Image Reconstruction via Variational Network for Real-Time Hand-Held Sound-Speed Imaging

Valery Vishnevskiy$^{(\boxtimes)}$, Sergio J. Sanabria, and Orcun Goksel

Computer-assisted Applications in Medicine Group, ETH Zurich, Zurich, Switzerland
vishnevskiy@biomed.ee.ethz.ch

Abstract. Speed-of-sound is a biomechanical property for quantitative tissue differentiation, with great potential as a new ultrasound-based image modality. A conventional ultrasound array transducer can be used together with an acoustic mirror, or so-called reflector, to reconstruct sound-speed images from time-of-flight measurements to the reflector collected between transducer element pairs, which constitutes a challenging problem of limited-angle computed tomography. For this problem, we herein present a variational network based image reconstruction architecture that is based on optimization loop unrolling, and provide an efficient training protocol of this network architecture on fully synthetic inclusion data. Our results indicate that the learned model presents good generalization ability, being able to reconstruct images with significantly different statistics compared to the training set. Complex inclusion geometries were shown to be successfully reconstructed, also improving over the prior-art by 23% in reconstruction error and by 10% in contrast on synthetic data. In a phantom study, we demonstrated the detection of multiple inclusions that were not distinguishable by prior-art reconstruction, meanwhile improving the contrast by 27% for a stiff inclusion and by 219% for a soft inclusion. Our reconstruction algorithm takes approximately 10 ms, enabling its use as a real-time imaging method on an ultrasound machine, for which we are demonstrating an example preliminary setup herein.

Keywords: Deep learning · Speed-of-sound · Image reconstruction

1 Introduction

Speed-of-sound (SoS) ultrasound computed tomography (USCT) is a promising image modality, which generates maps of speed of sound in tissue as an imaging biomarker. Potential clinical applications are differentiation of breast tumorous lesions [3], breast density assessment [13,15], staging of musculoskeletal [11] and non-alcoholic fatty liver disease [7], amongst others. For this, a set of time of flight (ToF) measurements through the tissue between pairs of transmit/receive elements of an ultrasonic array can be used for a tomographic reconstruction.

Various 2D ad 3D acquisition setups have been proposed, including circular or dome-shaped transducer geometries, which provide multilateral set of measurements that are convenient for reconstruction methods [8] but costly to manufacture and cumbersome in use. Hand-held reflector based setup [10,14] depicted in Fig. 1a uses a conventional portable ultrasound probe to measure ToF via wave reflections of a plate placed on the opposite side of the sample. Despite its simplicity, such a setup results in limited-angle (LA) CT, which requires prior assumptions and suitable regularization and numerical optimization techniques to produce meaningful reconstructions [14]. Such optimization techniques may not be guaranteed to converge, are often slow in runtime, and involve parameters that are difficult to set.

In this paper, we propose a problem-specific variational network [1,5] for limited-angle SoS reconstruction, with parameters learned from numerous forward simulations. Contrary to machine learning methods based on sinogram inpainting [16] and reconstruction artefact removal [6] for LA-CT, we learn reconstruction process end-to-end, and show that it allows to qualitatively improve conventional reconstruction.

2   Methods

Using the wave reflection tracking algorithm described in [14], we measure the ToF Δt between transmit (Tx) and receive (Rx) transducers in a $M = 128$ element linear ultrasound array (see Fig. 1a). Discretizing corresponding ray paths using a Gaussian sampling kernel, the inverse of ToF can be expressed as a linear combination of tissue slowness values x [s/m], i.e. $(\Delta t)^{-1} = \sum_{i \in \mathrm{Ray}} l_i x_i$. Considering a Cartesian $n_1 \times n_2 = P$ grid, we define the forward model

$$\mathbf{b} = \mathrm{diag}(\mathbf{m})\mathbf{L}\mathbf{x} + \mathcal{N}(\mathbf{0}, \sigma_N \mathbf{I}), \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^P$ is the inverse SoS (slowness) map, $\mathbf{L} \in \mathbb{R}^{M^2 \times P}$ is a sparse path matrix defined by acquisition geometry and discretization scheme, $\mathbf{m} \in \{0, 1\}^{M^2}$ is the undersampling mask with zeros indicating a missing (e.g., unreliable) ToF measurement between a corresponding Tx-Rx pair, and $\mathbf{b} \in \mathbb{R}^{M^2}$ is a zero-filled vector of measured inverse ToFs $(\Delta t)^{-1}$. Reconstructing a slowness map $\mathbf{x}$ is a process inverse to (1) and can be posed as the following convex optimization problem:

$$\hat{\mathbf{x}}(\mathbf{b}, \mathbf{m}; \lambda, \boldsymbol{\nabla}) = \underset{\mathbf{x}}{\mathrm{argmin}} \ \|\mathrm{diag}(\mathbf{m})\mathbf{L}\mathbf{x} - \mathbf{b}\|_1 + \lambda \|\boldsymbol{\nabla}\mathbf{x}\|_1, \tag{2}$$

which we solve using ADMM [2] algorithm with Cholesky factorization. Here $\boldsymbol{\nabla}$ is a matrix, and λ is the regularization weight.

It is common to choose regularization matrix $\boldsymbol{\nabla}_{\mathrm{TV}}$ that implements spatial gradients on Cartesian grid, yielding the total variation (TV) regularization [12], which allows to efficiently recover sharp image boundaries, but can introduce

signal underestimation and staircase artefacts that are amplified by the limited-angle acquisition. In attempt to remedy this problem, one can delicately construct a set of image filters that will penalize problem-specific reconstruction artefacts. We follow [14] and use regularization matrix $\boldsymbol{\nabla}_{\mathrm{MATV}}$ that implements convolution with the set of weighted directional gradient operators. This weights regularization according to known wave path information, such that the locations with information from a narrower angular range are regularized more.
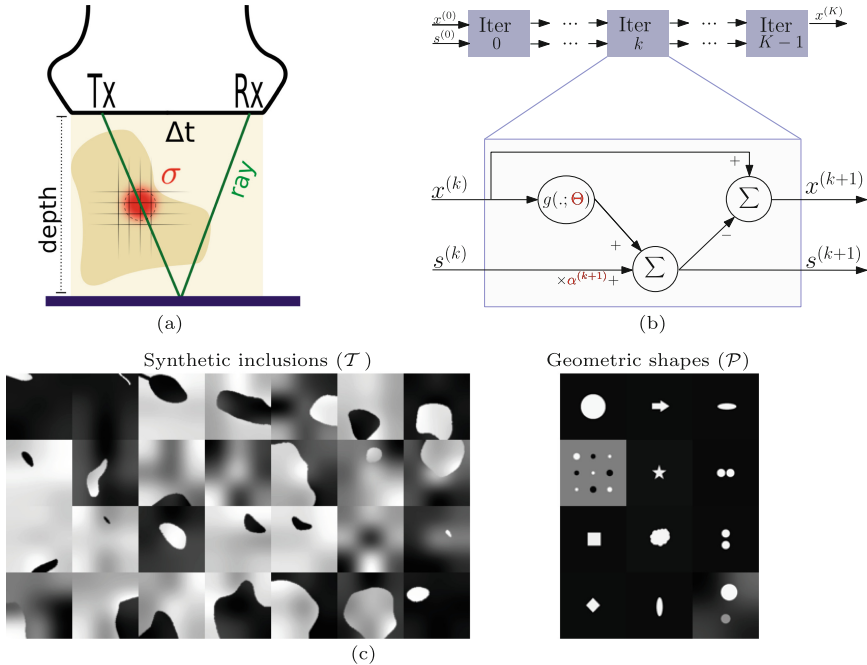
Fig. 1. (a) Acquisition setup and ray tracing discretization. (b) Structure of variational network; tunable parameters of the layer are highlighted in red. (c) Samples from synthetic training set $\mathcal{T}$ and testing set of geometric primitives $\mathcal{P}$. (Color figure online)

2.1   Variational Network

Variational networks (VN) is a class of deep learning methods that incorporate a parametrized prototype of a reconstruction algorithm in differentiable manner. A successful VN architecture proposed by Hammernik et al. in [5] for undersampled MRI reconstruction unrolls a fixed number of iterations of the gradient descent (GD) algorithm applied to a virtual optimization-based reconstruction problem. By unrolling the iterations of the algorithm into network layers (see Fig. 1b), the output is expressed as a formula parametrized by the regularization parameters and step lengths of this GD algorithm. The parameters are then tuned on retrospectively undersampled training data.

In contrast to discrete Fourier transform, the design matrix of LA-CT is poorly conditioned, which compromises the performance of conventional GD. Therefore, we propose to enhance the VN in the following ways: (i) unroll GD with momentum, (ii) add left diagonal preconditioner $\mathbf{p}^{(k)} \in \mathbb{R}^{M^2}$ for the path matrix $\mathbf{L}$, (iii) use adaptive data consistency term $\varphi_{\mathrm{d}}^{(k)}$, and (iv) allow spatial filter weighting $\mathbf{w}_i^{(k)} \in \mathbb{R}^P$. The resulting reconstruction network is defined in Algorithm 1 with tunable parameters Θ, where each of K variational layers contains N_f convolution matrices $\mathbf{D} = \mathbf{D}(\mathbf{d})$ with $N_c \times N_c$ kernels $\mathbf{d}$ that are ensured to be zero-centered unit-norm via re-parametrization: $\mathbf{d} = (\mathbf{d}' - \langle \mathbf{d}' \rangle)/\|(\mathbf{d}' - \langle \mathbf{d}' \rangle)\|_2$, where $\langle . \rangle$ denotes mean value of the vector. Each filter $\mathbf{D}$ is associated with its potential function $\varphi_{\mathrm{r}}\{.\}$ that is parametrized via cubic interpolation of control knots $\boldsymbol{\phi}_{\mathrm{r}} \in \mathbb{R}^{N_g}$ placed on Cartesian grid on $[-r, r]$ interval. Data term potentials $\varphi_{\mathrm{d}}\{.\}$ are defined in the same way. The network is trained to minimize ℓ_1-norm of the reconstruction error on the training set $\mathcal{T}$:

$$\min_{\Theta} \ \mathbb{E}_{\{\mathbf{b},\mathbf{m},\mathbf{x}^\star\} \in \mathcal{T}} \|\mathcal{V}(\mathbf{b}, \mathbf{m}; \Theta) - \mathbf{x}^\star\|_1. \tag{3}$$

Input: $\mathbf{b}$ — inverse ToF, $\mathbf{m}$ — undersampling mask
Parameters: $\Theta = \{\boldsymbol{\phi}_{\mathrm{d}}^{(k)}, \boldsymbol{\phi}_{\mathrm{r},i}^{(k)}, \mathbf{p}^{(k)}, \mathbf{w}_i^{(k)}, \mathbf{D}_i^{(k)}, \alpha^{(k)}\}_{i=1,\dots,N_f,\ k=1,\dots,K}$
$\mathbf{x}^{(0)} \leftarrow \alpha^{(0)} \mathbf{L}^\mathsf{T} \mathbf{b}$, $\mathbf{s}^{(0)} \leftarrow \mathbf{0}$
for $k := 0$ to $K-1$
    $\mathbf{g}^{(k)} \leftarrow \mathbf{L}^\mathsf{T} \mathrm{diag}(\mathbf{p}^{(k)}) \mathrm{diag}(\mathbf{m}) \varphi_{\mathrm{d}}^{(k)} \left\{ \mathrm{diag}(\mathbf{m}) \mathrm{diag}(\mathbf{p}^{(k)}) \left(\mathbf{L}\mathbf{x}^{(k)} - \mathbf{b} \right) \right\} +$
              $\sum_{i=1,\dots,N_f} \left(\mathbf{D}_i^{(k)} \right)^\mathsf{T} \mathrm{diag}(\mathbf{w}_i^{(k)}) \varphi_{\mathrm{r},i}^{(k)} \left\{ \mathrm{diag}(\mathbf{w}_i^{(k)}) \mathbf{D}_i^{(k)} \mathbf{x}^{(k)} \right\}$
    $\mathbf{s}^{(k+1)} \leftarrow \alpha^{(k+1)} \mathbf{s}^{(k)} + \mathbf{g}^{(k)}$
    $\mathbf{x}^{(k+1)} \leftarrow \mathbf{x}^{(k)} - \mathbf{s}^{(k+1)}$
Output: reconstructed image $\mathcal{V}(\mathbf{b}, \mathbf{m}; \Theta) := \mathbf{x}^{(K)}$

Algorithm 1: Proposed variational reconstruction network model VNv4.

Training. Dataset $\mathcal{T}$ is generated using fixed acquisition geometry with reflector depth equal to transducer array width. High-resolution (HR) 256×256 synthetic inclusion masks are produced by applying smooth deformation to an ellipse with random center, eccentricity, and radius. Two smooth slowness maps with random values from $[1/1650, 1/1350]$ interval are then blended with this inclusion mask, yielding a final slowness map $\mathbf{x}_{\mathrm{HR}}^\star$ (see Fig. 1c). The chosen range corresponds to observed SoS values for breast tissues of different densities and tumorous inclusions of different pathologies [4]. Forward path matrix $\mathbf{L}_{\mathrm{HR}}$ and random incoherent undersmapling mask $\mathbf{m}$ are used to generate noisy inverse ToF vector $\mathbf{b}$ according to model (1) with $\sigma_N = 2 \cdot 10^{-8}$. Finally, we downsample $\mathbf{x}_{\mathrm{HR}}^\star$ to $n_1 \times n_2$ size yielding the ground truth map $\mathbf{x}^\star$. About 10% of maps did not contain inclusions. For each reconstruction problem the path matrix $\mathbf{L}$ is normalized with its largest singular value, and inverse ToF are centered and scaled: $\mathbf{b}' = \mathbf{b} - (\langle \mathbf{b} \rangle / \langle \mathbf{L}\mathbf{1} \rangle) \mathbf{L}\mathbf{1}$, $\tilde{\mathbf{b}} = \mathbf{b}'/\mathrm{std}(\mathbf{b}')$.

The configuration of networks were the following: $K = 10$, $N_f = 50$, $N_c = 5$, $n_1 = n_2 = 64$, $N_g = 55$. All parameters were initialized from $\mathcal{U}(0, 1)$. We refer to this architecture as VNv4. Ablating spatial filter weighting $\mathbf{w}_i^{(k)}$ from VNv4, we get VNv3; additionally ablating adaptive data potentials $\varphi_d^{(k)}$, we get VNv2; further ablating preconditioner $\mathbf{p}^{(k)}$, we get VNv1; and eventually unrolling GD without momentum, VNv0. For tuning the aforementioned models we used 10^5 iterations of Adam algorithm [9] with learning rate 10^{-3} and batch size 25. Every 5000 iterations we readjust potential function's interval range r by setting it to the maximal observed value of the corresponding activation function argument.

3   Results

We compare TV and MA-TV against VN architectures on (i) 200 samples from $\mathcal{T}$ that were set aside and unseen during training, and (ii) a set $\mathcal{P}$ of 14 geometric primitives depicted in Fig. 1c, using following metrics:

$$\text{SAD}(\mathbf{x}, \mathbf{y}) = \frac{\|\mathbf{x} - \mathbf{y}\|_1}{P}, \quad \text{CR} = \frac{2\,|\mu_{\text{inc}} - \mu_{\text{bg}}|}{|\mu_{\text{inc}}| + |\mu_{\text{bg}}|}, \quad \text{CRf} = \frac{\text{estimated CR}}{\text{ground truth CR}}, \quad (4)$$

where μ_{inc} and μ_{bg} are mean values in the inclusion and background regions accordingly. The optimal regularization weight λ for TV and MA-TV algorithms was tuned to give the best (lowest) SAD on the P3 image (see Fig. 2). Similarly to training generation, the forward model for validation and test sets was computed on high resolution images with normal noise and 30% undersmapling.

Quantitative evaluation on synthetic data is reported in Table 1 and shows that the proposed VNv4 network outperforms conventional TV and MA-TV reconstruction methods both in terms of accuracy and contrast. Comparing VNv options, it can be seen that richer architectures performed better. Figure 2 shows qualitative evaluation of reconstruction methods. VNv4 is able to reconstruct multiple inclusions (P5), handle smooth SoS variation (T1), and generally maintain inclusion position and geometry without hallucinating nonexistent inclusions. Although for some geometries (e.g. P4) TV reconstruction has lower SAD value, VNv4 provides better contrast, which allows to separate the two inclusions. As expected from the limited-angle nature of the data, highly elongated inclusions that are parallel to the reflector either undergo axial geometric distortion (P1), or could not be adequately reconstructed (T3) by any presented method.

Breast Phantom Experiment. We also compared the reconstruction methods using a realistic breast elastography phantom (Model 059, CIRS Inc.) that mimics glandular tissue with two lesions of different density. Portable ultrasound system (UF-760AG, Fukuda Denshi Inc., Tokyo, Japan) streams full-matrix RF ultrasound data over a high bandwidth link to a dedicated PC, which is used to perform USCT reconstruction and output a live SoS video feedback (cf. Fig. 4).
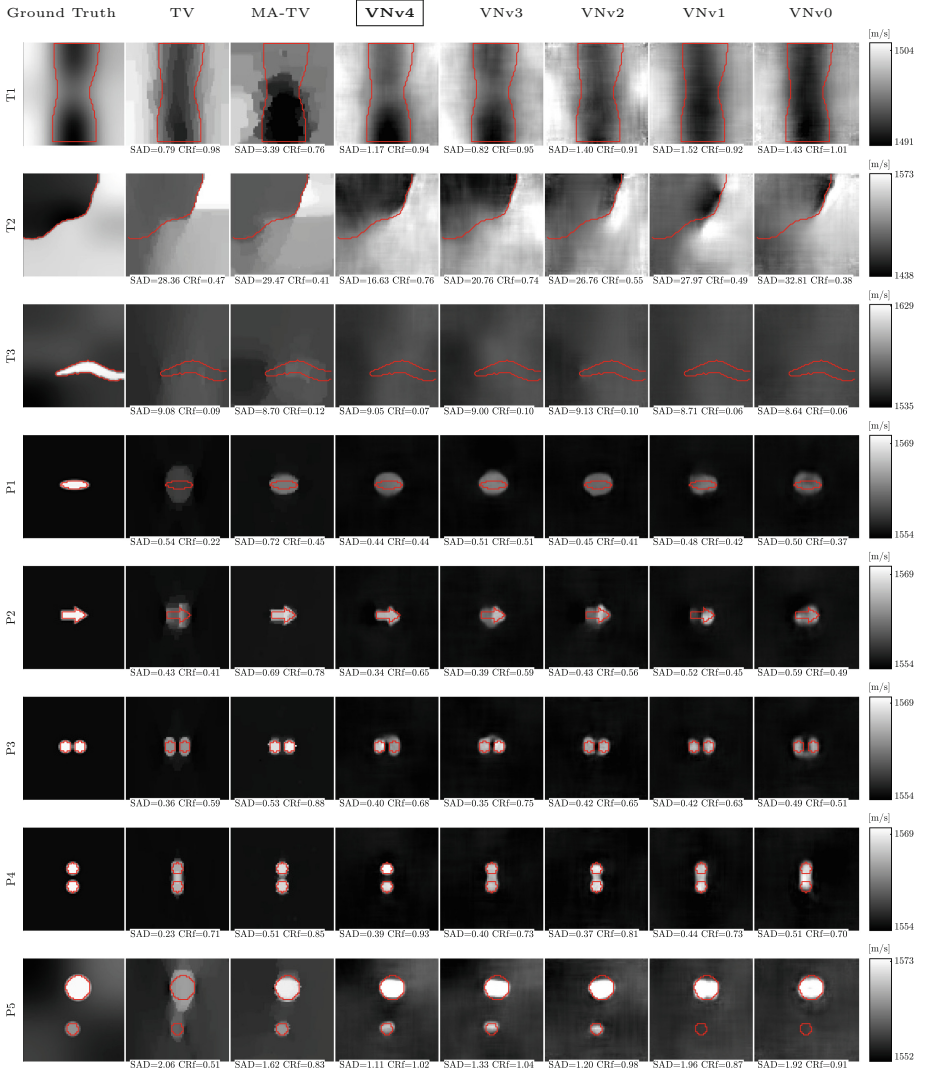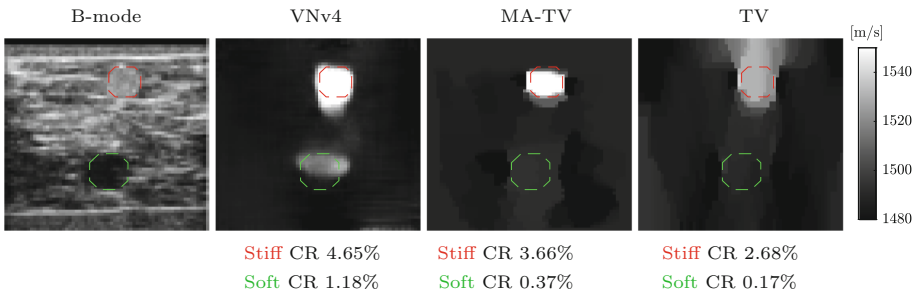
Fig. 2. Sound speed reconstructions of synthetic data from sets $\mathcal{T}$, $\mathcal{P}$ and single natural image N1. Inclusions are delineated with red curves. For each image transducer array is placed on top and reflector on bottom. (Color figure online)

We used an ultrasound probe (FUT-LA385-12P) with 128 piezoelectric transducer elements. For each frame a total of 128×128 RF lines are generated for all Tx/Rx combinations, at an imaging center frequency of 5 MHz digitized at 40.96 MHz. As seen in Fig. 3, VNv4 qualitatively outperforms both TV and MA-TV methods, showing clearly distinguishable hard and soft lesions. Run-time of MA-TV and TV algorithms on CPU is ∼30 s per image, while VN reconstruction takes ∼0.4 s on CPU and ∼0.01 s on GPU.

Table 1. SoS reconstruction measures computed on 200 validation images from training distribution $\mathcal{T}$, 14 test images from the set of geometric primitives $\mathcal{P}$.

Shape set	TV		MA-TV		VNv4		VNv3		VNv2		VNv1		VNv0	
	SAD	CRf	SAD	CRf	SAD	CRf	SAD	CRf	SAD	CRf	SAD	CRf	SAD	CRf
Synthetic inc. ($\mathcal{T}$)	7.27	0.49	7.64	0.53	**5.46**	**0.71**	5.91	0.66	6.77	0.59	7.56	0.47	7.96	0.43
Geometric shapes ($\mathcal{P}$)	0.54	0.63	0.72	0.84	**0.51**	**0.79**	0.60	0.77	0.62	0.73	0.77	0.60	0.78	0.57
Average	3.90	0.56	4.18	0.68	**2.99**	**0.75**	3.26	0.71	3.69	0.66	4.16	0.53	4.37	0.50

B-mode   VNv4   MA-TV   TV   [m/s]

Stiff CR 4.65%   Stiff CR 3.66%   Stiff CR 2.68%
Soft CR 1.18%   Soft CR 0.37%   Soft CR 0.17%

Fig. 3. Hand-held SoS mammography of the breast phantom. Stiff (red) and soft (green) inclusions were delineated in the B-mode image. (Color figure online)
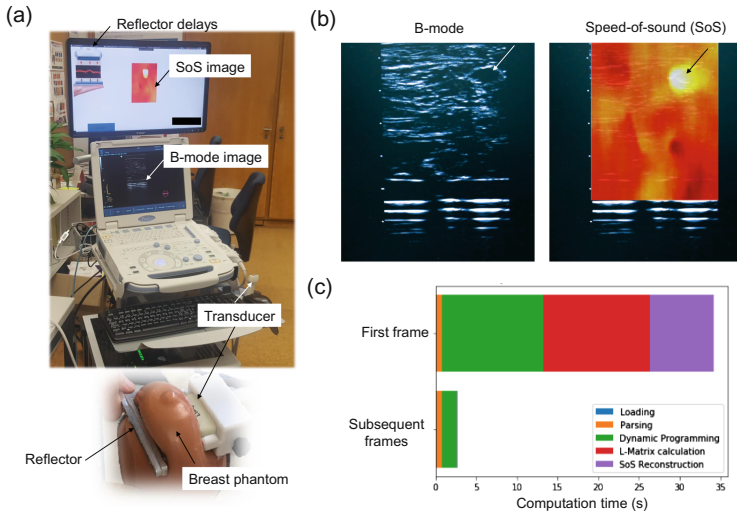
Fig. 4. Live SoS imaging demonstration. (a) Experimental setup. (b) Sample outputs of B-mode and SoS video feedback. A non-echogenic stiff lesion is clearly delineated in the SoS image. (c) Computational benchmarks, also showing initialization and memory allocation times. After initialization, SoS reconstruction time is negligible compared to data transfer and reflector ToF measurement via dynamic programming [14].

4 Discussion

In this paper we have proposed a deep variational reconstruction network for hand-held US sound-speed imaging. The method is able to reconstruct various inclusion geometries both in synthetic and phantom experiments. VN demonstrated good generalization ability, which suggests that unrolling even more sophisticated numerical schemes may be possible. Improvements over conventional reconstruction algorithms are both qualitative and quantitative. The ability of method to distinguish hard and soft inclusions has great diagnostic potential in characterizing lesions in real-time.

References

1. Adler, J., Öktem, O.: Solving ill-posed inverse problems using iterative deep neural networks. Inverse Probl. **33**(12), 124007 (2017)
2. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends ML **3**(1), 1–122 (2011)
3. Duric, N., et al.: In-vivo imaging results with ultrasound tomography: report on an on-going study at the Karmanos cancer institute. In: SPIE Med Imaging, pp. 76290M–1/M–9 (2010)
4. Duric, N., et al.: Detection of breast cancer with ultrasound tomography: first results with the computed ultrasound risk evaluation (cure) prototype. Med. Phys. **34**(2), 773–785 (2007)
5. Hammernik, K., Klatzer, T., Kobler, E., Recht, M.P., Sodickson, D.K., Pock, T., Knoll, F.: Learning a variational network for reconstruction of accelerated MRI data. MRM **79**, 3055–3071 (2018)
6. Hammernik, K., Würfl, T., Pock, T., Maier, A.: A deep learning architecture for limited-angle computed tomography reconstruction. Bildverarbeitung für die Medizin **2017**, 92–97 (2017)
7. Imbault, M., et al.: Robust sound speed estimation for ultrasound-based hepatic steatosis assessment. Phys. Med. Biol. **62**(9), 3582 (2017)
8. Jirik, R., et al.: Sound-speed image reconstruction in sparse-aperture 3-D ultrasound transmission tomography. IEEE Trans. Ultrason. Ferroelectr. Freq. Control **59**(2) (2012)
9. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
10. Krueger, M., Burow, V., Hiltawsky, K., Ermert, H.: Limited angle ultrasonic transmission tomography of the compressed female breast. In: Ultrasonics Symposium, vol. 2, pp. 1345–1348. IEEE (1998)
11. Qu, X., et al.: Limb muscle sound speed estimation by ultrasound computed tomography excluding receivers in bone shadow. In: SPIE Medical Imaging, p. 101391 (2017)
12. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena **60**(1–4), 259–268 (1992)
13. Sak, M., et al.: Comparison of breast density measurements made with ultrasound tomography and mammography. In: SPIE Medical Imaging, pp. 94190R–1/R–8 (2015)

14. Sanabria, S.J., Goksel, O.: Hand-held sound-speed imaging based on ultrasound reflector delineation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9900, pp. 568–576. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46720-7_66

15. Sanabria, S.J., et al.: Breast-density assessment with hand-held ultrasound: a novel biomarker to assess breast cancer risk and to tailor screening? Eur. Radiol. (2018)

16. Tovey, R., et al.: Directional sinogram inpainting for limited angle tomography. arXiv preprint arXiv:1804.09991 (2018)