

# Towards in-vivo ultrasound-histology: Plane-waves and generative adversarial networks for pixel-wise speed of sound reconstruction

Ivan Pavlov

*Computer Aided Medical Procedures,*  
*Technische Universität München*  
Boltzmannstraße 3,  
85748 Garching bei München, Germany  
ivan.pavlov@tum.de

Eduardo Prado

*Computer Aided Medical Procedures,*  
*Technische Universität München*  
Boltzmannstraße 3,  
85748 Garching bei München, Germany  
eduardo.prado@tum.de

Nassir Navab

*Computer Aided Medical Procedures,*  
*Technische Universität München*  
Boltzmannstraße 3,  
85748 Garching bei München, Germany  
*Computer Aided Medical Procedures,*  
*Johns Hopkins University,*  
3400 North Charles Street,  
Baltimore, MD 21218, USA  
nassir.navab@tum.de

Guillaume Zahnd

*Computer Aided Medical Procedures,*  
*Technische Universität München*  
Boltzmannstraße 3,  
85748 Garching bei München, Germany  
g.zahnd@tum.de

**Abstract**—Ultrasound imaging is a well-established modality, widely used for in vivo real time examination. Nevertheless, the ability of conventional ultrasound techniques is limited by the fact that different biological tissues are sometimes represented with the same image brightness, thus hindering visual — as well as automatic — identification. Especially valuable for tissue differentiation is the pressure wave velocity, which can be measured with ultrasound. Deep-learning-based methods carry a possibility to overcome such limitations and enable robust signal-based tissue identification. Such methods have been successfully applied to tackle various challenges of medical imaging research. The aim of the present work is to propose a Generative Adversarial Network (GAN) pipeline towards pixel-wise speed of sound (SoS) reconstruction from plane-wave ultrasound raw channel signals corresponding to three firing angles. The network is trained on a novel synthetic dataset focusing on complex geometry, generated with K-Wave. Results demonstrate a promising performance, with average ( $\pm$  STD) absolute SoS reconstruction errors of  $38 \pm 54$  m/s in real time at 114 fps. The proposed approach paves the way towards GAN-based ultrasound histology.

**Index Terms**—Deep learning, Ultrasound, Speed of Sound reconstruction

## I. INTRODUCTION

Pressure wave velocity (i.e. speed of sound, SoS) is an intrinsic property of biological tissues. Various anatomical components (e.g. muscle, tumor, atherosclerotic plaque) are characterised by their own specific mechanical density, which locally defines the SoS [1]. This property is often described as a potential biomarker for different forms of cancer [2]

This work was supported by the NVIDIA Corporation GPU grant program

and can be estimated using ultrasound imaging. Medical ultrasound B-mode imaging is a structural modality providing a detailed representation of the imaged region. However, such information can sometimes be ambiguous — and therefore lead to inaccurate diagnosis — when tissues of different nature result in the same pixel brightness.

There have been extensive developments in the application of deep learning to ultrasound imaging, mainly focused on segmentation, labelling and classification [3]. However, recent works have explored the problem of ultrasound inversion [4], namely the reconstruction of information regarding tissues acoustic properties via raw signals. Traditionally, this task was addressed with full-wave inverse scattering, which requires large computational resources and often limited to specific acquisition scenarios [5]. Current researches support the idea to replace conventional explicit approaches with novel implicit deep-learning-based techniques. A single-sided SoS inversion solution using a fully convolutional deep neural network was developed [6]. A deep-learning-based reconstruction framework was proposed for the scenario of limited-angle ultrasound tomography in prostate cancer [7]. In both cases, promising results are achieved, yet the samples geometry was rather simple. Introduction of more complex and realistic datasets can help the development of more efficient methods of reconstruction and address the problem of domain gap between synthetic and real data. Addressing this challenge, the aim of this study is to propose a GAN-based approach towards functional imaging via pixel-wise SoS reconstruction.

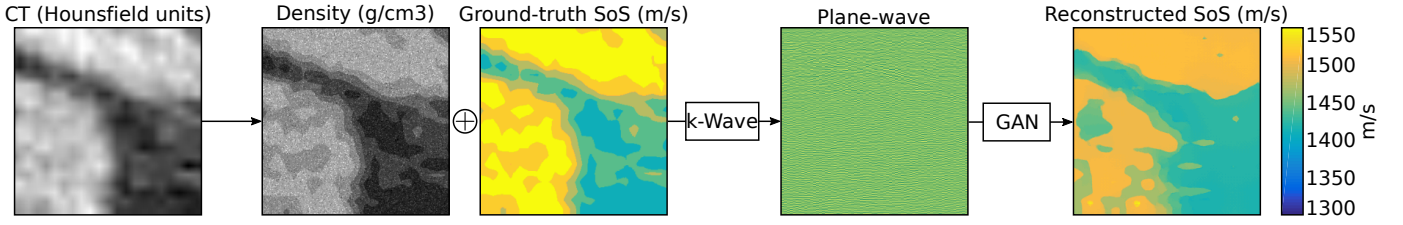


Fig. 1. Proposed reconstruction pipeline. The generative adversarial network (GAN) is used to reconstruct pixel-wise SoS values of tissues from plane-wave raw channel data. Training pairs are obtained using simulations from preprocessed CT data.

## II. METHODS

### A. Dataset preparation

The synthetic data approach was chosen in the described pipeline since collecting the real world data is practically impossible. The dataset consists of 9500 pairs of plane-wave raw channel data (3 firings for each pair) and corresponding ground truth medium data (SoS and density maps). When preparing the simulated dataset the goal was to obtain the realistically looking SoS maps. It is a complicated task to produce such maps via direct measurements or modelling. Therefore the assumption is that medium maps which yield realistically looking simulated B-mode images are realistic enough themselves. K-Wave MATLAB toolbox was chosen [8] as the ultrasound simulation software because it can account for acoustic heterogeneities while preserving a moderate run time on GPU. Experiments with speckle noise modelling and acoustic interface geometries were carried out, and based on that the data preparation protocol that yielded the best looking B-mode images was used. The dataset was prepared in accordance with this protocol as described below.

First cross-sectional regions from publicly available abdominal CT scans from the Visible Human Project [9] were extracted utilising both male and female volumes. CT scans contain the modality-specific noise [10] which can lead to incorrect results when they are used directly as the input to the ultrasound simulation software. In order to overcome this, scans were segmented using k-means [11] algorithm with 5 clusters. Next SoS and density values we obtained from segmented CT regions using the scale from the experimental data [12]. Due to the run time considerations, Hounsfield units of CT scans were truncated to the values corresponding to the SoS in the range of 1300 m/s and 1700 m/s. After that a random speckle is added in the density domain. The noise generation approach is based on [6] with the following modifications that helped achieve better looking B-mode images. 1) Heterogeneous noise with mass density variations between  $\pm 2\%$  and  $\pm 10\%$  depending on the SoS of the segmented region. 2) Noise distribution density of 21 reflectors per cubic wavelength. Resulting SoS and noisy density maps were used as the input for the simulations. The dataset preparation pipeline is showed in Figure 1.

Simulations were performed on cropped sections of size  $21.2 \times 45 \times 1.1$  mm (depth  $\times$  width  $\times$  elevation thickness). For the transmitter and the sensor the Cephasonics CPLA12875

(Cephasonics, CA, USA) linear transducer was modelled. It features 64 active elements, 128 elements in total and the central frequency of 5MHz. For each cropped section of medium maps three overlapping firings were simulated: one central and two  $15^\circ$  angled outside looking plane-waves. The simulation setup is showed in the Figure 2.

Raw channel data from the simulations was cropped to remove the transmit pulse and later the time gain compensation was applied. Finally both SoS maps and raw channel data were reshaped to the size of  $256 \times 256$  pixels for one firing or  $256 \times 512$  pixels for three overlapping firings with the normalised brightness between 0 and 1.

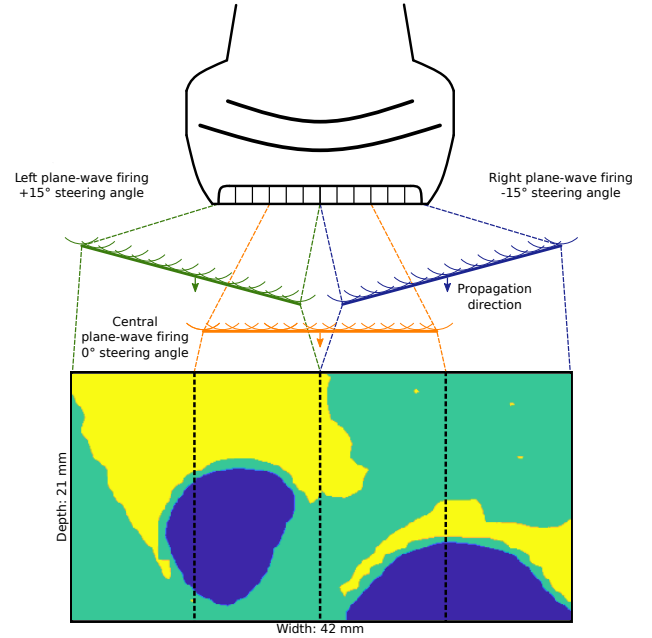


Fig. 2. Simulation setup. Three firings with steering angles of  $+15^\circ$ ,  $0^\circ$  and  $-15^\circ$  are used to simulate raw channel data for one training pair. Each firing covers a half-width part of the full reconstruction domain and uses a half of the transducer's elements.

### B. Network architecture

The architecture is based on the pix2pix conditional GAN [13]. Recent research shows outstanding performance of this type of architecture in the field of image to image mapping [14]. The most important aspect that defines the

decision to explore this architecture is the ability to learn complex generation quality criterion in the form of discriminator network. This helps to avoid excessive blurring present in other architectures such as direct use of auto-encoders.

The U-net generator network was modified by introducing individual downsampling branches for each of three input raw channel data. In the upsampling branch, the feature maps from different downsampling branches are combined at every size step in the following way. Two side feature maps are concatenated and then overlapped with the central feature map. After that, the resulting combined feature map is normalised. Such three-to-one combination gives the network a possibility to learn individual downsampling path for a differently angled input data while preserving their known spatial relation for the united upsampling path. The downsampling and upsampling are done through series of  $4 \times 4$  convolutions/deconvolutions with a stride of 2 and a padding of 1. This operation changes the resolution by a factor of two at each step. The modified generator architecture is shown in the Figure 3. In addition to that, the one-to-one architecture was tested. It features only one downsampling branch that corresponds to data from the central firing. Training and evaluation was done in PyTorch [15].

### III. RESULTS

The network was trained for a total of 300 epochs on our synthetic dataset. The same learning scheme as in the pix2pix paper [13] was used with a linear decay of the learning rate, the batch size of 10 and the least squares loss function. The reconstruction method was evaluated on a test dataset of 200 new data pairs, with an inference performing at 114

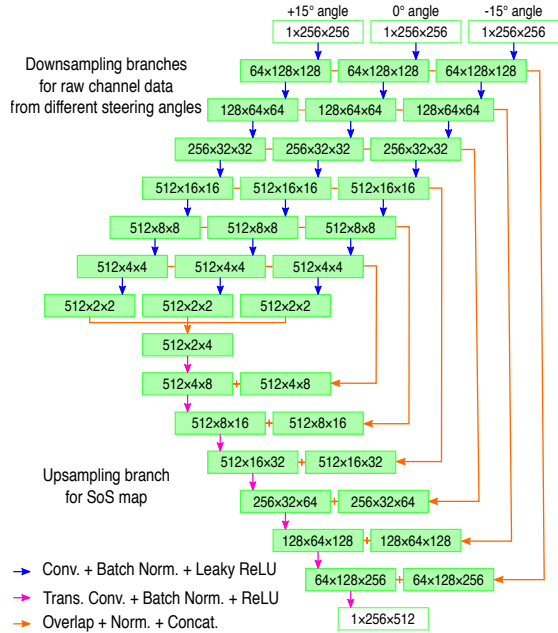


Fig. 3. Three-to-one variant of the generator architecture. Input layers take preprocessed and resized raw channel data. The output layer corresponds to the full-size reconstructed SoS map.

fps. Evaluation results are showed in the Figure 4. The mean absolute ( $\pm$ STD) error when quantifying the pixel-wise SoS was  $38 \pm 54$  m/s (bias of 7 m/s, Bland-Altman 95% limits of agreement of  $[-123, 136]$  m/s). Compared to the actual SoS magnitude, the mean absolute relative error was  $3 \pm 4\%$ .

### IV. DISCUSSION

The trained generator is able to reconstruct large structures with a good level of accuracy. However, smaller objects can be missed, especially when they are only slightly different from the surroundings in terms of the SoS and density. Additionally, the method can address the noise reduction task and produce reconstructed structures with clearly differentiable boundaries. One-to-one and three-to-one variations of the generator architectures were tested. Both of them achieve similar level of performance, however three-to-one variant offers two times wider reconstruction area. The central area of the reconstruction shows higher degree of pixel wise accuracy than areas on the side borders. This can be improved by the utilisation of data from additional angled firings. The pipeline is able to run in real time and can stay in this time frame even with additional downsampling branches.

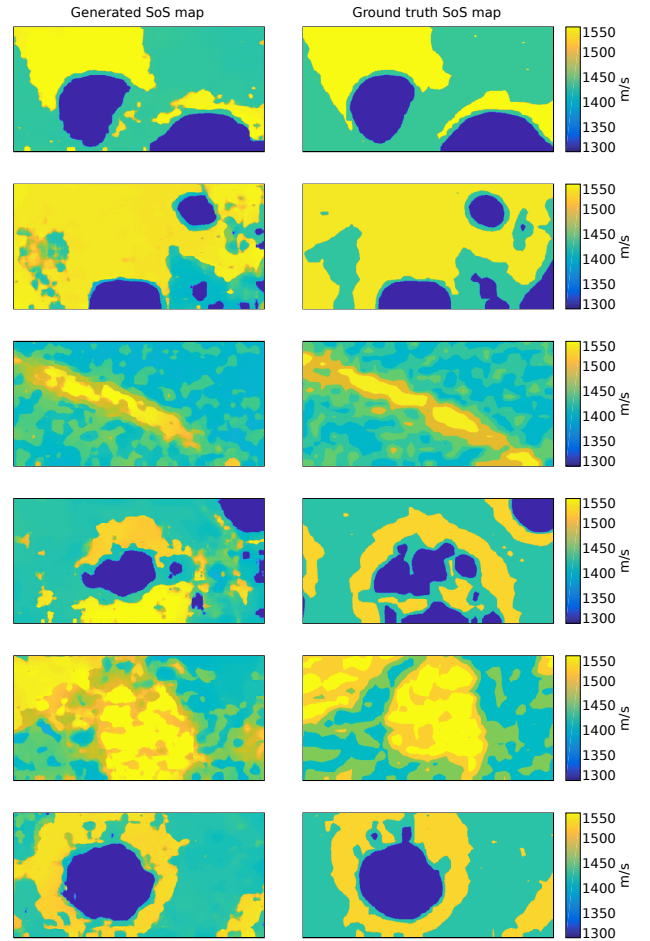


Fig. 4. Evaluation results. Generated SoS maps were produced with the three-to-one variant of the generator architecture.

## V. CONCLUSION

Paving the way towards fine-tuning of the beamforming delays, pixel-wise SoS has potential to decrease image distortion by refining the conventionally accepted value of 1540 m/s. Overall the high potential of generative adversarial networks in the task of reconstruction of complex heterogeneous structures is clearly seen. But such deep-learning-based methods are crucially dependant on the quality and the amount of training data. Thus, the creation of a standardised SoS reconstruction dataset will establish a good ground for the direct comparison of existing approaches and boost the overall development of new promising ideas.

## REFERENCES

- [1] S. F. Reis, *Characterisation of biological tissue: measurement of acoustic properties for Ultrasound Therapy*. PhD thesis, University of Lisbon, 2013.
- [2] B. Malik, J. Klock, J. Wiskin, and M. Lenox, "Objective breast tissue image classification using Quantitative Transmission ultrasound tomography," *Scientific Reports*, vol. 6, 2016.
- [3] G. Litjens, T. Kooi, B. Ehteshami Bejnordi, A. Setio, F. Ciompi, M. Ghafoorian, J. van der Laak, B. Ginneken, and C. Snchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, 02 2017.
- [4] H. Almansouri, S. V. Venkatakrishnan, G. T. Buzzard, C. A. Bouman, and H. Santos-Villalobos, "Deep neural networks for non-linear model-based ultrasound reconstruction," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 6–10, 2018.
- [5] K. Wang, T. Matthews, F. Anis, C. Li, N. Duric, and M. Anastasio, "Waveform inversion with source encoding for breast sound speed reconstruction in ultrasound computed tomography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 62, no. 3, pp. 475–493, 2015.
- [6] M. Feigin, D. Freedman, and B. W. Anthony, "A deep learning framework for single-sided sound speed inversion in medical ultrasound," *arXiv preprint 1810.00322*, 2018.
- [7] A. Cheng, Y. Kim, E. Anas, A. Rahmim, E. Boctor, R. Seifabadi, and B. Wood, "Deep learning image reconstruction method for limited-angle ultrasound tomography in prostate cancer," *Proceedings of SPIE*, vol. 10955, 2019.
- [8] B. E. Treeby, J. Jaros, D. Rohrbach, and B. T. Cox, "Modelling elastic wave propagation using the k-Wave MATLAB toolbox," *IEEE International Ultrasonics Symposium*, pp. 146–149, 2014.
- [9] M. Ackerman, "The visible human project," *Proceedings of the IEEE*, vol. 86, no. 3, pp. 504–511, 1998.
- [10] T. Mast, "Empirical relationships between acoustic parameters in human soft tissues," *Acoustics Research Letters Online*, vol. 1, no. 6, pp. 37–42, 2000.
- [11] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.
- [12] U. Schneider, E. Pedroni, and L. A., "The calibration of CT Hounsfield units for radiotherapy treatment planning," *Physics in Medicine & Biology*, vol. 41, pp. 111–124, 1996.
- [13] P. Isola, J. Zhu, T. Zhou, and E. A., "Image-to-image translation with conditional adversarial networks," *The Conference on Computer Vision and Pattern Recognition*, 2017.
- [14] H. Huang, P. S. Yu, and C. Wang, "An introduction to image synthesis with generative adversarial nets," *arXiv e-prints*, 2018.
- [15] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," in *NIPS Autodiff Workshop*, 2017.