Model-Aware Deep Learning Based Strategies for Clutter Suppression in Contrast Enhanced Ultrasound

Oren Solomon, *Student Member, IEEE,* Regev Cohen, *Student Member, IEEE,* Yi Zhang, Yi Yang, He Qiong, De Ma, Ruud J.G. van Sloun, *Member, IEEE*, Zhifei Dai, Jianwen Luo, and Yonina C. Eldar, *Fellow, IEEE*

***Abstract*—Ultrasonography is a low-cost, radiation-free in-vivo imaging modality. In particular, blood vessel imaging and the measurement of their physical parameters, such as blood flow velocities are of great clinical interest. One prominent solution involves the use of encapsulated gas microbubbles for improved visualization of the vascular bed, which has recently enabled imaging with unprecedented sub-wavelength spatial resolution. A typical preprocessing step to such operations is first separating the microbubble signal from the cluttering tissue signal. Here, we propose two main contributions. The first, is adopting a new model for the acquired contrast enhanced ultrasound signal, which was similarly suggested in the context of ultrasound Doppler processing and dynamic magnetic resonance imaging, and is typically solved via an iterative algorithm. This model leads to a better separation of microbubble signal from the tissue signal than commonly practiced methods. The second, is to utilize the recently proposed technique of learning fast sparse approximations from the field of deep learning to suggest a suitable network architecture for this task with improved convergence speed and accuracy over its iterative counterpart. We compare the performance of the suggested network on both simulations and *in-vivo* rat brain scans, with a commonly practiced deep-network architecture and the fast iterative shrinkage/thresholding algorithm. We show that our architecture exhibits better image quality and contrast.**

***Index Terms*—Ultrasound, Deep learning, Unfolding, Contrast agents, Blood.**

I. Introduction

Medical ultrasound (US) has been used extensively in the past decades for radiation-free diagnosis of a wide range of diseases. Due to its low-cost and real-time capabilities it impacts various clinical segments such as radiology, cardiology, vascular, obstetrics and emergency medicine. Ultrasound-based imaging modalities include brightness, motion, Doppler, harmonic modes, elastography and more [1].

One important imaging modality is contrast-enhanced ultrasound (CEUS) [2] which allows the detection and visualization of blood vessels whose different physical parameters such as relative blood volume (rBV) and velocity, shape and density are associated with different clinical conditions [3]. CEUS uses encapsulated gas microbubbles as ultrasound contrast agents (UCAs) which are administrated intravenously and are similar in size to red blood cells and thus can flow throughout the vascular system [4]. Among its many applications, CEUS is used for imaging of perfusion at the capillary level [5, 6], for estimating blood velocity in small vessels arterioles by applying Doppler processing [7, 8] and for sub-wavelength vascular imaging [9–12].

A major challenge in CEUS is to suppress clutter signals stemming from stationary and slowly moving tissue as they introduce significant artifacts in blood flow imaging [13]. Over the past few decades several approaches have been suggested for clutter removal. The simplest approach to remove tissue signal is to filter the ultrasonic signal along the temporal dimension using finite impulse response (FIR) or infinite impulse response (IIR) filters [14]. However, FIR filters must be high-order filters while IIR filters exhibit a long settling time which lead to a low number of temporal samples in each spatial location [15] when using focused transmission. Similar strategy is based on estimation of the tissue velocity and its subsequent removal by standard high-pass filtering [16]. However, the above methods rely on the assumption that tissue motion, if exists, is slow while blood flow is fast. Hence, this high-pass filtering approach is prone to failure in the presence of fast tissue motion, as in cardiology, or when imaging microvasculature in which blood velocities are low.

An alternative approach for tissue suppression is second harmonic imaging [17], which separates the blood and tissue signals by exploiting the non-linear response of the UCAs to low acoustic pressures, compared with the mostly linear tissue response. This technique, however, limits the frame-rate of the US scanner, and does not remove the tissue signal completely, as tissue can also exhibit a nonlinear response.

All of the above are based only on temporal information and neglect the high spatial coherence of the tissue, compared to the blood. To take advantage of these spatial characteristics of tissue, a method for clutter removal was presented in [18], based on singular value decomposition (SVD) of the spatiotemproal correlation matrix of successive temporal samples. SVD filtering operates by stacking the (typically

O. Solomon and R. Cohen contributed equally to this work.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 646804-ERC-COG-BNYQ.

O. Solomon (e-mail: orensol@campus.technion.ac.il), R. Cohen (e-mail: regev,cohen@campus.technion.ac.il) and Y. C. Eldar (e-mail: yonina@ee.technion.ac.il) are with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa 32000. R. J. G. van Sloun (e-mail: R.J.G.v.Sloun@tue.nl) and M. Mischi (e-mail: M.Mischi@tue.nl) are with the Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands. Y. Yang, Q. He and J Luo are with the Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China. D. Ma and Z. Dai are with the Department of Biomedical Engineering, Peking University, Beijing 100871, China.

beamformed) acquired frames as vectors in a matrix whose column index indicates frame number and then applies SVD decomposition to it. The largest eigen-values correspond to the highly correlated tissue, and are discarded. Thus, a new US movie is recomposed from the remaining singular values to produce the blood/UCA movie.

This latter approach led to the development of numerous SVD-based techniques [13, 19–23], such as the down-mixing technique proposed in [13] for tissue motion estimation, the real-time implementation of eigen-based clutter rejection proposed by Lovstakken *et al.* [24] and the principal component analysis for blood velocity estimation presented in [25]. Unfortunately, these methods are based on focused transmission beams which lead to a small number of temporal and spatial samples, thus limiting the impact of SVD-based filtering. To overcome this, SVD-based clutter removal was adapted for ultrafast plane-wave imaging [26–29] and demonstrated substantially improved clutter rejection and microvascular extraction. This strategy gained a lot of interest in recent years and nowadays it is used in numerous ultrafast US imaging applications such as functional ultrasound [30, 31], super-resolution ultrasound localization microscopy [27] and high-sensitivity microvessel perfusion imaging [26, 28].

However, this SVD-based approach suffers from two major limitations. First, direct application of SVD-based clutter filtering is computationally expensive [32], making it unsuitable for real-time application. To reduce computational cost, the authors of [33] proposed a stochastic approach based on randomized SVD (rSVD), developed by Halko *et al.* [34] and optimized by Voronin and Martinsson [35], and on decomposing the large samples matrix into smaller blocks that can be processed in parallel. Second, SVD-based filtering requires to determine a threshold, differentiating between tissue related and blood related singular-values, whose appropriate setting is typically unclear, as in typical acquisitions, the eigenvalue spectra of the tissue and contrast signals tend to overlap. This overlap and threshold uncertainty motivate the use of a different model for the acquired data, one that can clearly differentiate between tissue and contrast signals based on the aforementioned spatio-temporal information, as well as additional information unique to the contrast signal - their sparse distribution in the imaging plane.

Here, we propose two main contributions. The first, is the adaptation of a new model for the tissue/contrast separation problem. We show that similar to other applications such as MRI [36] and recent US Doppler applications [37], we can decompose the acquired, beamformed US movie as a sum of a low-rank matrix (tissue) and a sparse outlier signal (UCAs). This decomposition is also known as the robust principle component analysis (RPCA) [38]. We then propose to solve a convex minimization problem to retrieve the UCA signal, which leads to an iterative algorithm. Second, we propose to utilize recent ideas from the field of deep learning [39] to dramatically improve the convergence rate and image reconstruction quality of the proposed iterative algorithm. We do so by unfolding [40] the iterative algorithm as a fixed-length deep network which is trained on sets of separated tissue/UCA signals from both *in-vivo* and simulated data. Similar to [41], we utilize convolution layers instead of fully-connected (FC) layers, to exploit the shared spatial information between neighboring image pixels. Our training policy is a two stage process. We start by training the network on simulated data, and then train the resulting network on *in-vivo* data. This hybrid policy allows us to improve the network's performance and to achieve a fully-automated network, in which all the regularization parameters are also learned. We compare the performance of the proposed network with the commonly practiced SVD approach, the iterative algorithm and an adaptation of ResNet. We show that the unfolded network outperforms all other approaches in over all image quality and contrast.

Unfolding, or unrolling an iterative algorithm was suggested by Gregor and LeCun [40] to accelerate algorithm convergence. In the context of deep learning, typically one of the first dilemmas is what type of network architecture to use. Iterative algorithms provide a natural recurrent architecture, designed to solve a specific problem, such as sparse approximations, channel estimation [42] and more. The authors of [40] showed that by considering each iteration of an iterative algorithm as a layer in a deep network and subsequent concatenation of a few such layers it is possible to train such networks to achieve a dramatic improvement in convergence.

In the context of RPCA, a principled way to construct learnable pursuit architectures for structured sparse and robust low rank models was introduced in [43]. The proposed networks, derived from the iteration of proximal descent algorithms, were shown to faithfully approximate the solution of RPCA while demonstrating several orders of magnitude speed-up compared to standard optimization algorithms. However, this approach is based on a non-convex formulation in which the rank of the low-rank part (or an upper bound of it) is assumed to be known a priori. This poses a network design limitation, as the rank can vary between different applications or even different realizations of the same application, as in CEUS. Thus, for each choice of the rank upper bound, a new network needs to be trained, which can limit its applicability. On the contrary, our approach is rank free, and does not require a-priory knowledge of this upper bound.

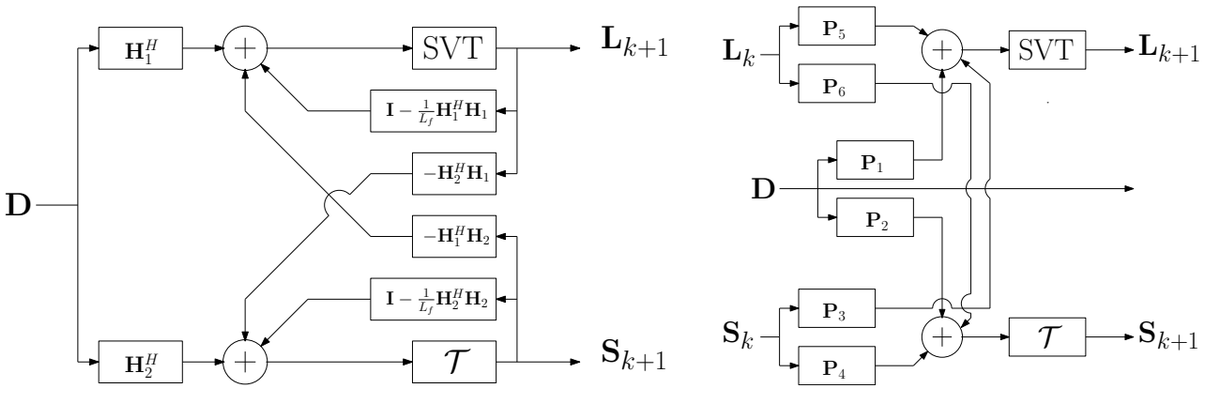
The rest of the paper is organized as follows. In Section II we introduce the mathematical formulation of the low-rank and sparse decomposition. Section V presents *in-silico* as well as *in-vivo* results of both the iterative algorithm and proposed deep networks. Lastly, we discuss the results, limitations and further research directions in Section VI.

Throughput the paper, x represents a scalar, $\mathbf{x}$ a vector, $\mathbf{X}$ a matrix and $\mathbf{I}_{N\times N}$ is the $N \times N$ identity matrix. The notation $||\cdot||_p$ represents the standard p-norm and $||\cdot||_F$ is the Frobenius norm. Subscript x_l denotes the lth element of $\mathbf{x}$ and $\mathbf{x}_l$ is the lth column of $\mathbf{X}$. Superscript $\mathbf{x}^{(p)}$ represents $\mathbf{x}$ at iteration p, T^* denotes the adjoint of T, and $\bar{\mathbf{A}}$ is the complex conjugate of $\mathbf{A}$.

II. Deep learning strategy for RPCA in US

A. Problem formulation

We start by first providing a model for the acquired US signal. In US imaging, typically a series of pulses are trans-



(a) Iterative algorithm for sparse and low-rank separation.

(b) Single layer from the corresponding unfolded algorithm.

Fig. 1: Architecture comparison between the iterative algorithm (panel (a)) and its unfolded counterpart (panel (b)). The learned network in panel (b) draws its architecture from the iterative algorithm, and is trained on examples from a given dataset. In both panels, $\mathbf{D}$ is the input measurements matrix, $\mathbf{S}_k$ and $\mathbf{L}_k$ are the estimated sparse and low-rank matrices in each iteration/layer, respectively.

mitted to the imaged medium. The resulting echoes from the sonicated medium are received in each transducer element and then combined in a process called beamforming to produce a focused image. As presented in [44], after demodulation the complex analytical (IQ) signal can be represented as

$$D(x,z,t) = I(x,z,t) + iQ(x,z,t),$$

where $I(x,z,t)$ and $Q(x,z,t)$ are the in-phase and quadrature components of the demodulated signal, x, z are the vertical and axial coordinates, and t indicates frame number. The signal $D(x,z,t)$ is a sum of echoes returned from the blood/CEUS signal $S(x,z,t)$ as well as from the tissue $L(x,z,t)$, contaminated by additive noise $N(x,z,t)$

$$D(x,z,t) = L(x,z,t) + S(x,z,t) + N(x,z,t).$$

By acquiring a series of movie frames $t = 1, \ldots, T$, and stacking them as vectors in a matrix leads to the following model

$$\mathbf{D} = \mathbf{L} + \mathbf{S} + \mathbf{N}. \tag{1}$$

Assuming that each movie frame is of size $M \times M$ pixels, each of the matrices of (1) are of size $M^2 \times T$. From here on, we consider a more general model, in which the acquired matrix $\mathbf{D}$ is composed as follows

$$\mathbf{D} = \mathbf{H}_1\mathbf{L} + \mathbf{H}_2\mathbf{S} + \mathbf{N}, \tag{2}$$

with $\mathbf{H}_1$ and $\mathbf{H}_2$ being the measurement matrices of appropriate dimensions. The model (2) can also be applied to MR imaging, video compression and additional US applications, as we discuss in Section VI. Similar to [36], we propose to solve the following minimization problem

$$\min_{\mathbf{L},\mathbf{S}} \frac{1}{2}||\mathbf{D} - (\mathbf{H}_1\mathbf{L} + \mathbf{H}_2\mathbf{S})||_F^2 + \lambda_1||\mathbf{L}||_* + \lambda_2||\mathbf{S}||_{1,2}, \tag{3}$$

where $||\cdot||_*$ stands for the nuclear norm and $||\cdot||_{1,2}$ is the mixed $l_{1,2}$ norm. We use the mixed $l_{1,2}$ norm since the pattern of the sparse outlier (blood or CEUS signal) is the same between different frames, and ultimately corresponds to the locations of the blood vessels, which are assumed to not change, or change very slowly during the acquisition period. On the other hand, the nuclear norm is known to promote low-rank solutions, and is the convex relaxation of the non-convex rank minimization constraint [45].

By defining

$$\mathbf{X} = \begin{bmatrix} \mathbf{L} \\ \mathbf{S} \end{bmatrix}, \ \mathbf{P}_1 = \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}, \ \mathbf{P}_2 = \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix}$$

and $\mathbf{A} = [\mathbf{H}_1, \mathbf{H}_2]$, (3) can be rephrased as

$$\min_{\mathbf{L},\mathbf{S}} \frac{1}{2}||\mathbf{D} - \mathbf{A}\mathbf{X}||_F^2 + h(\mathbf{X}), \tag{4}$$

where $h(\mathbf{X}) = \sum_{i=1}^{2} \lambda_i \rho_i(\mathbf{P}_i\mathbf{X})$ with $\rho_1 = ||\cdot||_*$ and $\rho_2 = ||\cdot||_{1,2}$. The minimization problem (4) is a regularized least-squares problem, for which numerous numerical minimization algorithms exist. Specifically, the (fast) iterative shrinkage/thresholding algorithm, (F)ISTA, [46, 47] involves finding *Moreau's proximal* (prox) mapping [48, 49] of h, defined as

$$\text{prox}_h(\mathbf{X}) = \underset{\mathbf{U}}{\text{argmin}} \left\{ h(\mathbf{U}) + \frac{1}{2}||\mathbf{U} - \mathbf{X}||_F^2 \right\}. \tag{5}$$

Plugging the definition of $\mathbf{X}$ into (5) yields

$$\text{prox}_h(\mathbf{X}) = \underset{\mathbf{U}_1,\mathbf{U}_2}{\text{argmin}} \left\{ \lambda_1\rho_1(\mathbf{U}_1) + \frac{1}{2}||\mathbf{U}_1 - \mathbf{L}||_F^2 \right.$$
$$\left. + \lambda_2\rho_2(\mathbf{U}_2) + \frac{1}{2}||\mathbf{U}_2 - \mathbf{S}||_F^2 \right\},$$

for which it can be shown that since $\text{prox}_h(\mathbf{X})$ separable in $\mathbf{L}$ and $\mathbf{S}$, it holds that

$$\text{prox}_h(\mathbf{X}) = \begin{bmatrix} \text{prox}_{\rho_1}(\mathbf{L}) \\ \text{prox}_{\rho_2}(\mathbf{S}) \end{bmatrix}. \tag{6}$$

The proximal mapping (6) is applied in each iteration to the gradient of the quadratic part of (4), given by

$$\frac{d}{d\mathbf{X}} \frac{1}{2}||\mathbf{D} - \mathbf{A}\mathbf{X}||_F^2 = \mathbf{A}^H(\mathbf{A}\mathbf{X} - \mathbf{D}),$$

and more specifically,

$$\begin{bmatrix} \frac{d}{d\mathbf{L}} \\ \frac{d}{d\mathbf{S}} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1^H(\mathbf{H}_1\mathbf{L} + \mathbf{H}_2\mathbf{S} - \mathbf{D}) \\ \mathbf{H}_2^H(\mathbf{H}_1\mathbf{L} + \mathbf{H}_2\mathbf{S} - \mathbf{D}) \end{bmatrix}.$$

Thus, the general iterative step of ISTA applied to minimizing (3) is given by

$$\begin{aligned} \mathbf{L}^{k+1} &= \text{SVT}_{\lambda_1/L_f}\left\{\left(\mathbf{I} - \tfrac{1}{L_f}\mathbf{H}_1^H\mathbf{H}_1\right)\mathbf{L}^k - \mathbf{H}_1^H\mathbf{H}_2\mathbf{S}^k + \mathbf{H}_1^H\mathbf{D}\right\} \\ \mathbf{S}^{k+1} &= \mathcal{T}_{\lambda_2/L_f}\left\{\left(\mathbf{I} - \tfrac{1}{L_f}\mathbf{H}_2^H\mathbf{H}_2\right)\mathbf{S}^k - \mathbf{H}_2^H\mathbf{H}_1\mathbf{L}^k + \mathbf{H}_2^H\mathbf{D}\right\} \end{aligned}, \tag{7}$$

where L_f is the Lipschitz constant of the quadratic term of (4), given by the spectral norm of $\mathbf{A}^H\mathbf{A}$ and SVT$(\cdot)$ and $\mathcal{T}(\cdot)$ are the singular value thresholding [50] and mixed $l_{1/2}$ soft thresholding operators [46]. The ISTA algorithm for minimizing (3) is summarized in Algorithm 1, while the diagram in Fig. 1(a) presents the iterative algorithm, which relies on knowledge of $\mathbf{H}_1, \mathbf{H}_2$ and selection of λ_1 and λ_2.

Algorithm 1 ISTA for minimizing (3)

Require: $\mathbf{D}$, $\lambda_1 > 0, \lambda_2 > 0$, maximum iterations K_{max}
  Initialize $\mathbf{S} = \mathbf{L} = \mathbf{0}$ and $k = 1$
  while $k \leq K_{\text{max}}$ or stopping criteria not fulfilled **do**
    1: $\mathbf{G}_{1_k} = \left(\mathbf{I} - \frac{1}{L_f}\mathbf{H}_1^H\mathbf{H}_1\right)\mathbf{L}^k - \mathbf{H}_1^H\mathbf{H}_2\mathbf{S}^k + \mathbf{H}_1^H\mathbf{D}$
    2: $\mathbf{G}_{2_k} = \left(\mathbf{I} - \frac{1}{L_f}\mathbf{H}_2^H\mathbf{H}_2\right)\mathbf{S}^k - \mathbf{H}_2^H\mathbf{H}_1\mathbf{L}^k + \mathbf{H}_2^H\mathbf{D}$
    3: $\mathbf{L}^{k+1} = \text{SVT}_{\lambda_1/L_f}\{\mathbf{G}_{1_k}\}$
    4: $\mathbf{S}^{k+1} = \mathcal{T}_{\lambda_2/L_f}\{\mathbf{G}_{2_k}\}$
    5: $k \leftarrow k + 1$
  end while
  return $\mathbf{S}_{K_{\text{max}}}, \mathbf{L}_{K_{\text{max}}}$

The dynamic range between returned echoes from the tissue and UCA/blood signal can range from 10dB to 60dB. As this dynamic range expands, more iterations are required to achieve good separation of the signals. This observation motivates the pursuit of a fixed complexity algorithm. Thus, we move in the next section to describe the concept of unfolding Algorithm 1. Additional description of learning fast sparse approximations is given in the supplementary materials.

B. Unfolding the iterative algorithm

As previously mentioned, an iterative algorithm can be considered as a recurrent neural network, in which the kth iteration can be regarded as the kth layer in a feedforward network [43]. On the other hand, convolutions are linear operators, thus considering (7), for each layer in the network we swap each of the matrices dependent on $\mathbf{H}_1$ and $\mathbf{H}_2$ with convolution layers (kernels) $\mathbf{P}_1^k, \ldots, \mathbf{P}_6^k$ of appropriate sizes. The kernels as well as the regularization parameters λ_1^k and λ_2^k, are learned in training. By doing so, the following equations for the kth layer are obtained

$$\begin{aligned} \mathbf{L}^{k+1} &= \text{SVT}_{\lambda_1^k}\left\{\mathbf{P}_5^k * \mathbf{L}^k + \mathbf{P}_3^k * \mathbf{S}^k + \mathbf{P}_1^k * \mathbf{D}\right\}, \\ \mathbf{S}^{k+1} &= \mathcal{T}_{\lambda_2^k}\left\{\mathbf{P}_6^k * \mathbf{L}^k + \mathbf{P}_4^k * \mathbf{S}^k + \mathbf{P}_2^k * \mathbf{D}\right\}, \end{aligned}$$

with $*$ being a convolution operator. The latter can be considered as a single layer in a multi-layer feedforward network. A diagram of a single layer from the unfolded architecture is given in Fig. 1(b), where the supposedly known model matrices were replaced by the 2D convolution kernels $\mathbf{P}_1^k, \ldots, \mathbf{P}_6^k$, which are learned as part of the training process of the overall network.

Contrary to previous works in unfolding RPCA which considered training fully connected (FC) layers [43], we employ convolution kernels in our implementation that allows us to achieve spatial invariance while reducing the number of learned parameters considerably. In many applications, the recovered matrices $\mathbf{S}$ and $\mathbf{L}$ represent a 3D volume, or movie, of dynamic objects imposed on a (quasi) static background. Each column in $\mathbf{S}$ and $\mathbf{L}$ is a vectorized frame from the recovered sparse and low-rank movies. Thus, we consider in practice our data as a 3D volume and apply 2D convolutions. The SVT operation at the kth layer is performed after reshaping the input 3D volume into a 2D matrix, by vectorizing and column-wise stacking each frame.

The thresholding coefficients are learned independently for each layer. Given the kth layer, the actual thresholding values for both the SVT and soft-thresholding operations are given by $\text{thr}_L^k = \sigma(\lambda_L^k) \cdot a_L \cdot \max(L^k)$ and $\text{thr}_S^k = \sigma(\lambda_S^k) \cdot a_S \cdot \text{mean}(S^k)$ respectively, where $\sigma(\cdot)$ is a sigmoid function, a_L and a_S are fixed scalars (in our application we chose $a_L = 0.4$ and $a_S = 1.8$) and λ_L^k and λ_S^k are learned in each layer by the network.

C. Training the unfolded network

The unfolded network presented in the previous section is trained using back-propagation in a supervised manner. Generally speaking, we obtain training examples $\mathbf{D}_i$ and corresponding sparse $\hat{\mathbf{S}}_i$ and low-rank $\hat{\mathbf{L}}_i$ decompositions. In practice, $\hat{\mathbf{S}}_i$ and $\hat{\mathbf{L}}_i$ can either be obtained from simulations or by decomposing $\mathbf{D}_i$ using iterative algorithms such as FISTA [47]. The loss function is chosen as the sum of the mean squared errors (MSE) between the predicted $\mathbf{S}$ and $\mathbf{L}$ values of the network and $\hat{\mathbf{S}}_i$, $\hat{\mathbf{L}}_i$, respectively,

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) = &\frac{1}{2N}\sum_{i=1}^{N}||f_S(\mathbf{D}_i, \boldsymbol{\theta}) - \hat{\mathbf{S}}_i||_F^2 \\ &+ \frac{1}{2N}\sum_{i=1}^{N}||f_L(\mathbf{D}_i, \boldsymbol{\theta}) - \hat{\mathbf{L}}_i||_F^2. \end{aligned} \tag{8}$$

In the latter equation, f_S/f_L is the sparse/low-rank output of the unfolded network with learnable parameters $\boldsymbol{\theta} = \{\mathbf{P}_1^k, \ldots, \mathbf{P}_6^k, \lambda_1^k, \lambda_2^k\}$, $k = 1, \ldots, K$, where K is the number of chosen layers.

Training a deep network typically requires a large amount of training examples, and in practice, US scans of specific organs are not available in abundance. To be able to train the unfolded network, we thus rely on two strategies: patch-based analysis and simulations. First, instead of training the network over entire scans of organs, we divide the US movie into 3D patches (axial coordinate, lateral coordinate and frame number) and apply Algorithm 1 on each of these 3D patches. The resulting separated UCA movie is then considered as the desirable outcome of the network and the network is trained over these pairs of extracted 3D patches from the acquired movie, and the resulting reconstructed UCA movies.

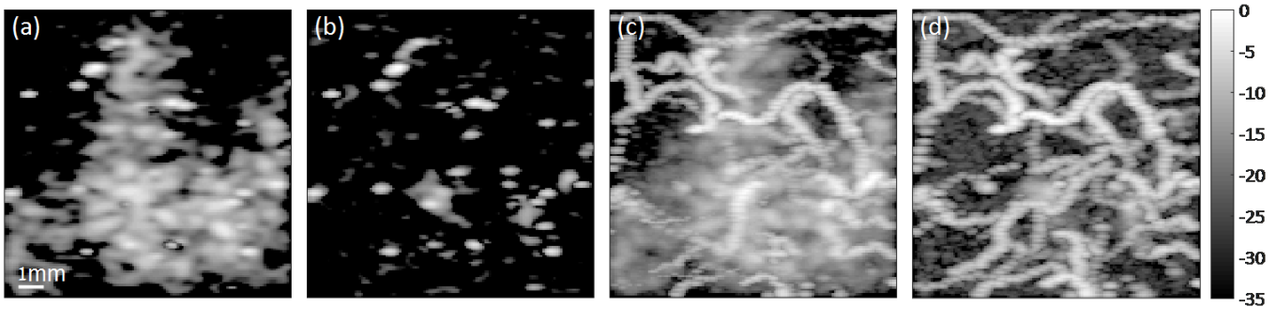


Fig. 2: Simulation results of the unfolded network. (a) A single recovered frame from a simulated movie, trained for 50 epochs on simulated data and for additional 20 epochs on experimental data. (b) A single recovered frame from a simulated movie, trained for 50 epochs on simulated data only. (c) MIP image for a recovered UCAs movie of 100 frames using the network in (a). (d) MIP image for a recovered UCAs movie of 100 frames using the network in (b). Color bar is in dB.

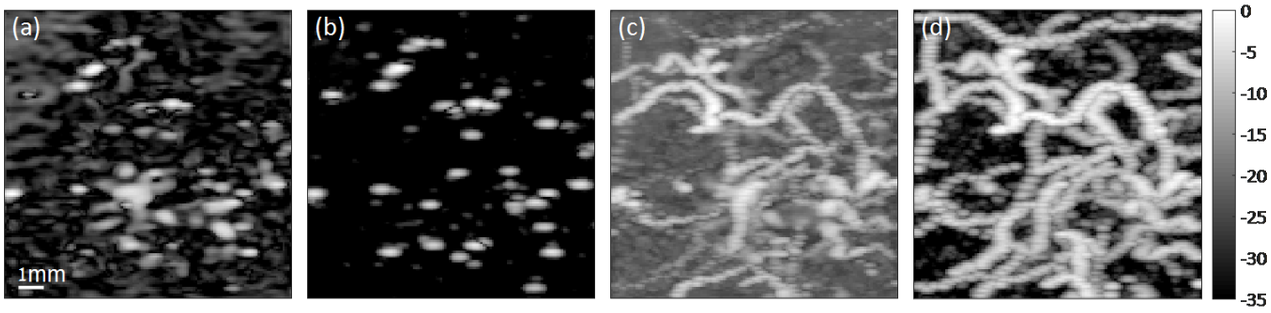


Fig. 3: Simulation results of ResNet. (a) A single recovered frame from a simulated movie, trained for 50 epochs on simulated data and for additional 20 epochs on experimental data. (b) A single recovered frame from a simulated movie, trained for 50 epochs on simulated data only. (c) MIP image for a recovered UCAs movie of 100 frames using the network in (a). (d) MIP image for a recovered UCAs movie of 100 frames using the network in (b). Color bar is in dB.

In practice each such patch is of size 32×32 pixels, over 20 consecutive frames and between consecutive patches there exists an overlap of 50%, which enables us to increase the number of training data. The regularization parameters of Algorithm 1, λ_1 and λ_2 are chosen empirically, but are chosen once for all the extracted patches.

A detailed description of the simulations is given in the supporting materials.

III. ResNet for clutter suppression

In recent years, several deep learning based architectures have been proposed and applied successfully to classification problems. One such popular network is the residual network, or ResNet [51]. ResNet utilizes convolution layers, along with batch normalization and skip connections. Skip connections allow the network to avoid the problem of vanishing gradients and reduce the overall number of network parameters.

We therefore decided to implemented ResNet using complex convolutions for the tissue clutter suppression task. The network does not recover the tissue signal, as the unfolded network, only the UCA signal. In Section V and in the supporting materials file, we compare both architectures and assess the advantages and disadvantages of each network. Network architecture and parameters are given in Section II of the supplementary materials.

IV. Experiments

The brains of two rats were scanned using a 256 system (Verasonics Inc., Kirkland, WA, USA). An L20-10 probe was utilized, with a central frequency of 15MHz. The rats underwent craniotomy after anesthesia to obtain an imaging window of $6 \times 2\text{mm}^2$. A bolus of 100μL SonoVue^{TM} (Bracco, Milan, Italy) contrast agent, diluted with normal saline with a ratio of 1:4, was administered intravenously to the rats tail vein. Plane-wave compounding of five steering angles (from $-12°$ to $12°$, with a step of $6°$) was adopted for ultrasound imaging. For each rat, over 6000 consecutive frames were acquired with a frame rate of 100Hz. 300 frames with relatively high B-mode intensity were manually selected for data processing in this work.

Both the complex ResNet and unfolded network were implemented it in Python 3.5.2, using the PyTorch 0.4.1 package. The unfolded network consists of 10 layers. Training was performed using the ADAM optimizer with a learning rate of 0.002. For the *in-vivo* experiments in Section V, we trained the network over 2400 simulated training pairs and additional 2400 *in-vivo* pairs taken only from the first rat. Training pairs were generated from the acquired US clips, after dividing each clip to $32 \times 32 \times 20$ patches. We then applied Algorithm 1 for each patch with $\lambda_1 = 0.02$ and $\lambda_2 = 0.001$ and $D_{\max} = 30000$ iterations to obtain the separated UCA signal

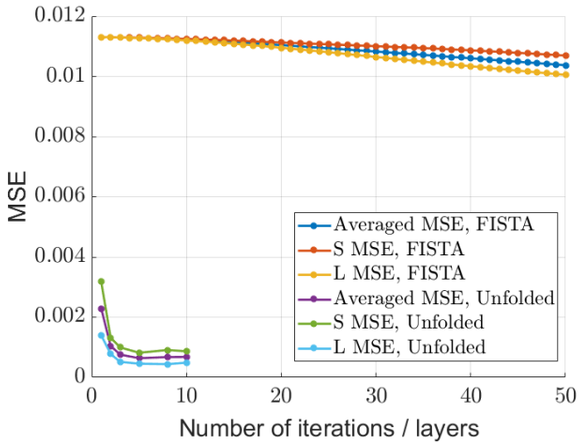


Fig. 4: MSE plot (Section II-C) for the FISTA algorithm and unfolded network as a function of the number of iterations/layers.

for the training process. Algorithm 1 was implemented in MATLAB (Mathworks Inc.) and was applied to the complex-valued IQ signal. PyTorch performs automatic differentiation and back-propagation using the Autograd functionality, and version 0.4.1 also support back-propagation through SVD, but only for real valued numbers. Thus, complex valued convolution layers and SVD operations were implemented, as described in Section III-B of the supplementary materials file.

V. Results

A. Simulation results

In this section we provide reconstruction results for both the unfolded and ResNet networks applied to simulated datasets. Figures 2 and 3 show reconstruction results from a simulated dataset for the unfolded network and ResNet, respectively. In both figures, panel (a) shows a single recovered frame when the networks are trained on simulated data for 50 epochs and additional 20 epochs on experimental data. Panel (b) shows reconstruction results when the networks are trained on simulated data only for 50 epochs. Panels (c) and (d) show maximum intensity projection (MIP) images from a simulated movie of 100 frames when the networks are trained on simulated data only and on both simulated and *in-vivo* data, respectively. In the case of recovering simulated data, it can be observed that when the networks are trained on both simulated and experimental data, their performance degrade, compared to their performance when trained on simulations only. This result is not surprising, since when the networks are trained on simulations only, their weights are optimized for simulated data only. When training on both types of data, the networks are required to generalize for both types, and are thus not optimized specifically only for simulated data. Section IV in the supporting materials provides additional simulation results, showing also the recovered tissue signal by the unfolded network.

As the unfolded network draws its architecture from the iterative ISTA algorithm, our second aim in this section is to assess the performance of both the unfolded network and the FISTA algorithm by calculating the MSE of each method as a function of iteration/layer number. Each layer in the unfolded network can be thought of as an iteration in the iterative algorithm. To that end, we next quantify the MSE over the simulated validation batch (sequence of 100 frames) as a function of layer number (unfolded network) and iteration number (FISTA), as presented in Fig. 4. For both methods, the MSE for the recovered sparse part (UCA signal) $\mathbf{S}$ and the low-rank part (tissue signal) $\mathbf{L}$ were calculated as a function of iteration/layer number, as well as the average MSE of both parts, according to (8) ($\alpha = 0.5$). For each layer number, we constructed an unfolded network with that number of layers, and trained it for 50 epochs on simulated data only.

Observing Fig. 4, it is clear that even when considering the unfolded network with only 1 layer, its performance in terms of MSE in an order of magnitude better than FISTA. Adding more layers improves the MSE, though after 5 layers, the performance remains roughly the same. However, it is also important to keep in mind that the FISTA algorithm is not designed to minimize (9), but rather (3).

B. In-vivo experiments

In Fig. 5, panel (a) depicts the resulting SVD based separation of the CEUS signal, while panel (b) shows the FISTA based separation and panel (c) shows the result of the unfolded network. The lower panels of Fig. 5 also compare the performance of the trained ResNet (panel (f)) on the *in-vivo* data as well as provide additional comparison to the commonly used wall filtering. Specifically, we use the 6th order Butterworth filter with two cutoff frequencies of 0.2π (panel (d)) and 0.9π (panel (e)) radians/samples. Two frequencies were chosen which represent two scenarios. The cutoff frequency of the recovery in panel (d) was chosen to suppress as much tissue signal, without rejecting slow moving UCAs. In panel (e), higher frequency was chosen, to suppress the slow moving tissue signal even further, but as can be seen, at a cost of removing also some of the slower bubbles. The result is a less consistent vascular image. Visually judging, all panels of Fig. 5 shows that ResNet outperforms both the SVD and wall filtering approaches. However, a more careful observation shows that the ResNet output, although more similar to the unfolded network's output, seems more grainy and less smooth than the unfolded network's image. The unfolded network's recovery exhibits the highest contrast, and produces the best visual result compared with all the other methods.

In each panel, the green and red boxes indicate selected areas, whose enlarged views are presented in the corresponding green and red boxes below each panel. Visual inspection of the entire panels (a)-(f) shows that both FISTA, ResNet and unfolded network achieve CEUS signal separation which is less noisy than the naive SVD approach and wall filtering. Considering the enlarged regions below the panels further supports this conclusion, showing better contrast of the FISTA and deep networks outputs. The enlarged panels below panels (d) and (e) show that indeed, as the cutoff frequency of the wall filter is increased, slow moving UCAs are also filtered

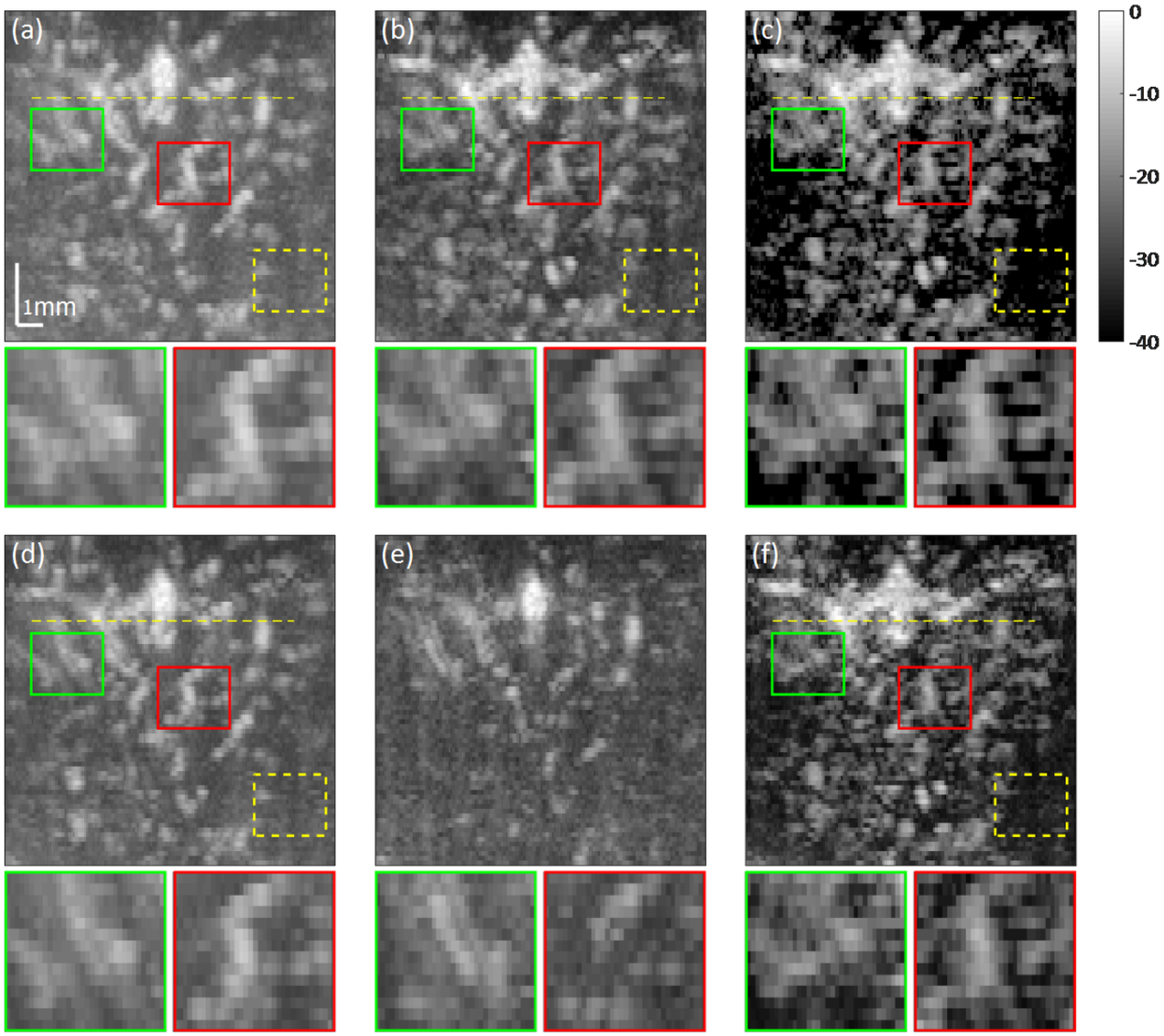


Fig. 5: Separation of *in-vivo* CEUS signal from tissue signal. (a) SVD based separation. (b) L+S FISTA separation. (c) Deep network separation, with the unfolded architecture of the FISTA algorithm. (d) Wall filtering with cutoff frequency of 0.2π (e) Wall filtering with cutoff frequency of 0.9π (f) ResNet. Color bar is in dB.

out. Both deep networks exhibit higher contrast than the other approaches.

TABLE I: CNR values for the selected green and red rectangles of Fig. 5, as compared with the dashed yellow background rectangle in each corresponding panel. All values are in dB.

	SVD	Wall filter	FISTA	ResNet	Unfolded
Green box	-1.65	-2.02	-1.67	-2.17	-0.3
Red box	-4.8	-5.55	-3.52	-2.95	-1.13

To further quantify the performance of each method, we provide two metrics to assess the contrast ratio of the output of each method. For each panel we calculate the contrast to

TABLE II: CR values for the selected green and red rectangles of Fig. 5, as compared with the dashed yellow background rectangle in each corresponding panel. All values are in dB.

	SVD	Wall filter	FISTA	ResNet	Unfolded
Green box	4.68	4.5	5.52	7.92	15.24
Red box	4.56	4.1	5.24	7.55	14.88

noise ratio (CNR), defined as

$$\text{CNR} = \frac{|\mu_s - \mu_b|}{\sqrt{\sigma_s^2 + \sigma_b^2}}.$$

CNR is calculated between a selected patch, e.g. the red or green boxes in panels (a)-(f) and a reference patch, marked by the dashed yellow patches, which represents the background,

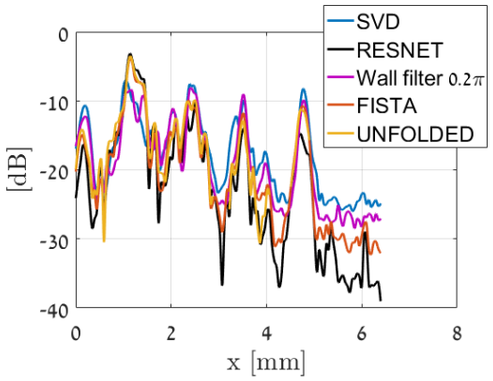


Fig. 6: Intensity profiles across the dashed yellow lines in panels (a)-(c) of Fig. 5. Regions in which the Unfolded networks' curve is missing indicate a value of $-\infty$. Values are in dB.

for the same image. That is, for each panel we estimate the CNR values of the red - yellow and green - yellow boxes, where μ_s is the mean of the red/green box with variance σ_s^2 and μ_b is the mean of the dashed yellow patch with variance σ_b^2. We also calculate the contrast ratio (CR), defined as

$$\text{CR} = \frac{\mu_s}{\mu_b},$$

and assess the performance of each method. Table II provides the estimated CR values of each method, respectively.

In both metrics, higher values imply higher contrast ratios, which suggest better noise suppression and better signal depiction. Considering both tables, the unfolded network outperforms all other approaches. In most cases, its performance is an order of magnitude better than the performance of the SVD filter. The CR values of ResNet are also better than the baseline SVD, though lower than those of the unfolded network. Its CNR values however, are not always higher than those of the SVD. In terms of CR, the FISTA results are better than those of the SVD filter, though lower than the deep-learning based approaches. In terms of CNR, for the green box, FISTA is comparable to SVD and better than ResNet, while for the red box, its performance is the worst.

Both metrics provide quantification to the performance of each method, supporting the previous conclusions, that by combining a proper model to the separation problem with a data-driven approach leads to a better separation of UCA and tissue signals, as well as noise reduction as opposed to the practiced SVD approach.

Lastly, we also provide intensity cross-sections, taken along the horizontal yellow dashed line for each method, as presented in Fig. 6. Considering the intensity cross-section of Fig. 6, its evident that all methods reconstruct all peaks with good correspondence. The FISTA and deep-learning networks' profiles exhibit higher contrast than the SVD and wall filter (deeper 'cavities'). In some areas, the unfolded (yellow) profiles seems to vanish. This is because the attained value is $-\infty$. The supporting materials file contains additional comparisons. Section V presents the training and validation losses of the networks, as well as the evolution of the regularization coefficients of the unfolded network as a function epoch number. Section VI discusses the importance of training the networks on both simulations and *in-vivo data*, while Section VII presents the training and execution times for both networks.

VI. Discussion and conclusions

In this work, the proposed deep-learning based frameworks clearly outperform the commonly practiced SVD filtering, wall filtering and the FISTA iterative algorithm. We attribute this improved performance for two main reasons. The first, is the fact that the networks are trained based on both *in-vivo* data and simulated data. The simulated data provides the networks an opportunity to learn from 'perfect' examples, without noise and with absolute separation of UCAs and their surroundings, which correspond to higher contrast images. The iterative algorithm, on the other hand, cannot learn or improve its performance on the *in-vivo* data from the simulated data. The second, is the fact that both networks rely on 2D complex convolutions. Contrary to FC layers, convolution layers reduce the number of learnable parameters considerably, thus help avoid over-fitting and achieve good performance even when the training sets are relatively low. Moreover, convolutions offer spatial invariance, which allows the network to capture spatially translated UCAs. As the input to the networks is a sequence of 2D complex frames (IQ signal), we implemented complex convolutions for both the unfolded network and ResNet.

Focusing on patch-based training over entire image training has several benefits. UCAs are used to image blood vessels, and as such entire images will include implicitly blood vessel structure. Thus, training over entire images may result in the network being biased towards the vessel trees presented in the (relatively small) training cohort. On the other hand, small patches are less likely to include meaningful structure, hence training on small patches will be less likely to bias the network towards specific blood vessel structures and enable the network to generalize better. Furthermore, as FISTA and the unfolded network employ SVD operations, processing the data in small batches improves execution time.

As was mentioned in the introduction, in the context of RPCA, a principled way to construct learnable pursuit architectures for structured sparse and robust low rank models was introduced in [43]. The proposed network was shown to faithfully approximate the RPCA solution with several orders of magnitude speed-up compared to its standard optimization algorithm counterpart. However, this approach is based on a non-convex formulation of the nuclear norm in which the rank (or an upper bound of it) is assumed to be known a priori.

The main idea in [43] is to majorize the non-differentiable nuclear norm with a differentiable term, such that the low-rank matrix is factorized as a product of two matrices, $\mathbf{L} = \mathbf{AB}$, where $\mathbf{A} \in \mathbb{R}^{n \times q}$ and $\mathbf{B} \in \mathbb{R}^{q \times m}$. Using this kind of factorization alleviates the need to compute the SVD product, but introduces another unknown parameter q which needs to be set (typically by hand), and corresponds to the rank of the low-rank matrix. This poses a network design limitation, as the rank can vary between different applications or even

different realizations of the same application, requiring the network to be re-trained per each new choice of q.

In fact, this is the same rank-thresholding parameter as in the standard SVD filtering technique, which we want to avoid hand-tuning in the first place. Moreover, this kind of factorization leads to a non-convex minimization problem, whose globally optimal stationary points depend on the choice of the regularization parameter λ_*. Since typically these parameters are chosen empirically, wrong choice of λ_* may lead to suboptimal reconstruction results of the RPCA problem, which are then used as training data for the fixed complexity learned algorithm. Overall, the algorithm will be trained for suboptimal performance. Since we operate on the original convex problem, we train against optimal reconstruction results of the RPCA algorithm, without the need to a-priori estimate the low-rank degree, q.

The proposed iterative and deep methods were demonstrated on the extraction of CEUS signal from an acquired IQ movie, but in principle can also be applied to dynamic MRI sequences, as well as to the separation of blood from tissue, e.g. for Doppler processing. In the latter case, the dynamic range between the tissue signal and the blood signal will be greater than that of the tissue and UCA signal. In terms of the iterative algorithm, this would lead to more iterations for the separation process, but once the iterative algorithm has finished, its learned version could be trained on its output to achieve faster execution.

Currently, the unfolded network and the ResNet offer a trade-off between them. By relying on convolutions, the unfolded network is trained with a considerable lower number of parameters (314 for 1 layer, 1796 for 10 layers) than the ResNet (25378). The unfolded network outperforms ResNet in both visual quality and quantifiable metrics, as presented in Section V. However, its training and execution times are slower (Section VII in the supporting materials file). This performance-runtime trade-off is attributed firstly to the fact that the unfolded network relies on SVD decomposition in each layer, which is a relatively computationally demanding operation. However, it allows the network to truly learn the rank of the low-rank matrix, without any need to upper bound it and restrict the architecture of the network. Incorporation of fast approximations for SVD computations can in principle be incorporated in the training and execution of the unfolded network, such as truncated or random SVD [52], which can potentially expedite the network's performance and achieve faster execution than ResNet. Its also important to keep in mind that the ResNet does not recover the tissue signal, only the UCA signal. In some applications, such as super-resolution CEUS imaging over long time durations, the tissue signal is used to correct for motion artifacts.

In this work, we proposed a low-rank plus sparse model for tissue/UCA signal separation, which exploits both spatio-temporal relations in the data, as well as the sparse nature of the UCA signal. This model leads to a solution in the form of an iterative algorithm, which outperforms the commonly practices SVD approach. We further suggested to improve both execution time and reconstructed image quality by unfolding the iterative algorithm into a deep network. The proposed architecture utilizes convolution layers instead of FC layers and a hybrid simulation-*in-vivo* training policy. Combined together, these techniques allow the unfolded network to achieve improved performance over its iterative counterpart, as well as over other popular architectures, such as ResNet. We demonstrated the performance of all methods on both simulated and *in-vivo* datasets, showing improved vascular depiction in a rat's brain.

References

[1] A. Fenster and J. C. Lacefield, *Ultrasound Imaging and Therapy*. Taylor & Francis, 2015.

[2] B. Furlow, "Contrast-enhanced ultrasound," *Radiologic technology*, vol. 80, no. 6, pp. 547S–561S, 2009.

[3] T. Opacic, S. Dencks, B. Theek, M. Piepenbrock, D. Ackermann, A. Rix, T. Lammers, E. Stickeler, S. Delorme, G. Schmitz *et al.*, "Motion model ultrasound localization microscopy for preclinical and clinical multiparametric tumor characterization," *Nature communications*, vol. 9, no. 1, p. 1527, 2018.

[4] N. De Jong, F. Ten Cate, C. Lancee, J. Roelandt, and N. Bom, "Principles and recent developments in ultrasound contrast agents," *Ultrasonics*, vol. 29, no. 4, pp. 324–330, 1991.

[5] N. Lassau, L. Chami, B. Benatsou, P. Peronneau, and A. Roche, "Dynamic contrast-enhanced ultrasonography (dce-us) with quantification of tumor perfusion: a new diagnostic tool to evaluate the early effects of antiangiogenic treatment," *European Radiology Supplements*, vol. 17, no. 6, pp. 89–98, 2007.

[6] J. M. Hudson, R. Williams, C. Tremblay-Darveau, P. S. Sheeran, L. Milot, G. A. Bjarnason, and P. N. Burns, "Dynamic contrast enhanced ultrasound for therapy monitoring," *European journal of radiology*, vol. 84, no. 9, pp. 1650–1657, 2015.

[7] C. Tremblay-Darveau, R. Williams, L. Milot, M. Bruce, and P. N. Burns, "Combined perfusion and doppler imaging using plane-wave nonlinear detection and microbubble contrast agents," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 61, no. 12, pp. 1988–2000, 2014.

[8] ——, "Visualizing the tumor microvasculature with a nonlinear plane-wave doppler imaging scheme based on amplitude modulation," *IEEE transactions on medical imaging*, vol. 35, no. 2, pp. 699–709, 2016.

[9] A. Bar-Zion, O. Solomon, C. Tremblay-Darveau, D. Adam, and Y. C. Eldar, "Sparsity-based ultrasound super-resolution hemodynamic imaging," *arXiv preprint arXiv:1712.00648*, 2017.

[10] R. J. van Sloun, O. Solomon, Y. C. Eldar, H. Wijkstra, and M. Mischi, "Sparsity-driven super-resolution in clinical contrast-enhanced ultrasound," in *Ultrasonics Symposium (IUS), 2017 IEEE International*. IEEE, 2017, pp. 1–4.

[11] R. J. van Sloun, O. Solomon, M. Bruce, Z. Z. Khaing, H. Wijkstra, Y. C. Eldar, and M. Mischi, "Super-resolution ultrasound localization microscopy through deep learning," *arXiv preprint arXiv:1804.07661*, 2018.

[12] O. Solomon, R. J. van Sloun, H. Wijkstra, M. Mischi, and Y. C. Eldar, "Exploiting flow dynamics for super-resolution in contrast-enhanced ultrasound," *arXiv preprint arXiv:1804.03134*, 2018.

[13] S. Bjaerum, H. Torp, and K. Kristoffersen, "Clutter filter design for ultrasound color flow imaging," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 49, no. 2, pp. 204–216, 2002.

[14] J. Willemetz, A. Nowicki, J. Meister, F. De Palma, and G. Pante, "Bias and variance in the estimate of the doppler frequency induced by a wall motion filter," *Ultrasonic Imaging*, vol. 11, no. 3, pp. 215–225, 1989.

[15] Y. M. Yoo, R. Managuli, and Y. Kim, "Adaptive clutter filtering for ultrasound color flow imaging," *Ultrasound in medicine & biology*, vol. 29, no. 9, pp. 1311–1320, 2003.

[16] L. Thomas and A. Hall, "An improved wall filter for flow imaging of low velocity flow," in *Ultrasonics Symposium, 1994. Proceedings., 1994 IEEE*, vol. 3. IEEE, 1994, pp. 1701–1704.

[17] P. J. Frinking, A. Bouakaz, J. Kirkhorn, F. J. Ten Cate, and N. De Jong, "Ultrasound contrast imaging: current and new potential methods," *Ultrasound in Medicine and Biology*, vol. 26, no. 6, pp. 965–975, 2000.

[18] L. A. Ledoux, P. J. Brands, and A. P. Hoeks, "Reduction of the clutter component in doppler ultrasound signals based on singular value decomposition: a simulation study," *Ultrasonic imaging*, vol. 19, no. 1, pp. 1–18, 1997.

[19] A. C. Yu and L. Lovstakken, "Eigen-based clutter filter design for ultrasound color flow imaging: a review," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 2010.

[20] F. W. Mauldin, F. Viola, and W. F. Walker, "Complex principal components for robust motion estimation," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 57, no. 11, 2010.

[21] F. W. Mauldin, D. Lin, and J. A. Hossack, "The singular value filter: a general filter design strategy for pca-based signal separation in medical ultrasound imaging," *IEEE transactions on medical imaging*, vol. 30, no. 11, pp. 1951–1964, 2011.

[22] C. M. Gallippi, K. R. Nightingale, and G. E. Trahey, "Bss-based filtering of physiological and arfi-induced tissue and blood motion," *Ultrasound in medicine & biology*, vol. 29, no. 11, pp. 1583–1592, 2003.

[23] J. S. Turek, M. Elad, and I. Yavneh, "Clutter mitigation in echocardiography using sparse signal separation," *Journal of Biomedical Imaging*, vol. 2015, p. 9, 2015.

[24] L. Lovstakken, S. Bjaerum, K. Kristoffersen, R. Haaverstad, and H. Torp, "Real-time adaptive clutter rejection filtering in color flow imaging using power method iterations," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 53, no. 9, pp. 1597–1608, 2006.

[25] D. E. Kruse and K. W. Ferrara, "A new high resolution color flow system using an eigendecomposition-based adaptive filter for clutter rejection," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 49, no. 10, pp. 1384–1399, 2002.

[26] C. Demené, T. Deffieux, M. Pernot, B.-F. Osmanski, V. Biran, J.-L. Gennisson, L.-A. Sieu, A. Bergel, S. Franqui, J.-M. Correas *et al.*, "Spatiotemporal clutter filtering of ultrafast ultrasound data highly increases doppler and fultrasound sensitivity," *IEEE transactions on medical imaging*, vol. 34, no. 11, pp. 2271–2285, 2015.

[27] C. Errico, J. Pierre, S. Pezet, Y. Desailly, Z. Lenkei, O. Couture, and M. Tanter, "Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging," *Nature*, vol. 527, no. 7579, pp. 499–502, 2015.

[28] P. Song, A. Manduca, J. D. Trzasko, and S. Chen, "Ultrasound small vessel imaging with block-wise adaptive local clutter filtering," *IEEE transactions on medical imaging*, vol. 36, no. 1, pp. 251–262, 2017.

[29] A. J. Chee and C. Alfred, "Receiver-operating characteristic analysis of eigen-based clutter filters for ultrasound color flow imaging," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 65, no. 3, pp. 390–399, 2018.

[30] A. Urban, C. Dussaux, G. Martel, C. Brunner, E. Mace, and G. Montaldo, "Real-time imaging of brain activity in freely moving rats using functional ultrasound," *nAture methods*, vol. 12, no. 9, p. 873, 2015.

[31] C. Errico, B.-F. Osmanski, S. Pezet, O. Couture, Z. Lenkei, and M. Tanter, "Transcranial functional ultrasound imaging of the brain using microbubble-enhanced ultrasensitive doppler," *NeuroImage*, vol. 124, pp. 752–761, 2016.

[32] G. H. Golub and C. F. Van Loan, "Matrix computations. 1996," *Johns Hopkins University, Press, Baltimore, MD, USA*, pp. 374–426, 1996.

[33] P. Song, J. D. Trzasko, A. Manduca, B. Qiang, R. Kadirvel, D. F. Kallmes, and S. Chen, "Accelerated singular value-based ultrasound blood flow clutter filtering with randomized singular value decomposition and randomized spatial downsampling," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 64, no. 4, pp. 706–716, 2017.

[34] N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM review*, vol. 53, no. 2, pp. 217–288, 2011.

[35] P.-G. Martinsson and S. Voronin, "A randomized blocked algorithm for efficiently computing rank-revealing factorizations of matrices," *SIAM Journal on Scientific Computing*, vol. 38, no. 5, pp. S485–S507, 2016.

[36] R. Otazo, E. Candès, and D. K. Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic mri with separation of background and dynamic components," *Magnetic Resonance in Medicine*, vol. 73, no. 3, pp. 1125–1136, 2015.

[37] M. Bayat and M. Fatemi, "Concurrent clutter and noise suppression via low rank plus sparse optimization for non-contrast ultrasound flow doppler processing in microvasculature," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 1080–1084.

[38] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, no. 3, p. 11, 2011.

[39] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[40] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*. Omnipress, 2010, pp. 399–406.

[41] H. Sreter and R. Giryes, "Learned convolutional sparse coding," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2191–2195.

[42] N. Samuel, T. Diskin, and A. Wiesel, "Deep mimo detection," *arXiv preprint arXiv:1706.01151*, 2017.

[43] P. Sprechmann, A. M. Bronstein, and G. Sapiro, "Learning efficient sparse and low rank models," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 9, pp. 1821–1833, 2015.

[44] A. Bar-Zion, C. Tremblay-Darveau, O. Solomon, D. Adam, and Y. C. Eldar, "Fast vascular ultrasound imaging with enhanced spatial resolution and background rejection," *IEEE transactions on medical imaging*, vol. 36, no. 1, pp. 169–180, 2017.

[45] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, p. 717, 2009.

[46] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge University Press, 2012.

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

[48] J.-J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bulletin de la Société mathématique de France*, vol. 93, pp. 273–299, 1965.

[49] Z. Tan, Y. C. Eldar, A. Beck, and A. Nehorai, "Analysis Sparse Recovery," *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1762–1774, 2014.

[50] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[51] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[52] Q. H. Yuanyuan, Wang and J. Luo, "Fast randomized singular value decomposition based clutter filtering for shear wave imaging," in *2018 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2018.