



HAL
open science

Ultrasound volume reconstruction from 2D freehand acquisitions using neural implicit representations

François Gaits, Nicolas Mellado, Adrian Basarab

► **To cite this version:**

François Gaits, Nicolas Mellado, Adrian Basarab. Ultrasound volume reconstruction from 2D freehand acquisitions using neural implicit representations. 21st IEEE International Symposium on Biomedical Imaging (ISBI 2024), May 2024, Athènes, Greece. . hal-04592493

HAL Id: hal-04592493

<https://hal.science/hal-04592493>

Submitted on 29 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ULTRASOUND VOLUME RECONSTRUCTION FROM 2D FREEHAND ACQUISITIONS USING NEURAL IMPLICIT REPRESENTATIONS



Université
de Toulouse



François Gaits¹, Nicolas Mellado¹, and Adrian Basarab²

¹Institut de Recherche en Informatique de Toulouse, UMR CNRS 5505, Université de Toulouse, France

²Université de Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS,

Inserm, CREATIS UMR 5220, U1206, Villeurbanne, France

Francois.Gaits@irit.fr



Overview

The objective of this work is to propose an unsupervised deep learning approach for 3D ultrasound reconstruction. We took inspiration from the neural implicit representations (NIR), a family of approaches that learn volumetric functions from 3D samples [1]. Inspired by NIR this work aims to use its idea to create a **3D volume** based on **freehand 2D ultrasound sweep**.

This work is partly inspired and motivated by existing article around the same idea: **ImplicitVol** [2] optimizes the positions of the slice along the volume using a NIR network, and **Ultra-NeRF** [3] centers its study around a sophisticated render process.

Using NIR we are able to produce:

- ◆ A denoised volume with high CNR and SNR
- ◆ A continuous lightweight representation in the form of the network

Neural Implicit Representation

NIR is a concept that revolves around **learning spatial properties** through example, and being able to quantify them **anywhere** in the volume of interest (VOI). It can be formulated as:

$$\Phi_{\theta}(\gamma(\mathbf{x})) = p_x, \quad (1)$$

where the network learns a function Φ with the optimized weights θ that maps \mathbf{x} the **coordinates** in the VOI (and optional additional information) to p_x a vector of **properties** at \mathbf{x} . The function γ is an **embedding**, giving higher dimensionality to \mathbf{x} . **Frequency embedding** [1] or **Hash grid embedding** [4] are usually employed. The impact of this embedding is one of the keys of this work.

Properties learned can be widely different, from occupancy to movement fields or semantic information. Of particular note is the recent work on novel view synthesis [1] that allows to learn the radiance emitted from a scene and create new views from a set of pictures.

The aim of this work is to apply the ideas of NIR to 3D ultrasound volume reconstruction, leveraging this **light**, **efficient** and **continuous** representation.

Material & Method

To show the potential of the proposed method, **simulated** and **experimental** images are used. The simulated images are obtained by generating synthetic ultrasound images from a volume for different positions of the simulated ultrasound probe [5]. The volumes considered are synthetic objects with perfectly **known geometry**, or 3D *in vivo* magnetic resonance volumes.

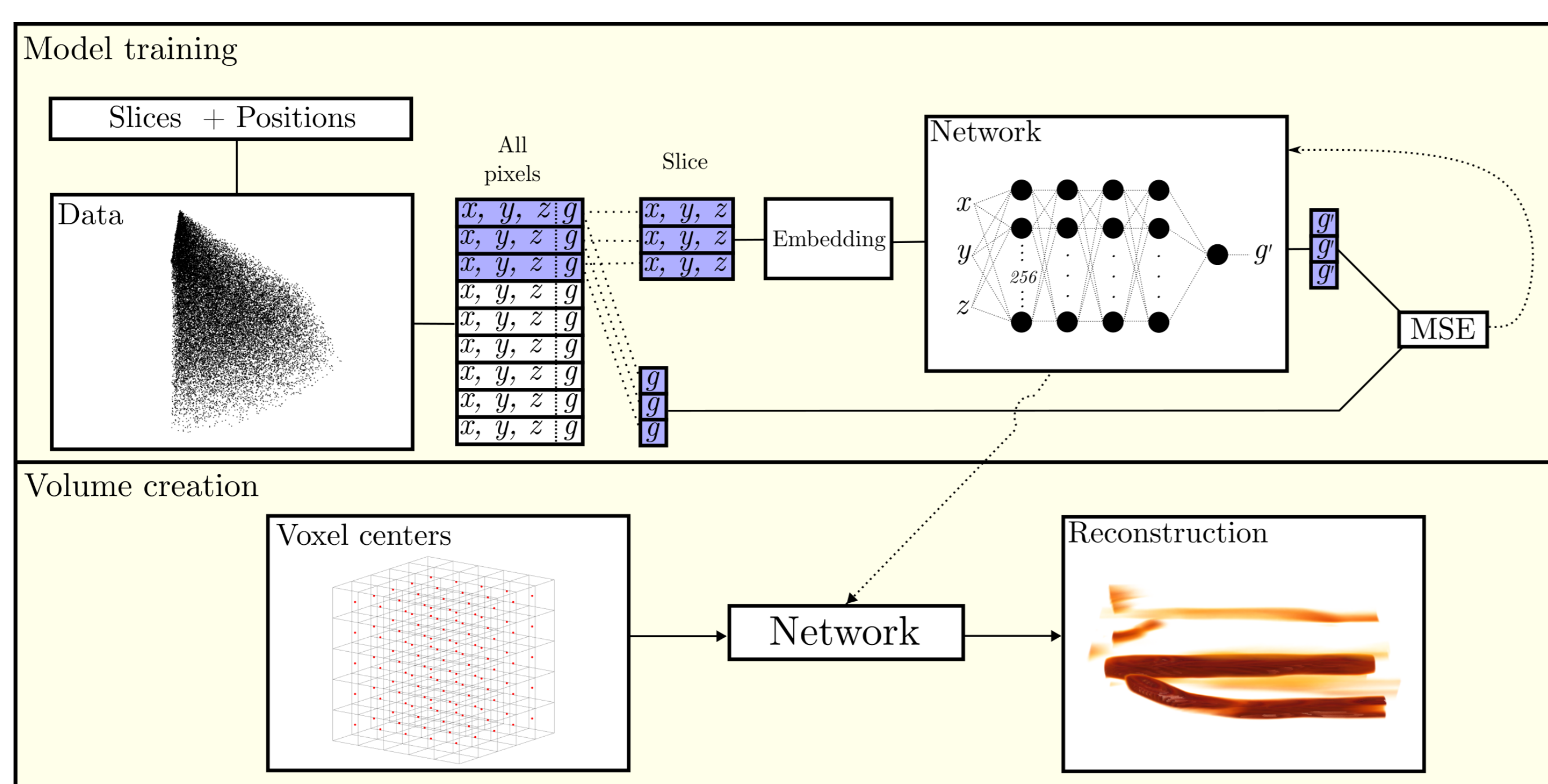


FIGURE 1 - Overview of the model training and usage. Each pixel of each slice represents a 4D point of the coordinates and the gray level of the image. An epoch consists in generating the gray level at the positions of a slice's pixels and comparing the results against the actual value using MSE, and use it to optimize the network's parameters. Once the training is done, the network can then be used to estimate the values in a voxel grid to create a 3D volume.

At each learning step, the network's results are compared against the expected value on a subset of known locations corresponding to the pixels of one of the ultrasound slices. The loss function is defined as the mean squared error between the network output and ultrasound slice pixel value, yielding to the following optimization process:

$$\theta = \arg \min_{\theta} \|\Phi_{\theta}(\gamma(\mathbf{x})) - g_x\|_2^2. \quad (2)$$

Impact of the Embedding

One of the key points in NIR is the **embedding of the input** through the function γ in 1. This encoding phase is the first layer of the network. Its role is to **increase** the dimension of the position vector \mathbf{x} , allowing the network to represent **more finely** spatial variations. In this work, we evaluate three encoding schemes:

◆ Frequency encoding:

Each component of the input vector \mathbf{x} , is mapped from \mathbb{R} to \mathbb{R}^{2L} (with L being a meta-parameter) following a function $\gamma(x)$:

$$\gamma(x) = [\sin(2^0\pi x), \cos(2^0\pi x), \dots, \sin(2^{L-1}\pi x), \cos(2^{L-1}\pi x)] \quad (3)$$

◆ Hash grid encoding:

Müller *et al.* [4] proposed a new approach to expand the input vector to **more meaningful** values through a process combining hierarchical acceleration structures and deep learning. In the general case, this approach tremendously speeds up the learning process and allows **finer details**.

◆ No encoding:

Encoding aims at enhancing the high frequencies in the learned medium. While this property is of high interest in optical applications, it is not necessarily the case in 3D ultrasound reconstruction, because of the **high level of speckle noise that affects the slices**.

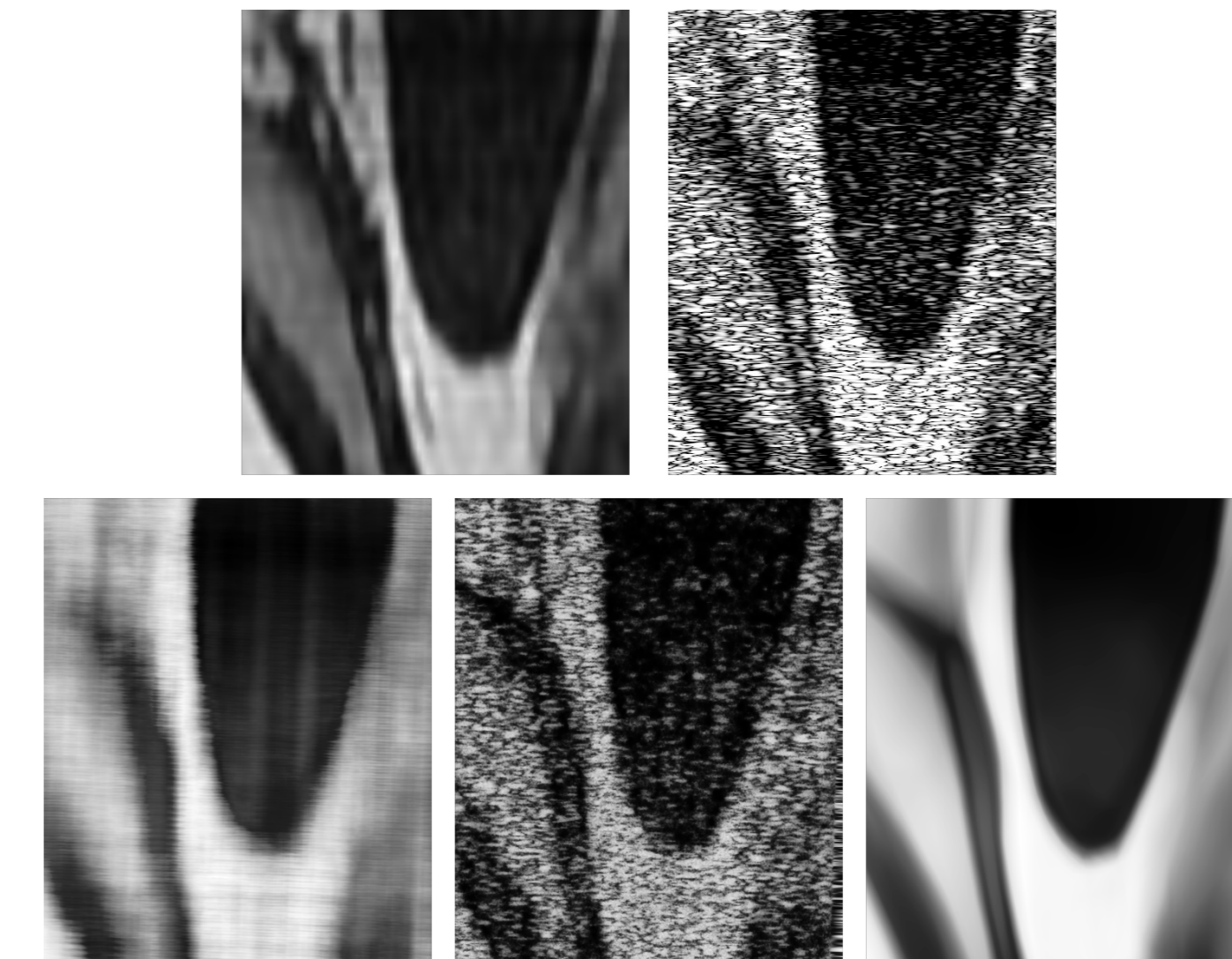


FIGURE 2 - Comparison of the reproduction of an ultrasound slice by a network trained on the same dataset for 10 minutes (reproduced slice not present in the learning set). Top: MRI slice, simulated ultrasound image. Bottom: reconstructed slice with frequency encoding, hash grid encoding, and without encoding.

Results

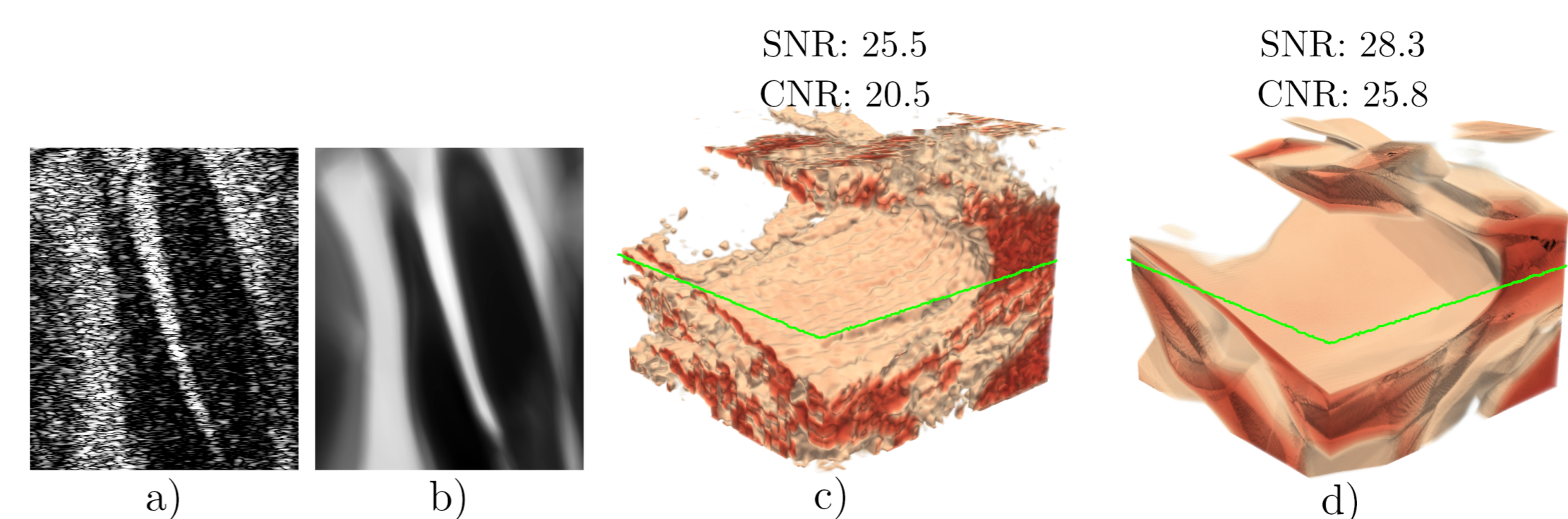


FIGURE 3 - Reconstruction based on a simulated double sweep (one horizontal, one vertical) of a T2 pelvic MRI based medium, with the probe trajectory shown in green. a) Ultrasound image given as an input to the network, b) reconstruction of the image in a) with the proposed method, c) DW reconstruction, d) 3D reconstruction with the proposed method.

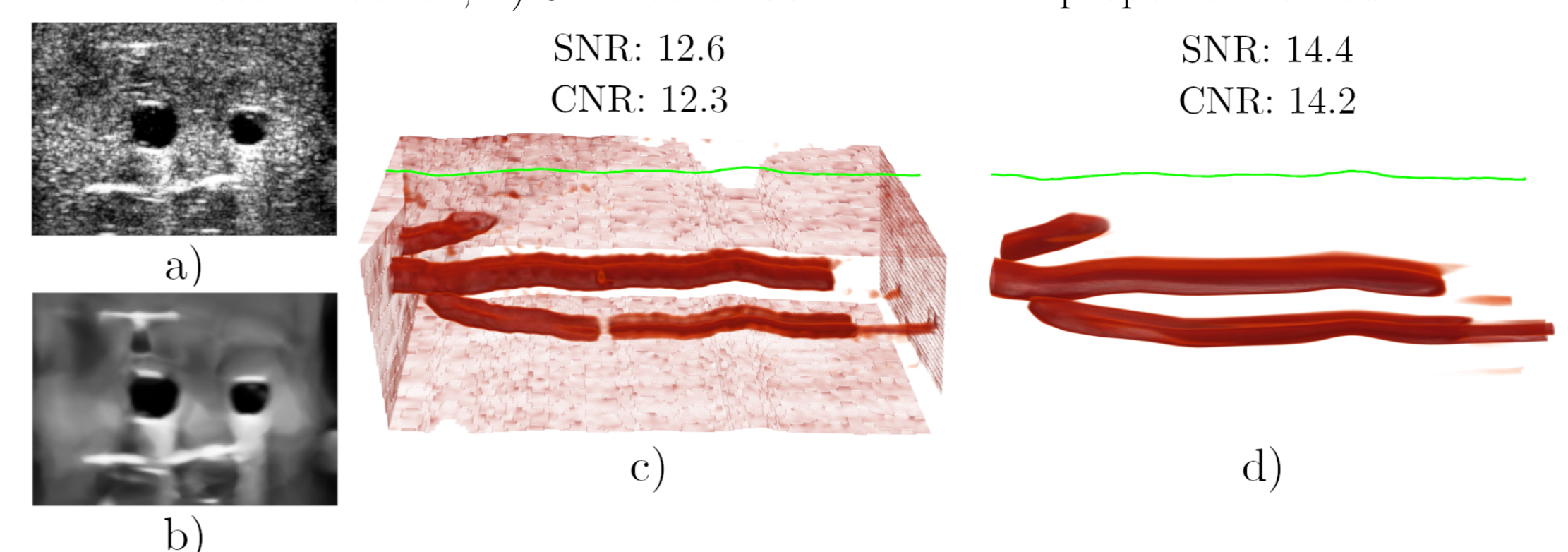


FIGURE 4 - Reconstruction based on a single horizontal freehand sweep acquired on a nerve phantom with probe trajectory (green). a) Ultrasound image given as an input to the network, b) Network reconstruction of the image in a), c) DW reconstruction, d) 3-D reconstruction.

References

- [1] Ben Mildenhall et al. "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis". In: *CoRR* abs/2003.08934 (2020). arXiv: 2003.08934. URL: <https://arxiv.org/abs/2003.08934>.
- [2] Pak-Hei Yeung et al. "Implicitvol: Sensorless 3d ultrasound reconstruction with deep implicit representation". In: *arXiv preprint arXiv:2109.12108* (2021).
- [3] Magdalena Wysocki et al. *Ultra-NeRF: Neural Radiance Fields for Ultrasound Imaging*. 2023. arXiv: 2301.10520 [eess.IV].
- [4] Thomas Müller et al. "Instant Neural Graphics Primitives with a Multiresolution Hash Encoding". In: *ACM Trans. Graph.* 41.4 (July 2022), 102:1–102:15. DOI: 10.1145/3528223.3530127. URL: <https://doi.org/10.1145/3528223.3530127>.
- [5] François Gaits et al. "Efficient stratified 3D scatterer sampling for freehand ultrasound simulation". In: 2023. DOI: 10.1109/TUFFC.2023.3324014.