare**, la plus grande base de données publique d'ULM *in vivo* à ce jour. Constituée de 99 acquisitions transcrâniennes provenant de 61 souris et totalisant plus de 30 TB de données brutes, cette ressource couvre une grande diversité de protocoles expérimentaux, de sondes et de conditions d'imagerie.

Ce chapitre correspond au quatrième article de cette thèse, soumis à *Scientific Data* le 12 décembre 2025. Il constitue une contribution fondamentale pour la reproductibilité de la recherche mais aussi fournit la fondation essentielle pour permettre à communauté d'étendre les méthodes proposées précédemment.

# ULMShare: A Large-Scale In Vivo Ultrasound Localization Microscopy Dataset for Microvascular Imaging

Brice Rauby[1,2,*], Nin Ghigo[1,*], Gerardo Ramos-Palacios[3], Alexis Leconte[1], Stephen A. Lee[1], Alice Wu[1], Paul Xing[1], Oleksandra Gulenko[1], Louis Caron[1], Antoine Malescot[4,5], Eric Martineau[4,5], Jonathan Porée[1], Maxime Gasse[2,6,7], Ravi Rungta[5,8,9], Abbas Sadikot[3], Jean Provost[1,9,10]

[1]Department of Engineering Physics, Polytechnique Montréal, Montreal, QC, Canada
[2]Mila – Quebec Artificial Intelligence Institute, Montreal, QC, Canada
[3]Montreal Neurological Institute, McGill University, Montreal, QC, Canada
[4]Department of Physiology and Pharmacology, Université de Montréal, QC, Canada
[5]Department of Stomatology, Université de Montréal, QC, Canada
[6]Microsoft AI, Montreal, QC, Canada
[7]Department of Computer Engineering and Software Engineering, Polytechnique Montréal, QC, Canada
[8]Department of Neuroscience, Université de Montréal, QC, Canada
[9]Centre Interdisciplinaire de Recherche sur le Cerveau et l'Apprentissage, Université de Montréal, QC, Canada
[10]Montreal Heart Institute, Montreal, QC, Canada

* Indicates equal contribution from both authors

## 6.1 Abstract

Ultrasound Localization Microscopy (ULM) enables microscopic imaging of the cerebral microvasculature *in vivo*, but relies on a multi-stage processing pipeline in which acquisition settings and reconstruction processes strongly influence the final output. Existing public datasets remain sparse, restricting rigorous evaluation and slowing progress in algorithm development, including emerging machine-learning approaches, which by design require large quantities of data to be robust and reliable. We introduce **ULMShare**, an open-access dataset of 99 whole-brain transcranial ULM acquisitions from 61 healthy mice (36 females, 22 males, 3 unknown; mean age: $8.2 \pm 5.5$ weeks; mean weight: $17.7 \pm 4.2$ g), for a total of 30TB of raw data. The dataset spans three experimental procedures, multiple injection

and anesthesia protocols, two ultrasound probes, and different imaging planes and orientations. Each acquisition includes raw ultrasonic data, detailed metadata, an illustrative reconstruction and the corresponding microbubble trajectories. Alongside the data, we report vascular saturation, Fourier Ring Correlation, and track-length statistics, plus expert visual gradings. ULMShare provides a broad, standardized and publicly available resource for method development, validation, and benchmarking. The full dataset is available on the Federated Research Data Repository and additional resources are hosted on the ULMShare Github repository.

## 6.2 Background & Summary

Ultrasound Localization Microscopy (ULM) is a super-resolution ultrasound technique that images microvasculature *in vivo*, at depth, and at resolutions breaching the diffraction limit [11, 13, 17]. By localizing and tracking clinically approved microbubbles (MB) injected into the bloodstream, ULM reconstructs detailed maps of the microvascular network and provides quantitative measurements of blood flow dynamics.

In the brain, ULM studies on animal models have been critical to develop new sequences and methodologies [24] enabling extraction of complex hemodynamic markers such as pulsatility imaging [24, 25, 84], or functional imaging [2, 26], as well as for resolving capillary-level vasculature [73, 242].

Despite these promising advances, ULM remains a modality that relies on experimental protocols requiring costly equipment and skilled technicians, which may induce high variability in acquisition quality even within the same protocol. Consequently, new methodological developments are often evaluated on a reduced number of *in vivo* acquisitions, which limits their generalization and overall impact. Additionally, some groups lack access to the required equipment and are therefore unable to validate their methods on *in vivo* data.

ULM relies on a complex reconstruction pipeline: beamforming, clutter filtering, MB localization and tracking—where each step can significantly impact the output quality. Open-access datasets are critical as they facilitate fair comparison between methodologies and improve reproducibility. Recent efforts were made within the community to release open-access datasets such as PALA [67] or companion datasets to animal studies [2, 22, 23, 26, 243]. The PALA benchmark has gained substantial traction [117, 118, 129, 149, 179, 184, 202, 244–266] demonstrating the strong demand for publicly available data to support the development and validation of ULM processing algorithms. However, existing public datasets remain scarce and are mostly limited to simulated data or a small number of *in vivo* acquisitions [67].

Large and diverse *in vivo* ULM datasets are essential for statistically rigorous method evaluation under varied protocols, reproducibility, fair benchmarking, and the development of robust machine-learning models [49]. By lowering the barrier to entry and enabling researchers without access to costly equipment to work directly with *in vivo* data, such resources can accelerate innovation and broaden the impact of ULM across the scientific community.

In this paper, we present **ULMShare**, an open-access dataset of ultrasound acquisitions, unique by its scale, two orders of magnitude larger than existing public ULM dataset. 99 acquisitions, each lasting several minutes, were acquired in 61 anesthetized wild-type mice (aged 3 to 16 weeks) through intact skin and skull, totaling over 30 TB. The full dataset is made available publicly as collection on the Federated Research Data Repository and ULM reconstructions and helper scripts are provided on a companion Github repository.

ULMShare comprises both previously unpublished data and data that have been used in prior publications. Continuous transcranial ULM acquisitions were first presented in Lee et al. [73], where they were used to demonstrate temporally continuous imaging for functional brain mapping. Data from Xing et al. [69, 164] were employed to investigate aberration correction techniques, highlighting methods to improve image quality and localization accuracy in transcranial ULM. Acquisitions reported in Ghigo et al. [84] were used for pulsatility evaluation, demonstrating the potential of ULM to capture dynamic vascular signals. Finally, datasets included in Leconte et al. [74] supported the development and validation of a spatiotemporal tracking algorithm for MB trajectories.

## 6.3 Methods

### 6.3.1 ULM acquisition

ULMShare consists of **99** transcranial brain ULM acquisitions obtained from 61 healthy mice (36 females, 22 males, 3 unknown; mean age: $8.2 \pm 5.5$ weeks; mean weight: $17.7 \pm 4.2$ g; see Fig. 6.1). Experimental procedures were approved by the Animal Care Ethics Committees of the University of Montreal (22-013), McGill University (AUP-4532), and the Montreal Heart Institute (2023-32-02 TAC-ultrasons).

#### Data Acquisition Protocols

The dataset aggregates acquisitions collected over three years using three distinct experimental protocols, previously described in published articles [69, 73, 74, 84, 267].

Common to all procedures, animals were secured in a stereotaxic frame to minimize motion.

For the majority of acquisitions ($n = 97$), the scalp was shaved and degassed ultrasound gel was applied. For 2 acquisitions (Protocol 3), the scalp was surgically removed to minimize attenuation. All data were acquired using a Vantage-256 system (Verasonics, WA, USA) emitting tilted plane waves ($-5°$ to $5°$, centered at $0°$). The specific acquisition parameters differentiating the three protocols are detailed in Table 6.1.

Table 6.1 Comparison of Acquisition Protocols used in ULMShare.

| **Parameter** | **Protocol 1** | **Protocol 2** | **Protocol 3** |
|---|---|---|---|
| Reference | [74, 84] | [73] | [164, 267] |
| **Acquisitions ($n$)** | **71** | **15** | **13** |
| Anesthesia | Isoflurane (1–2%) | Isoflurane (1–2%) | Ketamine/Medetomidine |
| MB Injection | Tail vein injection | Tail vein catheter [a] | Retro-orbital injection |
| Transducer | L22-14v | L8-18iD | L22-14v |
| Center Freq. | 15.625 MHz | 10.4667 MHz | 15.625 MHz |
| Plane Waves | 11 | 9 | 11 |
| Frame Rate | 1000 Hz | 1000 Hz | 1000 or 1600 Hz |
| Duration | 160 s | 300 s | 300 s |
| Sampling | 100% Bandwidth | 50% Bandwidth | 100% Bandwidth |

[a] *Note: 3 acquisitions followed Protocol 2 but used tail vein injection.*

### Dataset Composition and Variations

The 99 total acquisitions include specific subsets designed for sensitivity analysis and anatomical mapping, which are included within the protocol counts listed in Table 6.1. All the following details are systematically provided in the metadata.

Regarding Protocol 1, while the previously published configuration used a 1:10 saline dilution ($n = 6$), the majority of acquisitions employed a 1:50 dilution ($n = 64$); one acquisition used a 1:20 dilution. Additionally, sequential acquisitions were performed in 3 mice to image different anatomical planes, yielding 17 acquisitions (5–6 per mouse) taken at 0.5 mm intervals. Protocol 2 includes 3 acquisitions that combined the animal preparation of Protocol 1 (tail vein injection, 1:20 dilution, 4 $\mu$L/g) with the imaging sequence of Protocol 2 (L8-18iD probe, 300 s duration). In Protocol 3, a subset of acquisitions was collected during stimulation for functional imaging (treated here as resting-state due to lack of response), and for 2 acquisitions (1 mouse), the scalp was surgically removed to minimize attenuation.

Finally, for 23 acquisitions across the dataset, an expert manually identified the anatomical slice position, including striatum ($n = 11$), pons ($n = 2$), midbrain ($n = 4$), hippocampus ($n = 3$), and cerebellum ($n = 3$). Detailed metadata for every acquisition is provided in the

accompanying JSON files.

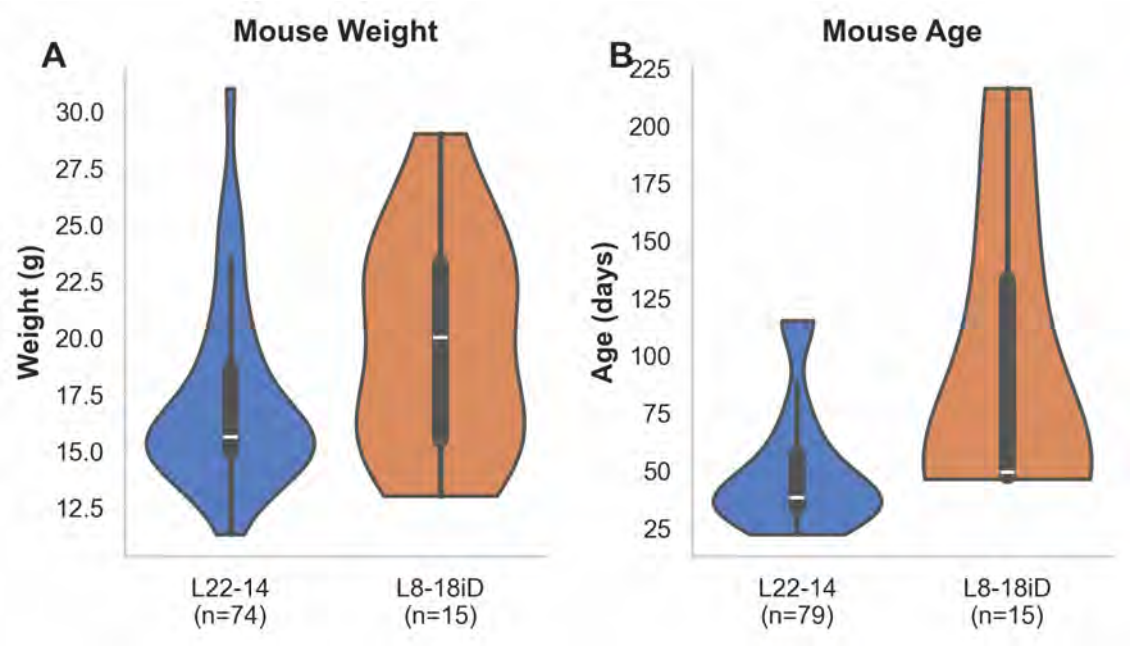



Figure 6.1 Overview of animal-specific information in ULMShare.

Table 6.2 Detailed composition of the ULMShare dataset grouped by experimental protocol

| Mice | Acq. | Injection Route | Dilution | Volume | Probe | Anatomical Coverage (Acquisitions) |
|---|---|---|---|---|---|---|
| ***Protocol 1*** *[74, 84]* | | | | | | |
| 36 | 64 | Tail vein | 1:50 | 4 $\mu$L/g | L22-14v | Striatum ($n = 11$), Midbrain ($n = 4$), Hippo. ($n = 3$), Cerebellum ($n = 3$), Pons ($n = 2$), Unreg. ($n = 41$) |
| 3 | 6 | Tail vein | 1:10 | 4 $\mu$L/g | L22-14v | Unregistered |
| 1 | 1 | Tail vein | 1:20 | 4 $\mu$L/g | L22-14v | Unregistered |
| ***Protocol 2*** *[73]* | | | | | | |
| 12 | 12 | Tail vein catheter | 1:10 | 4 $\mu$L/g | L8-18iD | Unregistered |
| 3 | 3 | Tail vein | 1:20 | 4 $\mu$L/g | L8-18iD | Bregma $-3$ mm ($n = 3$) |
| ***Protocol 3*** *[69, 267]* | | | | | | |
| 5 | 11 | Retro-orbital | Pure | 10 $\mu$L | L22-14v | Functional / Deep Brain targets |
| 1 | 2 | Retro-orbital | Pure | 10 $\mu$L | L22-14v | Functional (Scalp Removed) |

**Total:** 61 Mice, 99 Acquisitions.

### 6.3.2 ULM Processing

In addition to the raw data, ULMShare provides for each acquisition an illustrative signed ULM density map and the corresponding detected microbubble (MB) tracks. These ULM maps were generated using fixed, non-optimized processing parameters. Their purpose is to provide an example of the vascular information that can be obtained from each acquisition;

they should not be used as benchmarks for quantitative comparison. In phase-quadrature data were beamformed onto an isometric grid with a spacing of $\lambda/4$ using standard GPU-based delay-and-sum beamforming. An f-number of 1.4 was used for data acquired with the L22-14v probe, and 1.0 for acquisitions using the L8-18iD probe. Tissue signals were removed from the raw data using Singular Value Decomposition (SVD) clutter filtering [72]. The first 20 singular values were removed for L22-14v acquisitions and the first 30 for L8-18iD acquisitions.

Two signal-enhancement processes were then applied to the clutter-free data [268]: spatially dependent time-gain compensation (TGC) and lag-1 autocorrelation. The spatially dependent TGC aims to improve signal quality in shadowed regions. A 2D Gaussian filter with a standard deviation of $2.25\lambda$ (corresponding to 9 pixels of the beamforming grid) was applied to the Power Doppler (time average of the power across a set of hundred of frames) to compute a spatially dependent TGC map. This map represents an estimate of the spatial intensity distribution and is used to equalize intensity and reduce shadowing artifacts. Temporal lag-1 autocorrelation was then applied, reducing noise while enhancing MB signals.

On the resulting MB-enhanced data, a spatiotemporal tracking algorithm was employed to track MB trajectories in space and time [74]. Briefly, MBs moving through space and time resemble tubular structures with radii comparable to the point spread function (PSF), which can be enhanced using a spatiotemporal filter. The tubular radius was set to $\lambda$ across the image. A thinning algorithm was then applied to segment the centerlines of these tubular structures at pixel resolution [269]. Finally, subwavelength MB positions were estimated using a radial symmetry algorithm on regions of interest (ROI) of 15 pixels [67]. Tracks shorter than 20 frames were considered noise and discarded. Following [74], $\epsilon$ was fixed at 1.4 to reject tubular structures orthogonal to the temporal dimension, and $\tau$ was fixed at 0.6 to adapt the sensitivity of the function to the non-homogeneous intensity of MB PSF trajectories. All ULM processing parameters are summarized in Table 6.3.

## 6.4 Data Records

The ULMShare dataset is publicly available through the Federated Research Data Repository (FRDR). The collection includes 99 transcranial ultrasound acquisitions from 61 mice recorded between March 2022 and March 2025. Each record corresponds to a single acquisition and contains raw ultrasonic channel data and structured metadata. In total, the dataset represents approximately 30 TB of raw data. Some acquisitions correspond to datasets previously used in published studies; their associated DOIs are included in the acquisition-level metadata files and in the global `metadata.csv` file.

Table 6.3 Summary of fixed parameters used in ULM processing for each probe type.

| **Parameter** | **L22-14v** | **L8-18iD** |
|---|---|---|
| Beamforming | Grid spacing $\lambda/4$<br>f-number = 1.4 | Grid spacing $\lambda/4$<br>f-number = 1.0 |
| Clutter filtering (SVD) | First 20 singular values removed | First 30 singular values removed |
| TGC filter size | 2D Gaussian with $\sigma = 9\lambda$ | 2D Gaussian with $\sigma = 9\lambda$ |
| Tracking parameters | Tubular radius = $\lambda$<br>$\epsilon = 1.4$, $\tau = 0.6$ | Tubular radius = $\lambda$<br>$\epsilon = 1.4$, $\tau = 0.6$ |
| ROI for radial symmetry localization | 15 pixels | 15 pixels |
| Minimum number of frames per track | 20 | 20 |

To support transparent processing and reproducibility, we provide example ULM maps, detected microbubble trajectories, and postprocessing metrics for each dataset. These examples are hosted in a dedicated GitHub repository to enable accessibility, version control, and continuous updates, together with a dataset summary and MATLAB helper scripts.

### 6.4.1 Data directory

All raw acquisitions are stored under the `Data/` directory and organized by mouse index, as shown in Fig. 6.2. Each mouse folder contains a `mouse.json` file describing strain, sex, age, body weight, protocol number, and facility location. Each acquisition directory includes a `sequence.json` file specifying the imaging parameters (e.g., transducer type, number of plane-wave angles), an `acquisition.json` file containing acquisition-specific information such as injection parameters or procedure details, and a `raw/` folder that stores the Verasonics channel data files (`dataXXXX.bin`).

### 6.4.2 Code directory

All example outputs and supporting resources are hosted in a dedicated GitHub repository to enable accessibility, version control, and future updates. The repository mirrors the mouse/acquisition hierarchy of the FRDR archive.

The `examples_ulm/` directory replicates the FRDR mouse/acquisition structure. Each acquisition folder includes:

— an illustrative ULM density map (`density_map.png`),

— detected microbubble trajectories (`tracks.json.xz`),

```
Data/
├── Mouse_XX/
│   ├── mouse.json
│   ├── acquisition_01/
│   │   ├── sequence.json
│   │   ├── acquisition.json
│   │   └── raw/
│   │       ├── data0001.bin
│   │       ├── data0002.bin
│   │       └── ..
│   ├── acquisition_02/
│   └── ..
├── ..
└── README.txt
```

Figure 6.2 Directory structure of the raw data stored in the FRDR repository.

— metrics (`metrics.json`) computed for technical validation as described in 6.5.

The file `tracks.json` contains localized microbubble trajectories without interpolation. It includes a `metadata` section (pixel size, frame rate, units, acquisition identifier) and an `all_tracks` list, where each element corresponds to a processed buffer with its `buffer_index` and associated tracks.

The `code/` directory provides lightweight MATLAB scripts demonstrating the typical workflow from raw Verasonics buffers to a simple ULM reconstruction as explained in section 6.7. These examples rely on open-source libraries such as the MATLAB UltraSound Toolbox (MUST) [131, 270, 271] and the Tracking and Localization toolbox [74].

The `summary/` directory contains high-level overview material to facilitate navigation:

— an aggregated `metadata.csv` listing all mice and acquisitions,

— an illustrative density map in `images/` for rapid browsing,

— a human-readable `report.md` documenting data organization and provenance.

This structure allows users to explore example outputs, inspect metadata without downloading the raw FRDR archive.

```
ulmshare/
├── examples_ulm/
│   └── Mouse_XX/
│       ├── acquisition_01/
│       │   ├── density_map.png
│       │   ├── tracks.json.xz
│       │   └── metrics.json
│       ├── acquisition_02/
│       └── ..
├── code/
│   ├── matlab/
│   │   ├── load_raw_data.m
│   │   ├── example_processing_script.m
│   │   └── utils/
│   └── README.md
└── summary/
    ├── images/
    ├── metadata.csv
    └── report.md
```

Figure 6.3 Directory structure of ULM examples, summary, and helper codes hosted in the GitHub repository.

## 6.5 Technical Validation

To assess the technical quality and reproducibility of ULMShare, we performed quantitative and qualitative evaluations across all acquisitions. These analyses help readers evaluate the suitability of each acquisition for their intended applications. In addition to many high-quality examples, the dataset intentionally includes acquisitions of varying quality, reflecting the realistic and heterogeneous conditions of transcranial imaging. This diversity ensures that ULMShare offers both high-quality examples for benchmarking and more challenging cases for the development of robust reconstruction methods.

We evaluated vascular saturation as an indicator of the proportion of perfused vasculature, estimated spatial coherence using Fourier Ring Correlation (FRC) [272], reported average length of detected tracks, and carried out visual inspection of all super-resolution maps.

Together, these metrics provide a comprehensive assessment of acquisition quality and consistency across the dataset. FRC, vascular and mean track length distributions across the dataset are represented in Figure 6.4.

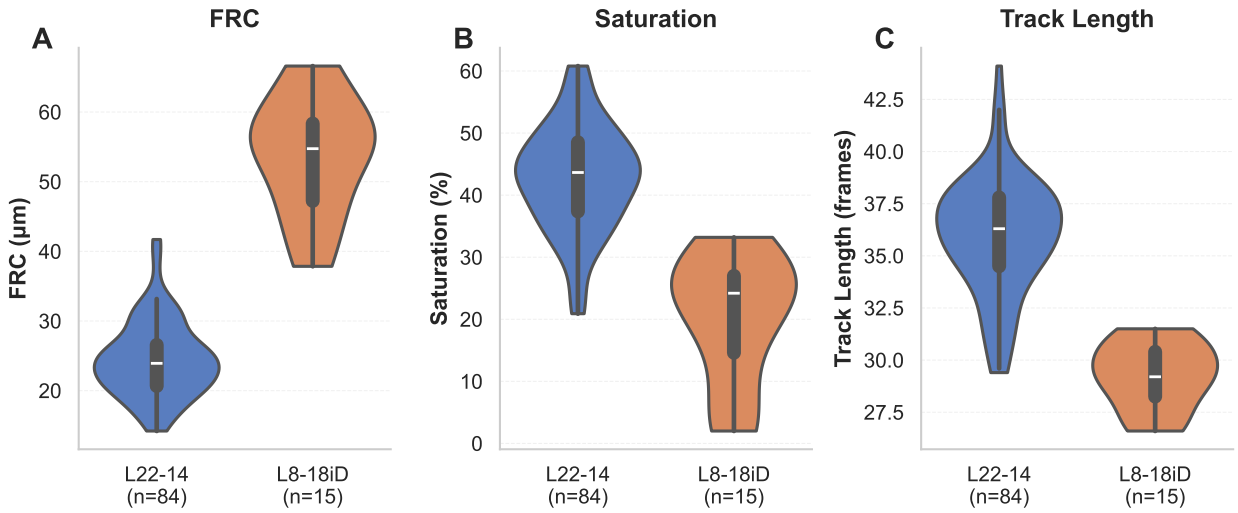



Figure 6.4 Distribution of ULM validation quantitative metrics across probe types. (A) FRC, (B) vascular saturation, and (C) mean track length. Data shown for L22-14 (n=84) and L8-18iD (n=15) probes.The white dot indicating the median, the thick black bar representing the interquartile range, and the thin black line extending to the data extremes.

### 6.5.1 Vascular saturation

Vascular saturation quantifies the fraction of the field of view visited by localized MBs throughout an acquisition. It provides a global measure of microvascular sampling density and reflects MB injection efficiency, perfusion, imaging depth, and potential shadowing from the skull.

Across the dataset, saturation values reflected expected differences between probe types. For the L22-14 probe (n=84), vascular saturation ranged from 20.9% to 61.5% (median 43.9%). For the L8-18iD probe (n=15), saturation values were lower, ranging from 0.5% to 33.2% (median 24.2%).

### 6.5.2 FRC - Spatial coherence

Spatial coherence was evaluated using FRC, which measures the frequency-dependent agreement between two statistically independent reconstructions. For each acquisition, the localized MBs were randomly split into two halves, reconstructed independently, and the FRC

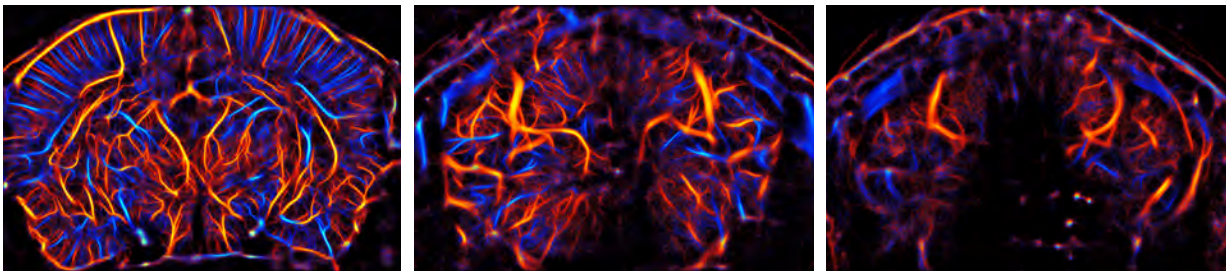
curve computed following Hingot *et al.* [76]. We report the spatial frequency at which the FRC curve crosses the half-bit threshold, which provides a data-driven estimate of the highest spatial frequency that is coherently sampled.

Across the dataset, FRC half-bit spatial coherence scales differed markedly between probes. For the L22-14 probe, values ranged from 14.21 to 41.70 $\mu$m (median 23.71$\mu$m $\sim \lambda/4$). For the L8-18iD probe, coherence scales were broader, ranging from 40.16 to 66.60 $\mu$m (median 54.74 $\mu$m $\sim \lambda/3$). This change reflects the larger wavelength associated with the lower center frequency.

### 6.5.3 Track length - Temporal coherence

Across acquisitions, track-length statistics exhibited moderate variability. For the L22-14v probe, mean track lengths ranged from 29.40 to 44.10 frames (median 36.30), whereas for the L8-18iD probe they ranged from 26.80 to 31.50 frames (median 29.20). Longer and more continuous trajectories generally indicate higher confidence in MB localization, as they require consistent detection across consecutive frames and are less likely to arise from noise, clutter, or spurious localizations [73].

### 6.5.4 Qualitative inspection



Grade A Grade B Grade C

Figure 6.5 Representative examples of ULM reconstructions (red: flowing downward; blue: flowing upward) illustrating the range of acquisition quality in the ULMShare dataset.

All acquisitions were visually inspected by two experienced researchers and assigned to a three-level quality scale based on the visibility of fine vascular structures—defined here as vessels with diameters below a fraction of the imaging wavelength (typically $< 0.1$–$0.3\lambda$). Grade A (high quality) corresponded to images in which these fine vessels were clearly visible and consistently delineated across most of the field of view, forming continuous microvascular patterns. Grade B (medium quality) was assigned when fine vessels were visible only in

certain regions, resulting in heterogeneous small-scale vascular detail. Grade C (low quality) was used when fine vessels were largely absent or indistinguishable, and the image was dominated by larger vascular structures.

After consensus between graders, 57 acquisitions were classified as A, 24 as B, and 18 as C. Representative examples of these three quality levels are shown in Fig. 6.5. Although lower-quality acquisitions exhibit reduced sampling or less favorable imaging conditions, they remain valuable for developing and benchmarking methods intended to operate robustly under realistic constraints.

## 6.6 Data Availability

The raw dataset, comprising approximately 30 TB of radio-frequency data and metadata from 99 acquisitions, is available via the Federated Research Data Repository (FRDR). Derived ULM maps, microbubble trajectories, and MATLAB helper scripts are hosted on the companion GitHub repository.

## 6.7 Code Availability

To support the use of ULMShare, we provide a set of lightweight MATLAB functions for loading raw data together with the associated `sequence.json` files. An example script illustrating the workflow from raw data to a simple super-resolution map is included. This example relies on existing open-source toolboxes and is intended solely as a demonstration of how to interact with the dataset, not as a reproduction of the ULM maps distributed in the repository.

Full reconstruction of the example ULM maps is not provided, as the internal processing pipeline is hardware-specific, under active development, and not currently suitable for full open-source release. Instead, users are encouraged to rely on established community tools. The MATLAB UltraSound Toolbox (MUST) [131, 270, 271] is used for beamforming, and the Tracking and Localization toolbox [74] is used for MB localization and tracking.

In the provided example, clutter is removed using SVD applied to beamformed data to reduce memory usage and computation time. All processing parameters in the example scripts match those summarized in Table 6.3.

# CHAPITRE 7 CONCLUSION

L'objectif général de cette thèse était d'explorer et de démontrer le potentiel de l'apprentissage profond pour surmonter les limitations intrinsèques de la microscopie de localisation ultrasonore (ULM), notamment, pour réduire le temps d'acquisition et améliorer la reproductibilité ainsi que la robustesse expérimentale.

## 7.1 Synthèse des travaux

Ces travaux s'articulent en quatre contributions majeures, allant de l'analyse théorique à la mise en œuvre pratique et au partage de ressources.

### 7.1.1 Analyse critique et structuration du domaine

Le premier apport de cette thèse (Chapitre 3) a été d'établir une analyse critique et structurée de l'état de l'art. Cette revue de la littérature a permis de cartographier les approches existantes et de souligner la prédominance des entraînements sur données simulées. En formalisant la difficulté à respecter la condition i.i.d. (indépendante et identiquement distribuée) entre les distributions d'entraînement synthétiques et les données expérimentales, elle a identifié le *domain shift* comme un frein majeur à la généralisation *in vivo*. De plus, cette analyse a mis en exergue l'absence d'approches d'apprentissage profond pour l'imagerie 3D ainsi que le manque de socle commun d'évaluation. Ce dernier complique notamment la comparaison objective des méthodes existantes.

### 7.1.2 Passage à l'échelle pour l'ULM 3D par parcimonie

Cette thèse a introduit l'utilisation de réseaux de neurones à tenseur parcimonieux (*Sparse Tensor Neural Networks*) en ULM 3D (Chapitre 4). En exploitant la parcimonie des signaux de microbulles pour réduire la complexité en mémoire de deux ordres de grandeur, cette approche a permis l'application de l'apprentissage profond en ULM volumique. L'évaluation *in silico* a démontré que l'ajout d'une dimension spatiale, combiné à la capacité de modélisation de l'apprentissage profond, permet de mieux séparer les signaux de bulles se chevauchant à haute concentration. Ces résultats ouvrent la voie à une application viable de l'ULM en 3D, promettant une réduction significative du temps d'acquisition et du volume de données générées. Toutefois, bien que la modélisation améliorée des signaux de bulles et les gains en efficacité mémoire soient théoriquement transposables *in vivo*, l'application directe de ce

modèle s'est heurtée à un écart trop important entre les simulations d'entraînement et les données réelles, motivant ainsi l'approche développée dans l'étude suivante.

### 7.1.3 Apprentissage in vivo et robustesse par distillation

Afin de combler l'écart entre simulation et réalité, cette thèse développe une méthode permettant l'entraînement directement *in vivo* via une approche *Teacher-Student* (Chapitre 5), tirant ainsi parti de la grande quantité de données expérimentales disponibles au laboratoire. L'utilisation de ce cadre, couplée à des mécanismes de perturbation, a rendu les modèles robustes aux dégradations du signal d'entrée. D'une part, cette stabilité accrue face à la variabilité expérimentale offre des perspectives prometteuses pour améliorer la reproductibilité des acquisitions ULM. D'autre part, la possibilité de réduire le nombre de canaux de réception (de 128 à 32) sans perte majeure de qualité constitue une piste sérieuse pour alléger les contraintes matérielles de l'ULM 3D. En s'affranchissant totalement des simulations, cette méthode établit un cadre d'entraînement respectant la condition i.i.d., identifiée comme critique dans notre revue de la littérature (Chapitre 3). Ce cadre permet enfin de distinguer les gains de performance liés à l'architecture du réseau de ceux liés à la qualité des simulations, offrant ainsi un standard rigoureux pour l'évaluation de futurs modèles d'apprentissage.

### 7.1.4 Base de donnée massive : ULMShare

Afin de rendre possible l'entraînement et l'évaluation des méthodes d'apprentissage dans des conditions i.i.d., cette thèse présente *ULMShare* (Chapitre 6), la plus grande base de données d'ULM *in vivo* publiée à ce jour. ULMShare constitue un jalon essentiel vers l'établissement de protocoles d'évaluation plus fiables pour l'ensemble des méthodes d'ULM, qu'elles soient basées sur l'apprentissage profond ou non. La diversité des acquisitions fournies permettra de caractériser plus finement les limites des mesures de performance actuelles et d'en développer de plus robustes. Enfin, ULMShare représente une ressource précieuse pour le développement de nouvelles approches, inspirées des travaux du Chapitre 5, visant à pallier plusieurs des limitations que nous explorerons dans la section suivante.

## 7.2 Limitations et recherches futures

Si les travaux présentés dans cette thèse marquent une avancée significative pour l'ULM, ils soulèvent également de nouvelles questions. Cette section discute les limites intrinsèques des solutions proposées et esquisse les pistes de recherche les plus prometteuses pour l'avenir de la modalité.

### 7.2.1 Limitations principales

Nous identifions ici quatre obstacles majeurs qui restreignent la portée actuelle de nos solutions.

#### Représentation des trajectoires de microbulles

Cette thèse s'est focalisée sur l'amélioration ou l'extension de formulations d'entraînements et d'architectures existantes. Cependant, elle n'a pas approfondi l'une des limitations majeures mentionnées au chapitre 3 : l'incorporation du contexte temporel et la représentation des trajectoires de bulles dans l'apprentissage. Actuellement, la majorité des approches, y compris celles présentées ici, séparent la localisation et le suivi (*tracking*) ou écrasent la temporalité des trajectoires. Des architectures capables de traiter la séquence temporelle complète et de représenter les trajectoires sans perte d'information seraient probablement plus performantes. Elles pourraient aussi faciliter la représentation du réseau vasculaire sous la forme d'un graphe. Il serait alors possible d'extraire directement des biomarqueurs complexes comme la vitesse, la tortuosité ou la pulsatilité, sans dépendre de la projection sur une grille et de la segmentation des vaisseaux. Une telle représentation permettrait une analyse approfondie et fiable des réorganisations du réseau vasculaire lors de maladies ou en réponse à des traitements thérapeutiques.

#### Propagation de biais de détection

L'approche TS-ULM propose une solution au défi majeur du transfert des modèles "sparse" 3D vers le *in vivo*. Cependant, l'approche *Teacher-Student* présentée au chapitre 5 repose sur des pseudo-labels générés par une méthode conventionnelle (TAL). Bien que l'approche proposée semble permettre une meilleure reconstruction que l'approche TAL en limitant la variance de l'erreur, il est fortement possible qu'elle en reproduise les biais. L'utilisation de simulation pourrait être un outil essentiel pour identifier la nature et le degré de ces biais. Finalement, l'exploitation d'un consensus entre différentes approches conventionnelles pourrait mitiger ce problème.

#### Défi d'évaluation *in vivo*

Une autre limitation majeure réside dans l'absence de vérité terrain absolue pour les données expérimentales, un défi identifié dès notre revue de littérature (Chapitre 3). Si la base de données *ULMShare* (Chapitre 6) fournit un référentiel de données brutes indispensable pour la communauté, elle ne pallie pas le manque de métriques de référence fiables pour juger de

la qualité d'une reconstruction. Nos travaux sur l'approche *Teacher-Student* (Chapitre 5) ont d'ailleurs mis en lumière les incohérences des métriques actuelles dans des régimes dégradés : nous avons notamment observé des cas paradoxaux présentant une résolution apparente élevée (bon FRC) alors même que la densité vasculaire était manifestement insuffisante (faible saturation). Cette décorrélation confirme que les indicateurs standards actuels ne suffisent pas à garantir la qualité de reconstruction, soulignant le besoin critique de développer de nouvelles métriques d'évaluation plus robustes. Par ailleurs, il semble peu probable de pouvoir quantifier la fidélité de la reconstruction anatomique sans validation croisée avec d'autres modalités.

#### Généralisation inter-organes et inter-espèces

Les développements méthodologiques et la base de données *ULMShare* se sont concentrés principalement sur l'imagerie cérébrale chez la souris. Bien que les principes méthodologiques soient transposables, la généralisation directe des modèles entraînés vers d'autres organes (cœur, rein, foie) ou vers l'application clinique chez l'homme reste à démontrer. Les mouvements physiologiques complexes (respiration, battements cardiaques) présents dans d'autres organes posent des défis supplémentaires de recalage que nos modèles actuels, entraînés sur des acquisitions stéréotaxiques stabilisées, ne gèrent pas explicitement.

### 7.2.2 Perspectives de recherche

Les outils méthodologiques et les ressources développés dans ce manuscrit posent les fondations nécessaires pour étendre les capacités de l'ULM. Nous proposons ici trois axes de recherche prioritaires visant à converger vers une imagerie volumique robuste et accessible.

#### Vers un modèle unifié 3D robuste in vivo

La suite logique de ces travaux consiste à fusionner les approches des chapitres 4 et 5. En appliquant le cadre d'entraînement *Teacher-Student* à des modèles utilisant la parcimonie des trajectoires de bulles, il devient envisageable de créer un modèle unifié capable de traiter des données volumiques *in vivo*. Un tel développement permettrait de gérer les très grands volumes de données générés et d'exploiter des concentrations de microbulles élevées. L'extension vers la 3D de la réduction du nombre d'éléments auraient un impact important sur les couts d'un systeme permettant l'ULM 3D. Une telle avancée, combinant réduction du temps d'acquisition, allègement du matériel et relaxation des contraintes de transfert et de stockage, rendrait l'ULM microscopique 3D nettement plus accessible.

### Extension du cadre Teacher-Student

Le paradigme *Teacher-Student* développé pour la localisation offre un cadre flexible qui pourrait être étendu bien au-delà des perturbations étudiées dans cette thèse. Une perspective majeure serait d'entraîner un modèle étudiant directement sur des données brutes (avant filtrage du signal tissulaire), en utilisant les prédictions d'un enseignant sur des données filtrées comme référence. Cette approche permettrait potentiellement de s'affranchir de l'étape de filtrage SVD, souvent nécessitant le réglage de parametre dépendant de l'acquisition et délétère pour la détection des microbulles les plus lentes. De plus, ce cadre pourrait être adapté pour gérer des concentrations de microbulles encore plus élevées, en apprenant à séparer des sources superposées que les méthodes conventionnelles échouent à distinguer. Enfin, la synergie de multiples perturbations (bruit, décimation, superposition) combinée à l'utilisation de modèles à plus grande capacité pourrait mener à des réseaux extrêmement robustes, capables de reconstruire des images de haute qualité dans des conditions d'acquisition dégradées.

### Réduction des contraintes matérielles pour la clinique et le préclinique

Au-delà de l'amélioration de la qualité d'image, l'apprentissage profond offre une opportunité de démocratiser l'accès à l'ULM. Les résultats du chapitre 5 ont montré qu'un modèle bien entraîné peut maintenir une haute qualité d'image même avec un nombre réduit de canaux de réception. Une perspective majeure est d'exploiter cette propriété pour faciliter le transfert clinique de l'ULM 3D. En couplant l'apprentissage profond avec des sondes matricielles à adressage clairsemé (*sparse arrays*), il deviendrait envisageable de réaliser de l'imagerie microvasculaire volumique avec des échographes avec moins de voies d'acquisition, réduisant ainsi drastiquement le coût et la complexité technique du système. Parallèlement, ces travaux ouvrent la voie à une démocratisation de l'ULM 2D dans les études précliniques. En rendant possible l'imagerie avec seulement 32 canaux, cette méthode positionne l'ULM comme une modalité d'imagerie transcrânienne haute résolution accessible à des laboratoires ne disposant pas d'échographes de recherche haut de gamme.

## 7.3 Conclusion générale

Cette thèse démontre que le potentiel de l'apprentissage profond en ULM est important pour en dépasser les limites intrinséques en temps d'acquisition et en améliorer la robustesse et la reproductabilité. En passant de la simulation à la réalité des données *in vivo* et du plan au volume, nous avons posé les fondations pour que l'apprentissage profond joue un rôle crucial dans le développement de l'ULM.

# RÉFÉRENCES


[1] B. Rauby, P. Xing, J. Porée, M. Gasse et J. Provost, "Pruning Sparse Tensor Neural Networks Enables Deep Learning for 3D Ultrasound Localization Microscopy," *IEEE Transactions on Image Processing*, vol. 34, p. 2367–2378, 2025.

[2] 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, n°. 1, p. 2932, avr. 2024.

[3] R. Nortley *et al.*, "Amyloid $\beta$ oligomers constrict human capillaries in Alzheimer's disease via signaling to pericytes," *Science*, vol. 365, n°. 6450, p. eaav9518, juill. 2019.

[4] P. Carmeliet et R. K. Jain, "Angiogenesis in cancer and other diseases," *Nature*, vol. 407, n°. 6801, p. 249–257, sept. 2000.

[5] A. Liesz, "The vascular side of Alzheimer's disease," *Science*, vol. 365, n°. 6450, p. 223–224, juill. 2019.

[6] C. Iadecola, "The Neurovascular Unit Coming of Age : A Journey through Neurovascular Coupling in Health and Disease," *Neuron*, vol. 96, n°. 1, p. 17–42, sept. 2017.

[7] R. Crumpler, R. J. Roman et F. Fan, "Capillary Stalling : A Mechanism of Decreased Cerebral Blood Flow in AD/ADRD," *Journal of experimental neurology*, vol. 2, n°. 4, p. 149–153, 2021.

[8] P. Song, J. M. Rubin et M. R. Lowerison, "Super-resolution ultrasound microvascular imaging : Is it ready for clinical use ?" *Zeitschrift für Medizinische Physik*, vol. 33, n°. 3, p. 309–323, août 2023.

[9] A. Y. Shih, J. D. Driscoll, P. J. Drew, N. Nishimura, C. B. Schaffer et D. Kleinfeld, "Two-photon microscopy as a tool to study blood flow and neurovascular coupling in the rodent brain," *Journal of Cerebral Blood Flow and Metabolism : Official Journal of the International Society of Cerebral Blood Flow and Metabolism*, vol. 32, n°. 7, p. 1277–1309, juill. 2012.

[10] H. G. Bezerra, M. A. Costa, G. Guagliumi, A. M. Rollins et D. I. Simon, "Intracoronary optical coherence tomography : A comprehensive review clinical and research applications," *JACC. Cardiovascular interventions*, vol. 2, n°. 11, p. 1035–1046, nov. 2009.

[11] O. Couture, V. Hingot, B. Heiles, P. Muleki-Seya et M. Tanter, "Ultrasound Localization Microscopy and Super-Resolution : A State of the Art," *IEEE Transactions on*

*Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, n°. 8, p. 1304–1320, août 2018.

[12] M. Tanter et M. Fink, "Ultrafast imaging in biomedical ultrasound," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 61, n°. 1, p. 102–119, janv. 2014.

[13] C. Errico *et al.*, "Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging," *Nature*, vol. 527, n°. 7579, p. 499–502, nov. 2015.

[14] K. Christensen-Jeffries, R. J. Browning, M. Tang, C. Dunsby et R. J. Eckersley, "In Vivo Acoustic Super-Resolution and Super-Resolved Velocity Mapping Using Microbubbles," *IEEE Transactions on Medical Imaging*, vol. 34, n°. 2, p. 433–440, févr. 2015.

[15] O. M. Viessmann, R. J. Eckersley, K. Christensen-Jeffries, M. X. Tang et C. Dunsby, "Acoustic super-resolution with ultrasound and microbubbles," *Physics in Medicine and Biology*, vol. 58, n°. 18, p. 6447–6458, sept. 2013.

[16] Y. Desailly, O. Couture, M. Fink et M. Tanter, "Sono-activated ultrasound localization microscopy," *Applied Physics Letters*, vol. 103, n°. 17, p. 174107, oct. 2013.

[17] K. Christensen-Jeffries *et al.*, "Super-resolution Ultrasound Imaging," *Ultrasound in Medicine & Biology*, vol. 46, n°. 4, p. 865–891, avr. 2020.

[18] A. Chavignon, V. Hingot, C. Orset, D. Vivien et O. Couture, "3D transcranial ultrasound localization microscopy for discrimination between ischemic and hemorrhagic stroke in early phase," *Scientific Reports*, vol. 12, n°. 1, p. 14607, août 2022.

[19] M. R. Lowerison *et al.*, "Super-Resolution Ultrasound Reveals Cerebrovascular Impairment in a Mouse Model of Alzheimer's Disease," *Journal of Neuroscience*, vol. 44, n°. 9, févr. 2024.

[20] ——, "Aging-related cerebral microvascular changes visualized using ultrasound localization microscopy in the living mouse," *Scientific Reports*, vol. 12, n°. 1, p. 619, déc. 2022.

[21] T. Opacic *et al.*, "Motion model ultrasound localization microscopy for preclinical and clinical multiparametric tumor characterization," *Nature Communications*, vol. 9, n°. 1, p. 1527, avr. 2018.

[22] L. Denis *et al.*, "Sensing ultrasound localization microscopy for the visualization of glomeruli in living rats and humans," *eBioMedicine*, vol. 91, p. 104578, avr. 2023.

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

[24] C. Bourquin, J. Porée, F. Lesage et J. Provost, “In Vivo Pulsatility Measurement of Cerebral Microcirculation in Rodents Using Dynamic Ultrasound Localization Microscopy,” *IEEE Transactions on Medical Imaging*, vol. 41, n°. 4, p. 782–792, avr. 2022.

[25] C. Bourquin *et al.*, “Quantitative pulsatility measurements using 3D dynamic ultrasound localization microscopy,” *Physics in Medicine & Biology*, vol. 69, n°. 4, p. 045017, févr. 2024.

[26] N. Renaudin, C. Demené, A. Dizeux, N. Ialy-Radio, S. Pezet et M. Tanter, “Functional ultrasound localization microscopy reveals brain-wide neurovascular activity on a microscopic scale,” *Nature Methods*, vol. 19, n°. 8, p. 1004–1012, août 2022.

[27] 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, n°. 1, p. 57–66, janv. 2024.

[28] 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, n°. 8, p. 08NT01, avr. 2021.

[29] S. Bodard *et al.*, “Ultrasound localization microscopy of the human kidney allograft on a clinical ultrasound scanner,” *Kidney International*, vol. 103, n°. 5, p. 930–935, mai 2023.

[30] J. Yan *et al.*, “Transthoracic ultrasound localization microscopy of myocardial vasculature in patients,” *Nature Biomedical Engineering*, p. 1–12, mai 2024.

[31] C. Demené *et al.*, “Transcranial ultrafast ultrasound localization microscopy of brain vasculature in patients,” *Nature biomedical engineering*, vol. 5, n°. 3, p. 219–228, mars 2021.

[32] V. Hingot, C. Errico, B. Heiles, L. Rahal, M. Tanter et O. Couture, “Microvascular flow dictates the compromise between spatial resolution and acquisition time in Ultrasound Localization Microscopy,” *Scientific Reports*, vol. 9, n°. 1, p. 2456, févr. 2019.

[33] 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, n°. 5, p. 803–814, mai 2018.

[34] V. Hingot, C. Errico, M. Tanter et O. Couture, “Subwavelength motion-correction for ultrafast ultrasound localization microscopy,” *Ultrasonics*, vol. 77, p. 17–21, mai 2017.

[35] B. Heiles *et al.*, “Ultrafast 3D Ultrasound Localization Microscopy Using a 32×32 Matrix Array,” *IEEE Transactions on Medical Imaging*, vol. 38, n°. 9, p. 2005–2015, sept. 2019.

[36] A. Chavignon, B. Heiles, V. Hingot, C. Orset, D. Vivien et O. Couture, "3D Transcranial Ultrasound Localization Microscopy in the Rat Brain with a Multiplexed Matrix Probe," *IEEE transactions on bio-medical engineering*, vol. PP, déc. 2021.

[37] 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, p. 261–282, 2022.

[38] I. Goodfellow, Y. Bengio et A. Courville, *Deep Learning*, ser. Adaptive Computation and Machine Learning Series, F. Bach, édit. Cambridge, MA, USA : MIT Press, nov. 2016.

[39] A. Krizhevsky, I. Sutskever et G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, n°. 6, p. 84–90, mai 2017.

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

[41] J. Ma, Y. He, F. Li, L. Han, C. You et B. Wang, "Segment anything in medical images," *Nature Communications*, vol. 15, n°. 1, p. 654, janv. 2024.

[42] R. J. G. van Sloun, R. Cohen et Y. C. Eldar, "Deep Learning in Ultrasound Imaging," *Proceedings of the IEEE*, vol. 108, n°. 1, p. 11–29, janv. 2020.

[43] O. Solomon *et al.*, "Deep Unfolded Robust PCA With Application to Clutter Suppression in Ultrasound," *IEEE Transactions on Medical Imaging*, vol. 39, n°. 4, p. 1051–1063, avr. 2020.

[44] R. J. G. van Sloun *et al.*, "Super-Resolution Ultrasound Localization Microscopy Through Deep Learning," *IEEE Transactions on Medical Imaging*, vol. 40, n°. 3, p. 829–839, mars 2021.

[45] L. Milecki *et al.*, "A Deep Learning Framework for Spatiotemporal Ultrasound Localization Microscopy," *IEEE Transactions on Medical Imaging*, vol. 40, n°. 5, p. 1428–1437, mai 2021.

[46] X. Liu, T. Zhou, M. Lu, Y. Yang, Q. He et J. Luo, "Deep Learning for Ultrasound Localization Microscopy," *IEEE Transactions on Medical Imaging*, vol. 39, n°. 10, p. 3064–3078, oct. 2020.

[47] J. Youn, M. L. Ommen, M. B. Stuart, E. V. Thomsen, N. B. Larsen et J. A. Jensen, "Detection and Localization of Ultrasound Scatterers Using Convolutional Neural Networks," *IEEE Transactions on Medical Imaging*, vol. 39, n°. 12, p. 3855–3867, déc. 2020.

[48] K. G. Brown, D. Ghosh et K. Hoyt, “Deep Learning of Spatiotemporal Filtering for Fast Super-Resolution Ultrasound Imaging,” *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 67, n°. 9, p. 1820–1829, sept. 2020.

[49] B. Rauby, P. Xing, M. Gasse et J. Provost, “Deep Learning in Ultrasound Localization Microscopy : Applications and Perspectives,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 71, n°. 12 : Breaking the Resolution Barrier in Ultrasound, p. 1765–1784, déc. 2024.

[50] M. Lerendegui *et al.*, “ULTRA-SR Challenge : Assessment of Ultrasound Localization and TRacking Algorithms for Super-Resolution Imaging,” *IEEE Transactions on Medical Imaging*, p. 1–1, 2024.

[51] D. D. Gutterman *et al.*, “The Human Microcirculation : Regulation of Flow and Physiological Functions,” *Circulation Research*, vol. 118, n°. 1, p. 157–172, janv. 2016.

[52] M. Brundel, J. de Bresser, J. J. van Dillen, L. J. Kappelle et G. J. Biessels, “Cerebral microinfarcts : A systematic review of neuropathological studies,” *Journal of Cerebral Blood Flow & Metabolism*, vol. 32, n°. 3, p. 425–436, mars 2012.

[53] K. A. Jellinger, “The pathology of "vascular dementia" : A critical update,” *Journal of Alzheimer’s disease : JAD*, vol. 14, n°. 1, p. 107–123, mai 2008.

[54] P. G. Camici et F. Crea, “Coronary Microvascular Dysfunction,” *New England Journal of Medicine*, vol. 356, n°. 8, p. 830–840, févr. 2007.

[55] E. M. Haacke, R. W. Brown, M. R. Thompson et R. Venkatesan, *Magnetic Resonance Imaging : Physical Principles and Sequence Design*. John Wiley & Sons, 1999.

[56] S. Trattnig *et al.*, “Clinical 7 T MRI : Are we there yet ? A review about magnetic resonance imaging at ultra-high field,” *European Radiology*, vol. 28, n°. 9, p. 3610–3627, sept. 2018.

[57] J. T. Bushberg, J. A. Seibert, E. M. Leidholdt et J. M. Boone, *The Essential Physics of Medical Imaging*, 4e éd. Lippincott Williams & Wilkins, 2020.

[58] P. Doyle, K. Mroz, A. Jodko-Wladzinska, A. Pawelec, A. Kotecki et P. Skrzypczyk, “Radiation Doses in Cardiovascular Computed Tomography : A Review,” *Journal of Clinical Medicine*, vol. 12, n°. 8, p. 2968, avr. 2023.

[59] W. Denk, J. H. Strickler et W. W. Webb, “Two-photon laser scanning fluorescence microscopy,” *Science*, vol. 248, n°. 4951, p. 73–76, 1990.

[60] J. Huisken, J. Swoger, F. Del Bene, J. Wittbrodt et E. H. Stelzer, “Optical sectioning deep inside live embryos by selective plane illumination microscopy,” *Science*, vol. 305, n°. 5686, p. 1007–1009, 2004.

[61] V. Ntziachristos, “Going deeper than microscopy : The optical imaging frontier in biology,” *Nature Methods*, vol. 7, n°. 8, p. 603–614, août 2010.

[62] D. W. Holdsworth et M. M. Thornton, “Micro-CT in small animal imaging,” *Trends in Biotechnology*, vol. 20, n°. 8, p. S34–S39, oct. 2002.

[63] E. Betzig *et al.*, “Imaging intracellular fluorescent proteins at nanometer resolution,” *Science (New York, N.Y.)*, vol. 313, n°. 5793, p. 1642–1645, sept. 2006.

[64] M. J. Rust, M. Bates et X. Zhuang, “Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM),” *Nature Methods*, vol. 3, n°. 10, p. 793–795, oct. 2006.

[65] T. L. Szabo, “Chapter 14 - Ultrasound Contrast Agents,” dans *Diagnostic Ultrasound Imaging : Inside Out (Second Edition)*, T. L. Szabo, édit. Boston : Academic Press, janv. 2014, p. 605–651.

[66] M. Versluis, E. Stride, G. Lajoinie, B. Dollet et T. Segers, “Ultrasound Contrast Agent Modeling : A Review,” *Ultrasound in Medicine & Biology*, vol. 46, n°. 9, p. 2117–2144, sept. 2020.

[67] B. Heiles, A. Chavignon, V. Hingot, P. Lopez, E. Teston et O. Couture, “Performance benchmarking of microbubble-localization algorithms for ultrasound localization microscopy,” *Nature Biomedical Engineering*, vol. 6, n°. 5, p. 605–616, mai 2022.

[68] J. Bercoff *et al.*, “Ultrafast compound doppler imaging : Providing full blood flow characterization,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 58, n°. 1, p. 134–147, janv. 2011.

[69] 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, n°. 2, p. 662–673, févr. 2024.

[70] G. Montaldo, M. Tanter, J. Bercoff, N. Benech et 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, n°. 3, p. 489–506, mars 2009.

[71] J. Brown, K. Christensen-Jeffries, S. Harput, C. Dunsby, M. X. Tang et R. J. Eckersley, “Investigation of microbubble detection methods for super-resolution imaging of microvasculature,” dans *2017 IEEE International Ultrasonics Symposium (IUS)*, sept. 2017, p. 1–4.

[72] C. Demené *et al.*, “Spatiotemporal Clutter Filtering of Ultrafast Ultrasound Data Highly Increases Doppler and fUltrasound Sensitivity,” *IEEE Transactions on Medical Imaging*, vol. 34, n°. 11, p. 2271–2285, nov. 2015.

[73] S. A. Lee *et al.*, “Functional Assessment of Cerebral Capillaries using Single Capillary Reporters in Ultrasound Localization Microscopy,” *arXiv :2407.07857*, juill. 2024.

[74] A. Leconte *et al.*, “A Tracking Prior to Localization Workflow for Ultrasound Localization Microscopy,” *IEEE Transactions on Medical Imaging*, vol. 44, n°. 2, p. 698–710, févr. 2025.

[75] X. Chen, M. R. Lowerison, Z. Dong, N. V. C. Sekaran, D. A. Llano et P. Song, “Localization free super-resolution microbubble velocimetry using a long short-term memory neural network,” *IEEE Transactions on Medical Imaging*, p. 1–1, 2023.

[76] V. Hingot, A. Chavignon, B. Heiles et O. Couture, “Measuring Image Resolution in Ultrasound Localization Microscopy,” *IEEE Transactions on Medical Imaging*, vol. 40, n°. 12, p. 3812–3819, déc. 2021.

[77] S. Schwarz *et al.*, “Ultrasound Super-Resolution Imaging of Neonatal Cerebral Vascular Reorganization,” *Advanced Science*, vol. 12, n°. 12, p. 2415235, 2025.

[78] P. Cormier, J. Porée, C. Bourquin et J. Provost, “Dynamic Myocardial Ultrasound Localization Angiography,” *IEEE Transactions on Medical Imaging*, p. 1–1, 2021.

[79] B. Beliard *et al.*, “Ultrafast Doppler imaging and ultrasound localization microscopy reveal the complexity of vascular rearrangement in chronic spinal lesion,” *Scientific Reports*, vol. 12, n°. 1, p. 6574, avr. 2022.

[80] Q.-Q. Zeng *et al.*, “Focal liver lesions : Multiparametric microvasculature characterization via super-resolution ultrasound imaging,” *European Radiology Experimental*, vol. 8, n°. 1, p. 138, déc. 2024.

[81] 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, n°. 2, p. 20230093, 2024.

[82] A. Wu *et al.*, “3D transcranial Dynamic Ultrasound Localization Microscopy in the mouse brain using a Row-Column Array,” juin 2024.

[83] R. M. Jones *et al.*, “Non-invasive 4D transcranial functional ultrasound and ultrasound localization microscopy for multimodal imaging of neurovascular response,” *Scientific Reports*, vol. 14, n°. 1, p. 30240, déc. 2024.

[84] N. Ghigo *et al.*, “Dynamic Ultrasound Localization Microscopy Without ECG-Gating,” *Ultrasound in Medicine and Biology*, vol. 50, n°. 9, p. 1436–1448, sept. 2024.

[85] O. Demeulenaere *et al.*, “In vivo whole brain microvascular imaging in mice using transcranial 3D Ultrasound Localization Microscopy,” *eBioMedicine*, vol. 79, mai 2022.

[86] P. Xing, V. Perrot, A. U. Dominguez-Vargas, S. Quessy, N. Dancause et J. Provost, "Towards Transcranial 3D Ultrasound Localization Microscopy of the Nonhuman Primate Brain," avr. 2024.

[87] 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, n°. 7, p. 1045–1057, juill. 2024.

[88] J. A. Jensen *et al.*, "Anatomic and Functional Imaging Using Row–Column Arrays," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, n°. 10, p. 2722–2738, oct. 2022.

[89] 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, n°. 6, p. 592–598, déc. 2022.

[90] O. Solomon, R. J. G. van Sloun, H. Wijkstra, M. Mischi et Y. C. Eldar, "Exploiting Flow Dynamics for Superresolution in Contrast-Enhanced Ultrasound," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, n°. 10, p. 1573–1586, oct. 2019.

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

[92] R. S. Sutton. (2019) The bitter lesson. [En ligne]. Disponible : www.incompleteideas.net/IncIdeas/BitterLesson.html

[93] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, n°. 3, p. 211–252, déc. 2015.

[94] K. He, X. Zhang, S. Ren et J. Sun, "Delving Deep into Rectifiers : Surpassing Human-Level Performance on ImageNet Classification," dans *2015 IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile : IEEE, déc. 2015, p. 1026–1034.

[95] S. Ren, K. He, R. Girshick et J. Sun, "Faster R-CNN : Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, n°. 6, p. 1137–1149, juin 2017.

[96] K. He, G. Gkioxari, P. Dollar et R. Girshick, "Mask R-CNN," dans *Proceedings of the IEEE International Conference on Computer Vision*, 2017, p. 2961–2969.

[97] I. Goodfellow *et al.*, "Generative adversarial networks," *Communications of the ACM*, vol. 63, n°. 11, p. 139–144, oct. 2020.

[98] J. Ho, A. Jain et P. Abbeel, “Denoising Diffusion Probabilistic Models,” dans *Advances in Neural Information Processing Systems*, vol. 33. Curran Associates, Inc., 2020, p. 6840–6851.

[99] D. P. Kingma et J. Ba, “Adam : A method for stochastic optimization,” 2017.

[100] I. Loshchilov et F. Hutter, “Decoupled weight decay regularization,” 2019.

[101] J. L. Ba, J. R. Kiros et G. E. Hinton, “Layer normalization,” 2016.

[102] Y. Wu et K. He, “Group normalization,” 2018.

[103] V. Nair et G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” dans *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, J. Fürnkranz et T. Joachims, édit., 2010, p. 807–814.

[104] P. Ramachandran, B. Zoph et Q. V. Le, “Searching for activation functions,” 2017.

[105] F. Gers, J. Schmidhuber et F. Cummins, “Learning to forget : continual prediction with lstm,” dans *1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)*, vol. 2, 1999, p. 850–855 vol.2.

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

[107] A. Dosovitskiy *et al.*, “An Image is Worth 16x16 Words : Transformers for Image Recognition at Scale,” dans *International Conference on Learning Representations*, oct. 2020.

[108] R. Bommasani *et al.*, “On the Opportunities and Risks of Foundation Models,” juill. 2022.

[109] M. Awais *et al.*, “Foundational Models Defining a New Era in Vision : A Survey and Outlook,” juill. 2023.

[110] A. Kirillov *et al.*, “Segment Anything,” avr. 2023.

[111] D. Hyun, L. L. Brickson, K. T. Looby et J. J. Dahl, “Beamforming and Speckle Reduction Using Neural Networks,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, nº. 5, p. 898–910, mai 2019.

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

[113] A. C. Luchies et B. C. Byram, “Deep Neural Networks for Ultrasound Beamforming,” *IEEE Transactions on Medical Imaging*, vol. 37, nº. 9, p. 2010–2021, sept. 2018.

[114] H. Strohm, S. Rothlübbers, K. Eickel et M. Günther, "Deep learning-based reconstruction of ultrasound images from raw channel data," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, n°. 9, p. 1487–1490, 2020.

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

[116] B. Luijten, N. Chennakeshava, Y. C. Eldar, M. Mischi et R. J. G. van Sloun, "Ultrasound Signal Processing : From Models to Deep Learning," *Ultrasound in Medicine and Biology*, vol. 49, n°. 3, p. 677–698, mars 2023.

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

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

[119] Z. Zhang, M. Hwang, T. J. Kilbaugh et J. Katz, "Improving sub-pixel accuracy in ultrasound localization microscopy using supervised and self-supervised deep learning," *Measurement Science and Technology*, vol. 35, n°. 4, p. 045701, avr. 2024.

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

[121] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira et J. W. Vaughan, "A theory of learning from different domains," *Machine Learning*, vol. 79, n°. 1-2, p. 151–175, mai 2010.

[122] B. H. Menze *et al.*, "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, n°. 10, p. 1993–2024, oct. 2015.

[123] 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, n°. 11, p. 2514–2525, nov. 2018.

[124] 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, p. 1–13, 2023.

[125] N. Blanken, J. M. Wolterink, H. Delingette, C. Brune, M. Versluis et G. Lajoinie, "Super-Resolved Microbubble Localization in Single-Channel Ultrasound RF Signals Using Deep Learning," *IEEE Transactions on Medical Imaging*, p. 1–1, 2022.

[126] X. Chen, M. R. Lowerison, Z. Dong, A. Han et P. Song, “Deep Learning-Based Microbubble Localization for Ultrasound Localization Microscopy,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, n°. 4, p. 1312–1325, avr. 2022.

[127] J. A. Jensen, “Field : A Program for Simulating Ultrasound Systems : 10th Nordic-Baltic Conference on Biomedical Imaging,” *Medical & Biological Engineering & Computing*, vol. 34, n°. sup. 1, p. 351–353, 1997.

[128] W. Gu, Z. Yan, B. Li, C. Liu, D. Ta et X. Liu, “GAN-Based Ultrasound Localization Microscopy,” dans *2022 IEEE International Ultrasonics Symposium (IUS)*, oct. 2022, p. 1–4.

[129] X. Liu et M. Almekkawy, “Ultrasound Localization Microscopy Using Deep Neural Network,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 70, n°. 7, p. 625–635, juill. 2023.

[130] S. Luan *et al.*, “Deep learning for fast super-resolution ultrasound microvessel imaging,” *Physics in Medicine & Biology*, vol. 68, n°. 24, p. 245023, déc. 2023.

[131] 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, mai 2022.

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

[133] X. Yu *et al.*, “Deep learning for fast denoising filtering in ultrasound localization microscopy,” *Physics in Medicine & Biology*, vol. 68, n°. 20, p. 205002, oct. 2023.

[134] 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, n°. 7, p. 075005, mars 2021.

[135] X. Liu et M. Almekkawy, “Ultrasound Super Resolution using Vision Transformer with Convolution Projection Operation,” dans *2022 IEEE International Ultrasonics Symposium (IUS)*, oct. 2022, p. 1–4.

[136] 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, n°. 6, p. 2551–2562, nov. 2019.

[137] M. Lerendegui, K. Riemer, B. Wang, C. Dunsby et M.-X. Tang, “BUbble Flow Field : A Simulation Framework for Evaluating Ultrasound Localization Microscopy Algo-

rithms," nov. 2022.

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

[139] B. Mehta, M. Diaz, F. Golemo, C. J. Pal et L. Paull, "Active Domain Randomization," dans *Proceedings of the Conference on Robot Learning*. PMLR, mai 2020, p. 1162–1176.

[140] B. E. Treeby, J. Budisky, E. S. Wise, J. Jaros et B. T. Cox, "Rapid calculation of acoustic fields from arbitrary continuous-wave sources," *The Journal of the Acoustical Society of America*, vol. 143, n°. 1, p. 529–537, janv. 2018.

[141] 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, n°. 6, p. 3499–3505, déc. 2005.

[142] J. N. Harmon, Z. Z. Khaing, J. E. Hyde, C. P. Hofstetter, C. Tremblay-Darveau et M. F. Bruce, "Quantitative tissue perfusion imaging using nonlinear ultrasound localization microscopy," *Scientific Reports*, vol. 12, n°. 1, p. 21943, déc. 2022.

[143] D. T. Blackstock, "Generalized Burgers equation for plane waves," *The Journal of the Acoustical Society of America*, vol. 77, n°. 6, p. 2050–2053, juin 1985.

[144] J. Jiménez-Fernández, "Nonlinear response to ultrasound of encapsulated microbubbles," *Ultrasonics*, vol. 52, n°. 6, p. 784–793, août 2012.

[145] W. T. Shi et F. Forsberg, "Ultrasonic characterization of the nonlinear properties of contrast microbubbles," *Ultrasound in Medicine & Biology*, vol. 26, n°. 1, p. 93–104, janv. 2000.

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

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

[148] 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, n°. 2, p. 523–533, févr. 2022.

[149] 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*, p. 1–1, 2024.

[150] G. Ng, S. Worrell, P. Freiburger et G. Trahey, "A comparative evaluation of several algorithms for phase aberration correction," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 41, n°. 5, p. 631–643, sept. 1994.

[151] L. Nock, G. E. Trahey et 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, n°. 5, p. 1819–1833, mai 1989.

[152] M. O'Donnell et S. Flax, "Phase-aberration correction using signals from point reflectors and diffuse scatterers : Measurements," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 35, n°. 6, p. 768–774, nov. 1988.

[153] G. Montaldo, M. Tanter et M. Fink, "Time Reversal of Speckle Noise," *Physical Review Letters*, vol. 106, n°. 5, p. 054301, févr. 2011.

[154] B.-F. Osmanski, G. Montaldo, M. Tanter et M. Fink, "Aberration correction by time reversal of moving speckle noise," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 59, n°. 7, p. 1575–1583, juill. 2012.

[155] W. Lambert, L. A. Cobus, M. Couade, M. Fink et A. Aubry, "Reflection Matrix Approach for Quantitative Imaging of Scattering Media," *Physical Review X*, vol. 10, n°. 2, p. 021048, juin 2020.

[156] W. Lambert, L. A. Cobus, T. Frappart, M. Fink et A. Aubry, "Distortion matrix approach for ultrasound imaging of random scattering media," *Proceedings of the National Academy of Sciences*, vol. 117, n°. 26, p. 14 645–14 656, juin 2020.

[157] W. Lambert, L. A. Cobus, J. Robin, M. Fink et A. Aubry, "Ultrasound Matrix Imaging—Part II : The Distortion Matrix for Aberration Correction Over Multiple Isoplanatic Patches," *IEEE Transactions on Medical Imaging*, vol. 41, n°. 12, p. 3921–3938, déc. 2022.

[158] M. Feigin, D. Freedman et B. W. Anthony, "A Deep Learning Framework for Single-Sided Sound Speed Inversion in Medical Ultrasound," *IEEE transactions on bio-medical engineering*, vol. 67, n°. 4, p. 1142–1151, avr. 2020.

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

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

[161] Z. Tian, M. Olmstead, Y. Jing et 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, n°. 1, p. 117–126, janv. 2024.

[162] J. R. McCall, F. Santibanez, H. Belgharbi, G. F. Pinton et P. A. Dayton, "Non-invasive transcranial volumetric ultrasound localization microscopy of the rat brain with continuous, high volume-rate acquisition," *Theranostics*, vol. 13, n°. 4, p. 1235–1246, févr. 2023.

[163] J. Robin *et al.*, "In vivo adaptive focusing for clinical contrast-enhanced transcranial ultrasound imaging in human," *Physics in Medicine & Biology*, vol. 68, n°. 2, p. 025019, janv. 2023.

[164] P. Xing, A. Malescot, E. Martineau, R. Rungta et J. Provost, "Inverse Problem Based on a Sparse Representation of Contrast-enhanced Ultrasound Data for in vivo Transcranial Imaging," janv. 2024.

[165] A. Beck et M. Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," *SIAM Journal on Imaging Sciences*, vol. 2, n°. 1, p. 183–202, janv. 2009.

[166] M. A. O'Reilly et K. Hynynen, "A super-resolution ultrasound method for brain vascular mapping," *Medical Physics*, vol. 40, n°. 11, p. 110701, nov. 2013.

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

[168] J. Foiret, H. Zhang, T. Ilovitsh, L. Mahakian, S. Tam et K. W. Ferrara, "Ultrasound localization microscopy to image and assess microvasculature in a rat kidney," *Scientific Reports*, vol. 7, n°. 1, p. 13662, oct. 2017.

[169] 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, n°. 5, p. 957–967, mai 2020.

[170] A. Coudert *et al.*, "3D Transcranial ultrasound localization microscopy reveals major arteries in the sheep brain," mars 2024.

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

[172] 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, n°. 2, p. 149–167, févr. 2018.

[173] R. Gu *et al.*, "Contrastive Semi-Supervised Learning for Domain Adaptive Segmentation Across Similar Anatomical Structures," *IEEE Transactions on Medical Imaging*, vol. 42, n°. 1, p. 245–256, janv. 2023.

[174] 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, n°. 9, p. 1738–1751, sept. 2020.

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

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

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

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

[179] Y. Sui, X. Guo, J. Yu, D. Ta et K. Xu, "Generative Adversarial Nets for Ultrafast Ultrasound Localization Microscopy Reconstruction," dans *2022 IEEE International Ultrasonics Symposium (IUS)*, oct. 2022, p. 1–4.

[180] T. Tong, G. Li, X. Liu et Q. Gao, "Image Super-Resolution Using Dense Skip Connections," dans *2017 IEEE International Conference on Computer Vision (ICCV)*. Venice : IEEE, oct. 2017, p. 4809–4817.

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

[182] C. Huang *et al.*, "Short Acquisition Time Super-Resolution Ultrasound Microvessel Imaging via Microbubble Separation," *Scientific Reports*, vol. 10, n°. 1, p. 6007, avr. 2020.

[183] B. Heiles, A. Chavignon, V. Hingot, P. Lopez, E. Teston et O. Couture, "Addendum : Performance benchmarking of microbubble-localization algorithms for ultrasound loca-

lization microscopy," *Nature Biomedical Engineering*, p. 1–1, oct. 2023.

[184] G. Tuccio, S. Afrakhteh, G. Iacca et L. Demi, "Time Efficient Ultrasound Localization Microscopy Based on A Novel Radial Basis Function 2D Interpolation," *IEEE Transactions on Medical Imaging*, vol. 43, n°. 5, p. 1690–1701, mai 2024.

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

[186] W. Gu, B. Li, J. Luo, Z. Yan, D. Ta et X. Liu, "Ultrafast Ultrasound Localization Microscopy by Conditional Generative Adversarial Network," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 70, n°. 1, p. 25–40, janv. 2023.

[187] C. Trabelsi *et al.*, "Deep Complex Networks," dans *International Conference on Learning Representations*, févr. 2018.

[188] J. Lu, F. Millioz, D. Garcia, S. Salles, D. Ye et 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, n°. 2, p. 592–603, févr. 2022.

[189] S. Hochreiter et J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, n°. 8, p. 1735–1780, nov. 1997.

[190] H. Lee, S.-H. Oh, M.-G. Kim, Y.-M. Kim, G. Jung et H.-M. Bae, "Optical Flow Assisted Super-Resolution Ultrasound Localization Microscopy using Deep Learning," dans *2022 IEEE International Ultrasonics Symposium (IUS)*, oct. 2022, p. 1–4.

[191] A. Speiser *et al.*, "Deep learning enables fast and dense single-molecule localization with high accuracy," *Nature Methods*, vol. 18, n°. 9, p. 1082–1090, sept. 2021.

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

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

[194] M. Arjovsky, S. Chintala et L. Bottou, "Wasserstein Generative Adversarial Networks," dans *Proceedings of the 34th International Conference on Machine Learning*. PMLR, juill. 2017, p. 214–223.

[195] 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, avr. 2023.

[196] X. Liu et M. Almekkawy, “Ultrasound Microbubbles Localization Using Object Detection Model,” dans *2023 IEEE International Ultrasonics Symposium (IUS)*, sept. 2023, p. 1–4.

[197] S. K. Gharamaleki, B. Helfield et H. Rivaz, “Transformer-Based Microbubble Localization,” dans *2022 IEEE International Ultrasonics Symposium (IUS)*, oct. 2022, p. 1–4.

[198] A. Aitken, C. Ledig, L. Theis, J. Caballero, Z. Wang et W. Shi, “Checkerboard artifact free sub-pixel convolution : A note on sub-pixel convolution, resize convolution and convolution resize,” juill. 2017.

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

[200] X. Liu et M. Almekkawy, “Ultrasound Super Resolution Using Deep Learning Based on Attention Mechanism,” dans *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, p. 1–5.

[201] S. K. Gharamaleki, B. Helfield et H. Rivaz, “Deformable-Detection Transformer for Microbubble Localization in Ultrasound Localization Microscopy,” dans *2023 IEEE International Ultrasonics Symposium (IUS)*, sept. 2023, p. 1–4.

[202] G. Zhang, Y. Yue, F. Dai, X. Liu et D. Ta, “Transformer for Ultrafast Ultrasound Localization Microscopy,” dans *2022 IEEE International Ultrasonics Symposium (IUS)*, oct. 2022, p. 1–4.

[203] T. Zhou, M. Lu, Y. Yang, Q. He, J. Luo et X. Liu, “qULM-DL : Quantitative Ultrasound Localization Microscopy via Deep Learning,” dans *2020 IEEE International Ultrasonics Symposium (IUS)*. Las Vegas, NV, USA : IEEE, sept. 2020, p. 1–4.

[204] Z. Liu *et al.*, “Swin Transformer : Hierarchical Vision Transformer using Shifted Windows,” dans *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, QC, Canada : IEEE, oct. 2021, p. 9992–10 002.

[205] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov et S. Zagoruyko, “End-to-End Object Detection with Transformers,” dans *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox et J.-M. Frahm, édit. Cham : Springer International Publishing, 2020, vol. 12346, p. 213–229.

[206] X. Zhu, W. Su, L. Lu, B. Li, X. Wang et J. Dai, “Deformable DETR : Deformable Transformers for End-to-End Object Detection,” dans *International Conference on Learning Representations*, oct. 2020.

[207] 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 Biological Engineering*, vol. 42, n°. 6, p. 767–779, déc. 2022.

[208] M. Piepenbrock, D. Koretskaia, G. Schmitz et S. Dencks, “3D Microbubble Localization with a Convolutional Neural Network for Super-Resolution Ultrasound Imaging,” dans *2021 IEEE International Ultrasonics Symposium (IUS)*. Xi’an, China : IEEE, sept. 2021, p. 1–4.

[209] C. Choy, J. Gwak et S. Savarese, “4D Spatio-Temporal ConvNets : Minkowski Convolutional Neural Networks,” dans *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA : IEEE, juin 2019, p. 3070–3079.

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

[211] 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, n°. 6, p. 937–946, juin 2023.

[212] 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, n°. 4, p. 445–455, 2011.

[213] R. G. Gosling et D. H. King, “Arterial Assessment by Doppler-shift Ultrasound,” *Proceedings of the Royal Society of Medicine*, vol. 67, n°. 6 Pt 1, p. 447–449, juin 1974.

[214] C.-P. Chung, H.-Y. Lee, P.-C. Lin et 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, n°. 2, p. 625–632, janv. 2017.

[215] A. Reinke *et al.*, “Understanding metric-related pitfalls in image analysis validation,” *Nature Methods*, vol. 21, n°. 2, p. 182–194, févr. 2024.

[216] L. Maier-Hein *et al.*, “Why rankings of biomedical image analysis competitions should be interpreted with care,” *Nature Communications*, vol. 9, n°. 1, p. 5217, déc. 2018.

[217] 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*, p. 1355–68, 2023.

[218] M. Ashikuzzaman, A. Héroux, A. Tang, G. Cloutier et 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, p. 842–871, juill. 2024.

[219] B. Peng, Y. Xian et J. Jiang, "A Convolution Neural Network-Based Speckle Tracking Method for Ultrasound Elastography," dans *2018 IEEE International Ultrasonics Symposium (IUS)*, oct. 2018, p. 206–212.

[220] B. Peng, Y. Xian, Q. Zhang et 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, p. 74–91, mars 2020.

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

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

[223] A. K. Z. Tehrani, M. Sharifzadeh, E. Boctor et 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, p. 1181–1190, avr. 2022.

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

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

[226] 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, p. 2236–2247, juill. 2022.

[227] 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, p. 1041–1052, mars 2022.

[228] P. Song, A. Manduca, J. D. Trzasko, R. E. Daigle et S. Chen, "On the Effects of Spatial Sampling Quantization in Super-Resolution Ultrasound Microvessel Imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, nº. 12, p. 2264–2276, déc. 2018.

[229] A. Bar-Zion, O. Solomon, C. Tremblay-Darveau, D. Adam et Y. C. Eldar, "SUSHI : Sparsity-Based Ultrasound Super-Resolution Hemodynamic Imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, n°. 12, p. 2365–2380, déc. 2018.

[230] A. Paszke *et al.*, "PyTorch : An Imperative Style, High-Performance Deep Learning Library," dans *Advances in Neural Information Processing Systems*, vol. 32. Curran Associates, Inc., 2019.

[231] J. Gwak, C. Choy et S. Savarese, "Generative Sparse Detection Networks for 3D Single-Shot Object Detection," dans *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox et J.-M. Frahm, édit. Cham : Springer International Publishing, 2020, vol. 12349, p. 297–313.

[232] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang et Z. Tu, "Deeply-Supervised Nets," dans *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*. PMLR, févr. 2015, p. 562–570.

[233] D. P. Kingma et J. Ba, "Adam : A Method for Stochastic Optimization," janv. 2017.

[234] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, n°. 1-2, p. 83–97, 1955.

[235] E. S. Marquez, J. S. Hare et M. Niranjan, "Deep Cascade Learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, n°. 11, p. 5475–5485, nov. 2018.

[236] N. Blanken *et al.*, "PROTEUS : A Physically Realistic Contrast-Enhanced Ultrasound Simulator-Part I : Numerical Methods," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 72, n°. 7, p. 848–865, juill. 2025.

[237] B. Rauby *et al.*, "Ulmshare : A large-scale in vivo ultrasound localization microscopy dataset for microvascular imaging," 2025, submitted.

[238] M. Phuong et C. Lampert, "Towards Understanding Knowledge Distillation," dans *Proceedings of the 36th International Conference on Machine Learning*. PMLR, mai 2019, p. 5142–5151.

[239] Q. Xie, M.-T. Luong, E. Hovy et Q. V. Le, "Self-Training With Noisy Student Improves ImageNet Classification," dans *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, p. 10 687–10 698.

[240] R. Das et S. Sanghavi, "Understanding Self-Distillation in the Presence of Label Noise," dans *Proceedings of the 40th International Conference on Machine Learning*. PMLR, juill. 2023, p. 7102–7140.

[241] O. Ronneberger, P. Fischer et T. Brox, "U-Net : Convolutional Networks for Biomedical Image Segmentation," dans *Medical Image Computing and Computer-Assisted Inter-*

*vention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells et A. F. Frangi, édit. Cham : Springer International Publishing, 2015, p. 234–241.

[242] B. Heiles *et al.*, "Nonlinear sound-sheet microscopy : Imaging opaque organs at the capillary and cellular scale," *Science*, vol. 388, nº. 6742, p. eads1325, avr. 2025.

[243] F. Bureau, L. Denis, A. Coudert, M. Fink, O. Couture et A. Aubry, "Ultrasound matrix imaging for 3D transcranial in vivo localization microscopy," *Science Advances*, vol. 11, nº. 31, p. eadt9778, juill. 2025.

[244] Y. Wang, Y. Wang, C. Zheng, H. Peng et C. Zhang, "Ultrasound Contrast Microbubbles Reconstruction Using a Joint Enhanced Mean-to-Standard-Deviation Factor and Minimum Variance Beamformer," dans *2022 IEEE International Ultrasonics Symposium (IUS)*, oct. 2022, p. 1–4.

[245] A. Corazza, P. Muleki-Seya, A. W. Aissani, O. Couture, A. Basarab et B. Nicolas, "Microbubble detection with adaptive beamforming for Ultrasound Localization Microscopy," dans *2022 IEEE International Ultrasonics Symposium (IUS)*, oct. 2022, p. 1–4.

[246] D. H. Pham, V. Pustovalov et D. Kouamé, "The Performance Improvement of Ultrasound Localization Microscopy (ULM) Using the Robust Principal Component Analysis (RPCA)," dans *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, juill. 2023, p. 1–4.

[247] C. Hahne et R. Sznitman, "Geometric Ultrasound Localization Microscopy," juill. 2023.

[248] A. Corazza, P. Muleki-Seya, A. Basarab et B. Nicolas, "Microbubble Identification Based on Decision Theory for Ultrasound Localization Microscopy," *IEEE Open Journal of Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 3, p. 41–55, 2023.

[249] A. Corazza, P. Muleki-Seya, A. Chavignon, O. Couture, A. Basarab et B. Nicolas, "Adaptive beamforming combined with decision theory-based detection for ultrasound localization microscopy," dans *2023 IEEE International Ultrasonics Symposium (IUS)*, sept. 2023, p. 1–4.

[250] C. Hahne, G. Chabouh, O. Couture et R. Sznitman, "Learning Super-Resolution Ultrasound Localization Microscopy from Radio-Frequency Data," nov. 2023.

[251] C. Hahne, M. Hayoz et R. Sznitman, "StofNet : Super-resolution Time of Flight Network," déc. 2023.

[252] Y. Xiao, W. Han et B. Peng, "Deep-Learning-Based Video Frame Interpolation Method for Ultrasound Localization Microscopy : A Preliminary Study," dans *2023 IEEE 6th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*, août 2023, p. 669–674.

[253] G. Tuccio, S. Afrakhteh, G. Iacca et L. Demi, “Relaxing Super Localization Frame Rate Requirements Utilizing a Novel 2D Interpolation Technique,” dans *2023 IEEE International Ultrasonics Symposium (IUS)*, sept. 2023, p. 1–3.

[254] B. Pialot, L. Augeul, L. Petrusca et F. Varray, “A simplified and accelerated implementation of SVD for filtering ultrafast power Doppler images,” *Ultrasonics*, vol. 134, p. 107099, sept. 2023.

[255] S. Afrakhteh, G. Tuccio et L. Demi, “A Novel 2x2D Radial Basis Function-Based Interpolation for Short Acquisition Time and Relaxed Frame Rate Ultrasound Localization Microscopy,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 71, nº. 12 : Breaking the Resolution Barrier in Ultrasound, p. 1855–1867, déc. 2024.

[256] H. Lan *et al.*, “Deep Power-Aware Tunable Weighting for Ultrasound Microvascular Imaging,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 71, nº. 12 : Breaking the Resolution Barrier in Ultrasound, p. 1701–1713, déc. 2024.

[257] Y. Chen, B. Fang, F. Meng, J. Luo et X. Luo, “Competitive Swarm Optimized SVD Clutter Filtering for Ultrafast Power Doppler Imaging,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 71, nº. 4, p. 459–473, avr. 2024.

[258] G. Tuccio, S. Afrakhteh et L. Demi, “Towards sub-100Hz Super-Resolution Imaging Through a Novel Bi-Directional Interpolation Technique,” dans *2024 IEEE Ultrasonics, Ferroelectrics, and Frequency Control Joint Symposium (UFFC-JS)*, sept. 2024, p. 1–4.

[259] Y. Qiang *et al.*, “An adaptive spatiotemporal filter for ultrasound localization microscopy based on density canopy clustering,” *Ultrasonics*, vol. 144, p. 107446, déc. 2024.

[260] B. Fang, H. Li et Y. Chen, “Total Variational Robust PCA for Ultrasound Microvascular Clutter Filtering,” dans *2024 IEEE Ultrasonics, Ferroelectrics, and Frequency Control Joint Symposium (UFFC-JS)*, sept. 2024, p. 1–4.

[261] H. Li, B. Fang, Z. Ye, F. Meng et Y. Chen, “Attention USR-Net : An End-to-End Mapped Ultrasound Localization Microscopy,” dans *2024 IEEE Ultrasonics, Ferroelectrics, and Frequency Control Joint Symposium (UFFC-JS)*, sept. 2024, p. 1–4.

[262] W. Han, W. Zhou, L. Huang, J. Luo et B. Peng, “Tissue Clutter Filtering Methods in Ultrasound Localization Microscopy Based on Complex-Valued Networks and Knowledge Distillation,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 72, nº. 4, p. 440–453, avr. 2025.

[263] V. Pustovalov, D. H. Pham, C. Alix, J.-P. Remeniéras et D. Kouamé, “Computational Super-Resolution for Ultrasound Localization Microscopy Through Solving an Inverse Problem,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 72, nº. 5, p. 636–645, mai 2025.

[264] D. H. Pham, “Leveraging Autoencoder Neural Networks to Improve Microbubble Detection and Localization in Ultrasound Localization Microscopy,” dans *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*, avr. 2025, p. 1–4.

[265] B. Pialot *et al.*, “Computationally Efficient SVD Filtering for Ultrasound Flow Imaging and Real-Time Application to Ultrafast Doppler,” *IEEE Transactions on Biomedical Engineering*, vol. 72, nº. 3, p. 921–929, mars 2025.

[266] H. Cho, J. Lee, S. Park et Y. Yoo, “Numerical investigation of optimal transmission-reception conditions for aliasing-free ultrasound localization microscopy,” *Ultrasonics*, vol. 154, p. 107704, oct. 2025.

[267] P. Xing, A. Malescot, E. Martineau, R. L. Rungta et J. Provost, “Inverse Problem Approach to Aberration Correction for In Vivo Transcranial Imaging Based on a Sparse Representation of Contrast-Enhanced Ultrasound Data,” *IEEE Transactions on Biomedical Engineering*, vol. 72, nº. 11, p. 3196–3209, nov. 2025.

[268] P. Xing *et al.*, “3D ultrasound localization microscopy of the nonhuman primate brain,” *eBioMedicine*, vol. 111, janv. 2025.

[269] M. G. Wagner, “Real-time thinning algorithms for 2D and 3D images using GPU processors,” *Journal of Real-Time Image Processing*, vol. 17, nº. 5, p. 1255–1266, oct. 2020.

[270] A. Cigier, F. Varray et D. Garcia, “SIMUS : An open-source simulator for medical ultrasound imaging. Part II : Comparison with four simulators,” *Computer Methods and Programs in Biomedicine*, vol. 220, p. 106774, juin 2022.

[271] D. Garcia, “Make the most of MUST, an open-source Matlab UltraSound Toolbox,” dans *2021 IEEE International Ultrasonics Symposium (IUS)*, sept. 2021, p. 1–4.

[272] R. P. J. Nieuwenhuizen *et al.*, “Measuring image resolution in optical nanoscopy,” *Nature Methods*, vol. 10, nº. 6, p. 557–562, juin 2013.

# ANNEXE A LISTE DES CONTRIBUTIONS SCIENTIFIQUES

Cette thèse a donné lieu à plusieurs publications dans des revues à comité de lecture, ainsi qu'à des présentations lors de conférences internationales et locales.

**Articles de journaux (Premier auteur)**

- **B. Rauby**, P. Xing, J. Porée, M. Gasse, J. Provost. "Pruning Sparse Tensor Neural Networks Enables Deep Learning for 3D Ultrasound Localization Microscopy." *IEEE Transactions on Image Processing*, vol. 34, pp. 2367–2378, 2025.
- **B. Rauby**, P. Xing, M. Gasse, J. Provost. "Deep Learning in Ultrasound Localization Microscopy : Applications and Perspectives." *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 71, no. 12, pp. 1765–1784, 2024.
- **B. Rauby**, A. Leconte, A. Wu, G. Ramos-Palacios, S. A. Lee, J. Porée, A. F. Sadikot, M. Gasse, J. Provost. "Teacher-Student models for robust in vivo deep-learning in Ultrasound Localization Microscopy." *Soumis*.
- **B. Rauby***, N. Ghigo*, G. Ramos-Palacios, A. Leconte, S. A. Lee, A. Wu, P. Xing, O. Gulenko, L. Caron, A. Malescot, E. Martineau, J. Porée, M. Gasse, R. Rungta, A. Sadikot, J. Provost. "ULMShare : A Large-Scale In Vivo Ultrasound Localization Microscopy Dataset for Microvascular Imaging." *Soumis*.

**Articles de journaux (Co-auteur)**

- A. Leconte, J. Porée, **B. Rauby**, A. Wu, N. Ghigo, P. Xing, S. A. Lee, C. Bourquin, G. Ramos-Palacios, A. F. Sadikot, J. Provost. "A Tracking Prior to Localization Workflow for Ultrasound Localization Microscopy." *IEEE Transactions on Medical Imaging*, vol. 44, no. 2, pp. 698–710, 2025.
- P. Xing, J. Porée, **B. Rauby**, A. Malescot, E. Martineau, V. Perrot, R. L. Rungta, J. Provost. "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, 2024.
- C. Bourquin, J. Porée, **B. Rauby**, V. Perrot, N. Ghigo, H. Belgharbi, S. Bélanger, G. Ramos-Palacios, N. Cortes, H. Ladret, et al. "Quantitative pulsatility measurements using 3D dynamic ultrasound localization microscopy." *Physics in Medicine & Biology*, vol. 69, no. 4, 045017, 2024.

**Conférences Internationales (premier auteur)**

— **B. Rauby**, A. Leconte, J. Porée, M. Gasse, J. Provost. "Pseudo-Labels and Input Perturbation Improve in Vivo Applications of Deep Learning Ultrasound Localization Microscopy." *IEEE International Ultrasonics Symposium (IUS)*, 2025, Affiche.
— **B. Rauby**, J. Porée, H. Belgharbi, C. Bourquin, M. Gasse, J. Provost. "SparseNeST-ULM : Sparse Tensor Neural Network for ND-Ultrasound Localization Microscopy." *IEEE International Ultrasonics Symposium (IUS)*, 2022, Affiche.
— **B. Rauby**, J. Porée, H. Belgharbi, C. Bourquin, M. Gasse, J. Provost. "3D Spatiotemporal Ultrasound Localization Microscopy Using Deep Learning." *IEEE International Ultrasonics Symposium (IUS)*, 2021, Présentation Orale.

**Conférences Institutionnelles et Locales (premier auteur)**

— **B. Rauby**, J. Porée, A. Leconte, M. Gasse, J. Provost. "Improving Ultrasound Localization Microscopy with Complex Valued Neural Networks." *Colloque d'imagerie médicale de Québec (CIMQ)*, 2025, Affiche.
— **B. Rauby**, J. Porée, H. Belgharbi, C. Bourquin, M. Gasse, J. Provost. "Sparse Neural Networks for Ultrasound Localization Microscopy." *IVADO Octobre Numérique*, Oct. 2021, Présentation Orale.
— **B. Rauby**, J. Porée, H. Belgharbi, C. Bourquin, J. Provost. "Sous échantillonnage temporel en apprentissage profond pour la microscopie de localisation ultrasonore." *Imagining Imaging Symposium - Sherbrooke University*, Déc. 2020, Présentation courte.
— **B. Rauby**, J. Porée, H. Belgharbi, C. Bourquin, J. Provost. "Sous échantillonnage temporel en apprentissage profond pour la microscopie de localisation ultrasonore." *Digital October IVADO*, Oct. 2020, Présentation Orale.