Deep Unfolded Robust PCA With Application to Clutter Suppression in Ultrasound

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

***Abstract*— Contrast enhanced ultrasound is a radiation-free imaging modality which uses encapsulated gas microbubbles for improved visualization of the vascular bed deep within the tissue. It has recently been used to enable imaging with unprecedented subwavelength spatial resolution by relying on super-resolution techniques. A typical preprocessing step in super-resolution ultrasound is to separate the microbubble signal from the cluttering tissue signal. This step has a crucial impact on the final image quality. Here, we propose a new approach to clutter removal based on robust principle component analysis (PCA) and deep learning. We begin by modeling the acquired contrast enhanced ultrasound signal as a combination of low rank and sparse components. This model is used in robust PCA and was previously suggested in the context of ultrasound Doppler processing and dynamic magnetic resonance imaging. We then illustrate that an iterative algorithm based on this model exhibits improved separation of microbubble signal from the tissue signal over commonly practiced methods. Next, we apply the concept of deep unfolding to suggest a deep network architecture tailored to our clutter filtering problem which exhibits improved convergence speed and accuracy with respect to its iterative counterpart. We compare the performance of the suggested deep network on both simulations and in-vivo rat brain scans, with a commonly practiced deep-network architecture and with the fast iterative shrinkage algorithm. We show that our architecture exhibits better image quality and contrast.**

***Index Terms*— Deep unfolding, inverse problems, machine learning, neural network, robust PCA, ultrasound imaging.**

Manuscript received July 4, 2019; revised September 1, 2019; accepted September 8, 2019. Date of publication September 13, 2019; date of current version April 1, 2020. This work was supported by the European Union's Horizon 2020 Research and Innovation Program under Grant 646804-ERC-COG-BNYQ. *(Oren Solomon and Regev Cohen contributed equally to this work.) (Corresponding author: Regev Cohen.)*

O. Solomon and R. Cohen are with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel (e-mail: orensol@campus.technion.ac.il; regev.cohen@campus.technion.ac.il).

Y. Zhang is with the Department of Electrical Engineering, Tsinghua University, Beijing 100084, China (e-mail: yizhang.ch.2015@gmail.com).

Y. Yang, Q. He, and J. Luo are with the Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China.

R. J. G. van Sloun is with the Department of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands (e-mail: r.j.g.v.sloun@tue.nl).

Y. C. Eldar is with the Faculty of Math and Computer Science, Weizmann Institute of Science, Rehovot 7610001, Israel (e-mail: yonina.eldar@weizmann.ac.il).

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2019.2941271

I. Introduction

MEDICAL ultrasound (US) is a radiation-free imaging modality used extensively for diagnosis in a wide range of 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 physical parameters such as relative blood volume (rBV), 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, allowing them to 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 by applying Doppler processing [7], [8] and for sub-wavelength vascular imaging [9]–[14].

A major challenge in ultrasonic vascular imaging such as CEUS is to suppress clutter signals stemming from stationary and slowly moving tissue as they introduce significant artifacts in blood flow imaging [15]. Over the past few decades several approaches have been suggested for clutter removal. The simplest method of tissue signal removal is to filter the ultrasonic signal along the temporal dimension using high-pass finite impulse response (FIR) or infinite impulse response (IIR) filters [16]. However, FIR filters need to have a high order whereas IIR filters exhibit a long settling time. This leads to a low number of temporal samples in each spatial location [17] when using focused transmission. The above methods rely on the assumption that tissue motion, if exists, is slow whereas 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 method for tissue suppression is second harmonic imaging [18], 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

0278-0062 © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Weizmann Institute of Science. Downloaded on April 04,2020 at 17:31:02 UTC from IEEE Xplore. Restrictions apply.

US scanner, and does not remove the tissue signal completely, as tissue can also exhibit a nonlinear response.

The above techniques are based only on temporal information and neglect the high spatial coherence of the tissue, compared to the blood. To exploit these spatial characteristics of tissue, a method of clutter removal was presented in [19], based on the singular value decomposition (SVD) of the correlation matrix of successive temporal samples. SVD filtering operates by stacking the (typically beamformed) acquired frames as vectors in a matrix whose column index indicates frame number. Then, an SVD of the matrix is performed and the largest singular values, which correspond to the highly correlated tissue, are zeroed out. Finally, a new matrix is composed based on the remaining singular values and reshaped to produce the blood/UCA movie.

Several SVD-based techniques have been proposed [15], [20]–[23] , such as down-mixing [15] for tissue motion estimation, adaptive clutter rejection for color flow proposed by Lovstakken *et al.* [24] and the principal component analysis (PCA) for blood velocity estimation presented in [25]. However, these methods are based on focused transmission schemes which limit the frame rate and the field of view. This in turn leads to a small number of temporal and spatial samples, reducing the effectiveness of SVD-based filtering. To overcome this limitation, SVD clutter removal was extended to ultrafast plane-wave imaging [13], [26]–[28], demonstrating substantially improved clutter rejection and microvascular extraction. This strategy has gained a lot of interest in recent years and nowadays it is used in numerous ultrafast US imaging applications such as functional ultrasound [29], [30], super-resolution ultrasound localization microscopy [13], [14] and high-sensitivity microvessel perfusion imaging [26], [27].

A major limitation of SVD-based filtering is the requirement to determine a threshold which discriminates between tissue related and blood related singular values. The appropriate setting of this threshold is based on empirical observations, rather than on minimizing an optimality criteria [31], and thus it is often unclear, especially when the eigenvalue spectra of the tissue and blood signals overlap. This threshold uncertainty motivates the use of a different model for the acquired data, one that can differentiate between tissue and contrast signals based on the spatio-temporal information, as well as additional information unique to the contrast signal: its 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 magnetic resonance imaging (MRI) [32] and recent US Doppler applications [33], 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 robust principle component analysis (RPCA) [34]. We then propose to solve a convex minimization problem to retrieve the UCA signal, which leads to an iterative principal component pursuit (PCP) [34]. Second, we utilize recent ideas from the field of deep learning [35] to dramatically improve the convergence rate and image reconstruction quality of the iterative algorithm. We do so by unfolding [36] the algorithm into a fixed-length deep network which we term Convolutional rObust pRincipal cOmpoNent Analysis (CORONA). This approach harnesses the power of both deep learning and model-based frameworks, and leads to improved performance in various fields [37]–[41].

CORONA is trained on sets of separated tissue/UCA signals from both *in-vivo* and simulated data. Similar to [38], 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 CORONA with the commonly practiced SVD approach, the iterative RPCA algorithm and an adaptation of the residual network (ResNet), which is considered to be one of the leading deep architectures for a wide variety of problems [42]. We show that CORONA outperforms all other approaches in terms of image quality and contrast.

Unfolding, or unrolling an iterative algorithm, was first suggested by Gregor and LeCun [36] to accelerate algorithm convergence. In the context of deep learning, an important question 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 [43] and more. The authors of [36] showed that by considering each iteration of an iterative algorithm as a layer in a deep network and then concatenating few such layers, it is possible to train such networks to achieve a dramatic improvement in convergence, i.e., to reduce the number of iterations significantly.

In the context of RPCA, a principled way to construct learnable pursuit architectures for structured sparse and robust low rank models was introduced in [37]. 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 on 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. In contrast, our approach does not require a-priori knowledge of the rank. Moreover, the use of convolutional layers exploits spatial invariance and facilitates our training process as it reduces the number of learnable parameters dramatically.

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 III describes the protocol of the experiments and technical details regarding the realizations of CORONA and ResNet. Section IV presents *in-silico* as well as *in-vivo* results of the iterative algorithm and the proposed

Authorized licensed use limited to: Weizmann Institute of Science. Downloaded on April 04,2020 at 17:31:02 UTC from IEEE Xplore. Restrictions apply.

deep networks. Finally, we discuss the results, limitations and further research directions in Section V.

Throughput the paper, x represents a scalar, $\mathbf{x}$ a vector, $\mathbf{X}$ a matrix, $\mathbf{I}_{N\times N}$ is the $N \times N$ identity matrix, $\|\cdot\|_p$ 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, $(\cdot)^H$ denotes the hermitian operator, 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 providing a low-rank plus sparse (L + S) model for the acquired US signal. In US imaging, typically a series of pulses are transmitted to the imaged medium. The resulting echoes from the 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 expressed as

$$D(x,z,t) = I(x,z,t) + i\,Q(x,z,t),$$

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 the 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 $N(x,z,t)$:

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

Acquiring a series of movie frames $t = 1,\ldots,T$, and stacking them as vectors in a matrix $\mathbf{D}$, leads to the following model

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

Here, we assume that the tissue matrix $\mathbf{L}$ can be described as a low-rank matrix, due to its high spatio-temporal coherence. The CEUS echoes matrix $\mathbf{S}$ is assumed to be sparse, as blood vessels typically sparsely populate the imaged medium. Assuming that each movie frame is of size $M \times M$ pixels, the matrices in (1) are of size $M^2 \times T$. Henceforth, we consider a more general model, in which the acquired matrix $\mathbf{D}$ is composed as

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

where $\mathbf{H}_1$ and $\mathbf{H}_2$ are the measurement matrices of appropriate dimensions, which relate the measurements to the unknown quantities we wish to recover. These matrices typically represent the physical acquisition mechanism and are dictated by the physics of the measurement device (in this work $\mathbf{H}_1 = \mathbf{H}_2 = \mathbf{I}$). The model (2) can have been applied to MRI, video compression and additional US applications, as we discuss in Section V. Our goal is to formulate a minimization problem to extract both $\mathbf{L}$ and $\mathbf{S}$ from $\mathbf{D}$ under the corresponding assumptions on the matrices.

B. RPCA Formulation

RPCA [32] aims at solving 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, the sum of the singular values of $\mathbf{L}$, and $\|\cdot\|_{1,2}$ is the mixed $l_{1,2}$ norm, which equals the sum of the l_2 norms of each row of $\mathbf{S}$. 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 be fixed, or change very slowly during the acquisition period. The nuclear norm is known to promote low-rank solutions, and is the convex relaxation of the rank minimization constraint [45].

By defining

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

and $\mathbf{A} = [\mathbf{H}_1, \mathbf{H}_2]$, (3) can be rewritten 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 the *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

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

Since $\text{prox}_h(\mathbf{X})$ is separable in $\mathbf{L}$ and $\mathbf{S}$, it holds that

$$\text{prox}_h(\mathbf{X}) = \begin{bmatrix}\text{prox}_{\rho_1}(\mathbf{L})\\ \text{prox}_{\rho_2}(\mathbf{S})\end{bmatrix} = \begin{bmatrix}\text{SVT}_{\lambda_1}(\mathbf{L})\\ \mathcal{T}_{\lambda_2}(\mathbf{S})\end{bmatrix}, \tag{6}$$

where the operators $\mathcal{T}_\alpha(\mathbf{x}) = \max(0, 1 - \alpha/\|\mathbf{x}\|_2)\mathbf{x}$ and

$$\text{SVT}_\alpha(\mathbf{X}) = \mathbf{U}\text{diag}(\max(0,\sigma_i - \alpha))\mathbf{V}^H, \quad i = 1,\ldots,r$$

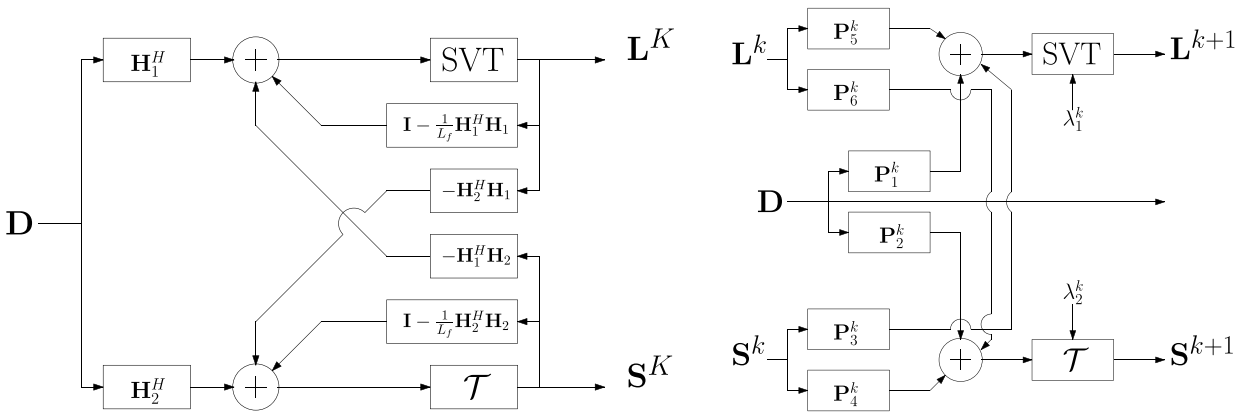
are the soft-thresholding and singular value thresholding [50] operators. Here $\mathbf{X}$ is assumed to have an SVD given by $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^H$ with $\Sigma = \text{diag}(\sigma_i,\ldots,\sigma_r)$, a diagonal matrix of the eigenvalues of $\mathbf{X}$. The proximal mapping (6) is applied in each iteration to the gradient of the quadratic part of (4), given by

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

and more specifically,

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

Authorized licensed use limited to: Weizmann Institute of Science. Downloaded on April 04,2020 at 17:31:02 UTC from IEEE Xplore. Restrictions apply.



(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 applied for K iterations (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 measurement matrix, and $\mathbf{S}_k$ and $\mathbf{L}_k$ are the estimated sparse and low-rank matrices in each iteration/layer, respectively.

The general iterative step of (L + S) ISTA applied to minimizing (3) is thus given by

$$\mathbf{X}^{k+1} = \text{prox}_h\left(\mathbf{X}^k - \frac{1}{L_f}g(\mathbf{X}^k)\right),$$

or

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

where L_f is the Lipschitz constant of the quadratic term of (4), given by the spectral norm of $\mathbf{A}^H\mathbf{A}$.

The L + S ISTA algorithm for minimizing (3) is summarized in Algorithm 1. 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 L + S 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 $k \le K_{\max}$ or stopping criteria not fulfilled **do**
    $\mathbf{G}_1^k = \left(\mathbf{I}-\frac{1}{L_f}\mathbf{H}_1^H\mathbf{H}_1\right)\mathbf{L}^k - \mathbf{H}_1^H\mathbf{H}_2\mathbf{S}^k + \mathbf{H}_1^H\mathbf{D}$
    $\mathbf{G}_2^k = \left(\mathbf{I}-\frac{1}{L_f}\mathbf{H}_2^H\mathbf{H}_2\right)\mathbf{S}^k - \mathbf{H}_2^H\mathbf{H}_1\mathbf{L}^k + \mathbf{H}_2^H\mathbf{D}$
    $\mathbf{L}^{k+1} = \text{SVT}_{\lambda_1/L_f}\left\{\mathbf{G}_1^k\right\}$
    $\mathbf{S}^{k+1} = \mathcal{T}_{\lambda_2/L_f}\left\{\mathbf{G}_2^k\right\}$
    $k \leftarrow k+1$
  end while
  return $\mathbf{S}^{K_{\max}}, \mathbf{L}^{K_{\max}}$

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. In the next section we propose CORONA which is based on unfolding Algorithm 1. Background on learning fast sparse approximations is given in Section I of the supplementary materials.

C. Unfolding the Iterative Algorithm

An iterative algorithm can be considered as a recurrent neural network, in which the kth iteration is regarded as the kth layer in a feedforward network [37]. To form a convolutional network, one may consider convolutional layers instead of matrix multiplications. With this philosophy, we form a network from (7) by replacing 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. These will be learned from training data. Contrary to previous works in unfolding RPCA which considered training fully connected (FC) layers [37], we employ convolution kernels in our implementation which allows us to achieve spatial invariance while reducing the number of learned parameters considerably. By doing so, the following equations for the kth layer are obtained

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

in which $*$ denotes a convolution operator. The latter can be considered as a single layer in a multi-layer feedforward network, which we refer to as CORONA: Convolutional rObust pRincipal cOmpoNent Analysis. A diagram of a single layer from the unfolded architecture is given in Fig. 1(b), where the supposedly known model matrices are 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.

In many applications, the recovered matrices $\mathbf{S}$ and $\mathbf{L}$ represent a 3D volume, or a 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. In practice, we treat our data as a 3D volume and apply 2D convolutions. The SVT operation (which has similar complexity as the SVD 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.

Authorized licensed use limited to: Weizmann Institute of Science. Downloaded on April 04,2020 at 17:31:02 UTC from IEEE Xplore. Restrictions apply.

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 $\sigma(x) = 1/(1 + \exp(-x))$ 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.

D. Training CORONA

CORONA 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 [51]. 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,

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

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

Training a deep network typically requires a large amount of training examples. In practice, US scans of specific organs are not available in abundance. To be able to train CORONA, we thus rely on two strategies: simulations and patch-based analysis. We created a simulator, described in Section II of the supplementary materials, to generate numerous examples of UCA and tissue signals, including ground truth, for training CORONA. Furthermore, instead of training the network over entire scans, we perform data augmentation by dividing each US movie used for training into 3D patches (axial coordinate, lateral coordinate and frame number). For *in-vivo* data, we then apply Algorithm 1 on each of these 3D patches. The SVD operations in Algorithm 1 become computationally tractable since we work on relatively small 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. In practice, the CEUS movie used for training is divided into 3D patches of size $32 \times 32 \times 20$ (32×32 pixels over 20 consecutive frames) with 50% overlap between neighboring patches. The regularization parameters of Algorithm 1, λ_1 and λ_2, are chosen manually by assessing the visual quality of the resulting recovery, but are chosen once for all the extracted patches.

III. Experiments

The brains of three rats were scanned using a Vantage 256 system (Verasonics Inc., Kirkland, WA, USA). An L20-10 probe was utilized, with a central frequency of 15 MHz. The rats underwent craniotomy after anesthesia to obtain an imaging window of 12.7×12.93 mm^2 (pixel size of 49.8×77 μm^2). A bolus of 100 μL SonoVue$^{\text{TM}}$ (Bracco, Milan, Italy) contrast agent, diluted with normal saline with a ratio of 1:4, was administered intravenously to the rat's 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 100 Hz. 300 frames with relatively high B-mode intensity were selected visually for data processing in this work. All procedures were approved by the ethics committee of Peking University (Beijing, China); approval number for the animal study is COE-DaiZF-1.

In Section II of the supplementary materials, we provide a detailed description of how the simulations of the UCA and tissue movies are generated. In particular, we detail how individual UCAs are modeled and propagated in the imaging plane, and describe the cluttering tissue signal model as well as tissue motion. We demonstrate the importance of training on both simulations and *in-vivo* data in Section III-A of the supplementary materials.

In recent years, several deep learning based architectures have been proposed and applied successfully to classification problems. One such approach is the residual network, or ResNet [42]. ResNet utilizes convolution layers, along with batch normalization and skip connections, which allow the network to avoid vanishing gradients and reduce the overall number of network parameters.

To compare with CORONA, we implemented ResNet using complex convolutions for the tissue clutter suppression task. The network does not recover the tissue signal, as CORONA, but only the UCA signal. In Section IV and in the supplementary materials, we compare both architectures and assess the advantages and disadvantages of each network, and show that CORONA outperforms ResNet in terms of image quality (contrast) of the CEUS signal.

Both ResNet and CORONA are implemented in Python 3.5.2, using the PyTorch 0.4.1 package. CORONA consists of 10 layers. The first three layers use convolution kernels of size $5 \times 5 \times 1$ with strides $(1, 1, 1)$, paddings $(2, 2, 0)$ and a bias, whereas the last seven layers use filters of size $3 \times 3 \times 1$ with strides $(1, 1, 1)$, paddings $(1, 1, 0)$ and a bias. Training was performed using the ADAM optimizer with a learning rate of 0.002. For the *in-vivo* experiments in Section IV, 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 $\mathbf{H}_1 = \mathbf{H}_2 = \mathbf{I}$, $\lambda_1 = 10^{-6}\sigma_{\max}$ and $\lambda_2 = 5 \cdot 10^{-8}\sigma_{\max}$ where $\sigma_{\max}$ is the maximum singular value of the data, and $D_{\max} = 2000$ iterations to obtain the separated UCA signal 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 supports back-propagation through SVD, but it is limited for real-valued numbers. Therefore, we implemented custom complex-valued convolution layers and SVD operations. All of these

Authorized licensed use limited to: Weizmann Institute of Science. Downloaded on April 04,2020 at 17:31:02 UTC from IEEE Xplore. Restrictions apply.

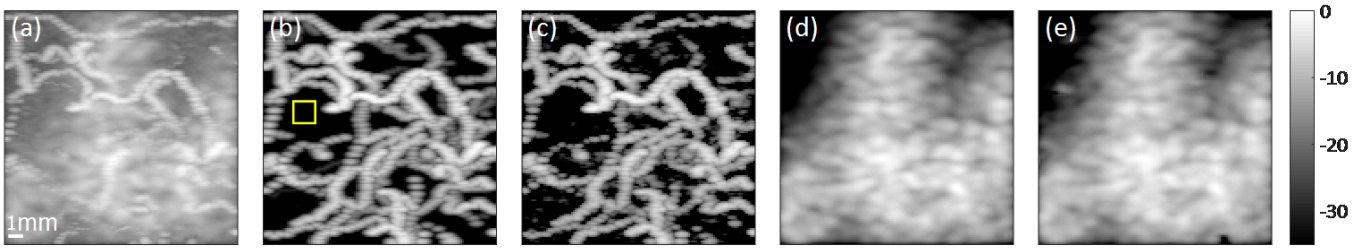


Fig. 2. Simulation results of CORONA. (a) MIP image of the input movie, composed from 50 frames of simulated UCAs cluttered by tissue. (b) Ground-truth UCA MIP image. (c) Recovered UCA MIP image via CORONA. (d) Ground-truth tissue MIP image. (e) Recovered tissue MIP image via CORONA. Color bar is in dB.

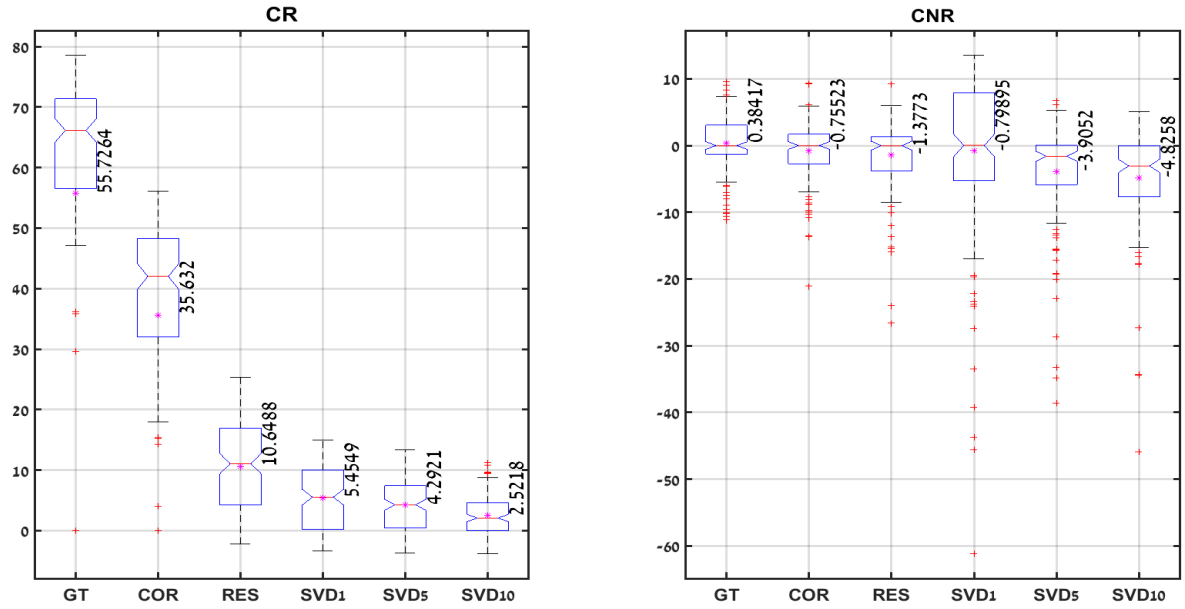


Fig. 3. CR and CNR comparison for the different recoveries of Fig. 2, and Fig. 7 in the supplementary materials. Red horizontal lines indicate the median, blue shapes indicate first (Q1) to third (Q3) quartiles and the black lines indicate the entire range of data samples, per category. Red markers indicate outliers per category, which were excluded from the statistical calculations. Purple stars represent averages over the entire ensemble for each method, with values written vertically. 'COR' stands for CORONA and 'RES' for ResNet. All values are in dB.

parameters are the same for both the simulation and *in-vivo* analysis, unless explicitly stated otherwise. *In-vivo* SVD recoveries were performed by associating the first 15 eigenvectors of the input movie to the tissue signal and the rest to the UCA signal.

IV. Results

A. Simulation Results

In this section we provide reconstruction results for CORONA applied to a simulated dataset, and trained on simulations. Fig. 2 presents reconstruction results of the UCA signal $\mathbf{S}$ and the low-rank tissue $\mathbf{L}$ against the ground truth images. Panel (a) shows a representative image in the form of maximum intensity projection (MIP)[1] of the input cluttered movie (50 frames). It is evident that the UCA signal, depicted as randomly twisting lines, is masked considerably by the simulated tissue signal. Panel (b) illustrates the ground truth MIP image of the UCA signal, whereas panel (c) presents the MIP image of the recovered UCA signal via CORONA. Panels (d) and (e) show MIP images of the ground truth and CORONA recovery, respectively.

Observing all panels, it is clear that CORONA is able to recover reliably both the UCA signal and the tissue signal. Additionally, since in the simulation we have the ground truth data, we can objectively compare the performance of CORONA, ResNet and the different SVD reconstructions. Fig. 3 presents a box-plot comparison of measured contrast ratio (CR) and contrast to noise ratio (CNR) values for each of the different reconstructions as well as for the ground truth (GT). CNR is calculated between a selected patch which represents the signal (UCAs) and a reference patch which represents the background, taken from the same image. We denote by μ_s the mean of the selected signal patch with variance σ_s^2 and by μ_b the mean of the background patch with variance σ_b^2. The CNR and CR are defined as

$$\mathrm{CNR} = \frac{|\mu_s - \mu_b|}{\sqrt{\sigma_s^2 + \sigma_b^2}}, \quad \mathrm{CR} = \frac{\mu_s}{\mu_b}.$$

For each method (and ground truth), CR and CNR values were calculated by considering the 12×12 pixels yellow

[1]In order to present a single representative image, we take the pixel-wise maximum from each movie. This process is also referred to as maximum intensity projection, and is a common method to visualize CEUS images.

Authorized licensed use limited to: Weizmann Institute of Science. Downloaded on April 04,2020 at 17:31:02 UTC from IEEE Xplore. Restrictions apply.

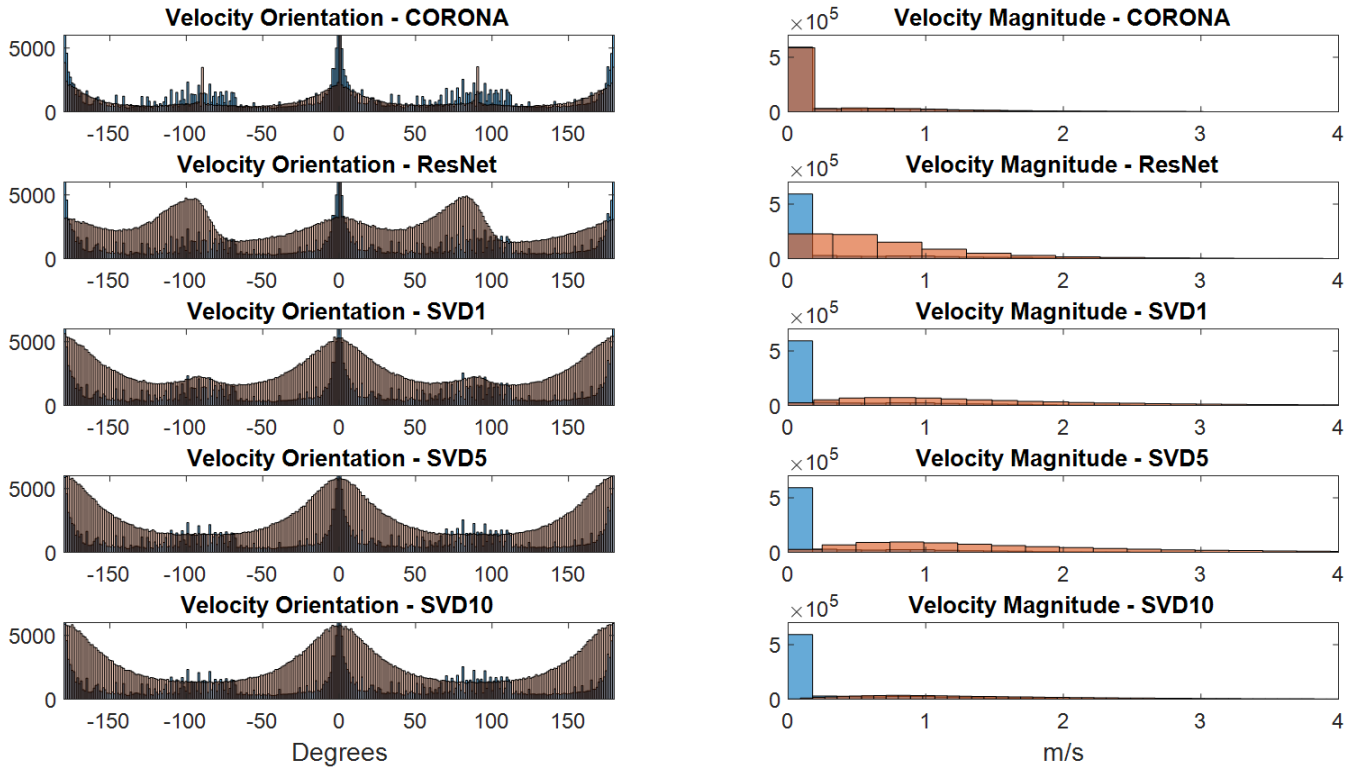


Fig. 4. Panels on the left column show velocity orientation histogram for each method (orange) against the ground truth velocity orientation histogram (blue). Panels on the right column present velocity magnitude histogram for each method (blue) compared with the ground truth velocity magnitude histogram (orange). SVD1, 5, 10 implies that the tissue/UCAs cutoff eigenvalue was chosen as 1, 5, and 10, respectively.

patch depicted in panel (b) of Fig. 2 as the background patch (for each reconstruction method, the patch was taken from the reconstruction itself). Then, we divided each MIP image (ground truth, CORONA, ResNet and the three SVD recoveries) into 144 12×12 non-overlapping patches and computed the corresponding CR and CNR values per each technique and ground truth with its corresponding background patch, as presented in Fig. 3. Corresponding SVD recoveries are presented in Fig. 7 in the supplementary materials, showing SVD recoveries of different tissue/UCA threshold values.

Examining Fig. 3, clearly the CR values of CORONA are the closest to the ground truth values by a large margin, indicating higher contrast than those of the SVD recoveries and ResNet (which in turn exhibits higher CR values than the SVD recoveries). The CNR values and distribution of CORONA are very similar to those of the ground truth, as well as those of ResNet, unlike the CNR values of the SVD recoveries, which also show lower median values. The average CNR value of CORONA is higher than the value of ResNet by a factor of almost two, and is the closest to the ground truth CNR. This indicates that CORONA recovery is more similar to the ground truth UCA MIP image than all of the SVD reconstructions. Although the CNR distribution of ResNet (corresponding MIP image is given in Fig. 9 of the supplementary materials, panel (c)) is relatively similar to that of CORONA and the ground truth, the distribution of CORONA still exhibits closer similarity to the ground truth than ResNet.

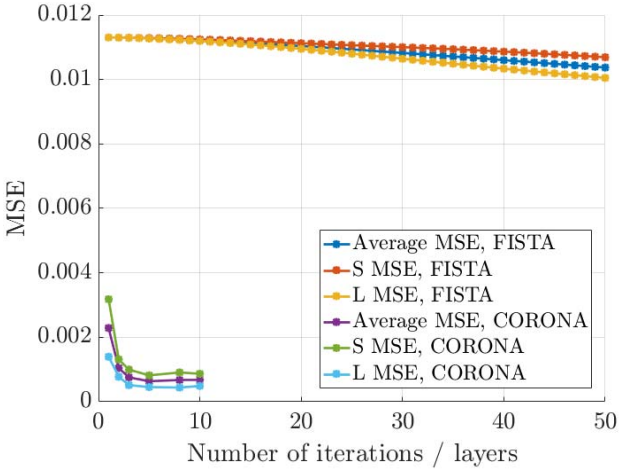


Fig. 5. MSE plot for the FISTA algorithm and CORONA as a function of the number of iterations/layers.

Clutter suppression in US is often used as a pre-processing step for further analysis, such as the extraction of blood flow parameters (e.g. velocities and relative blood volume [3]) as well as advanced processing such as super-resolution [12]. If the tissue signal is suppressed adequately, quantitative parameters such as blood velocity distribution can be estimated more precisely. On the other hand, if substantial tissue signal is still present in the extracted UCA signal, then it is very likely that extraction of quantitative parameters will be erroneous.

Authorized licensed use limited to: Weizmann Institute of Science. Downloaded on April 04,2020 at 17:31:02 UTC from IEEE Xplore. Restrictions apply.

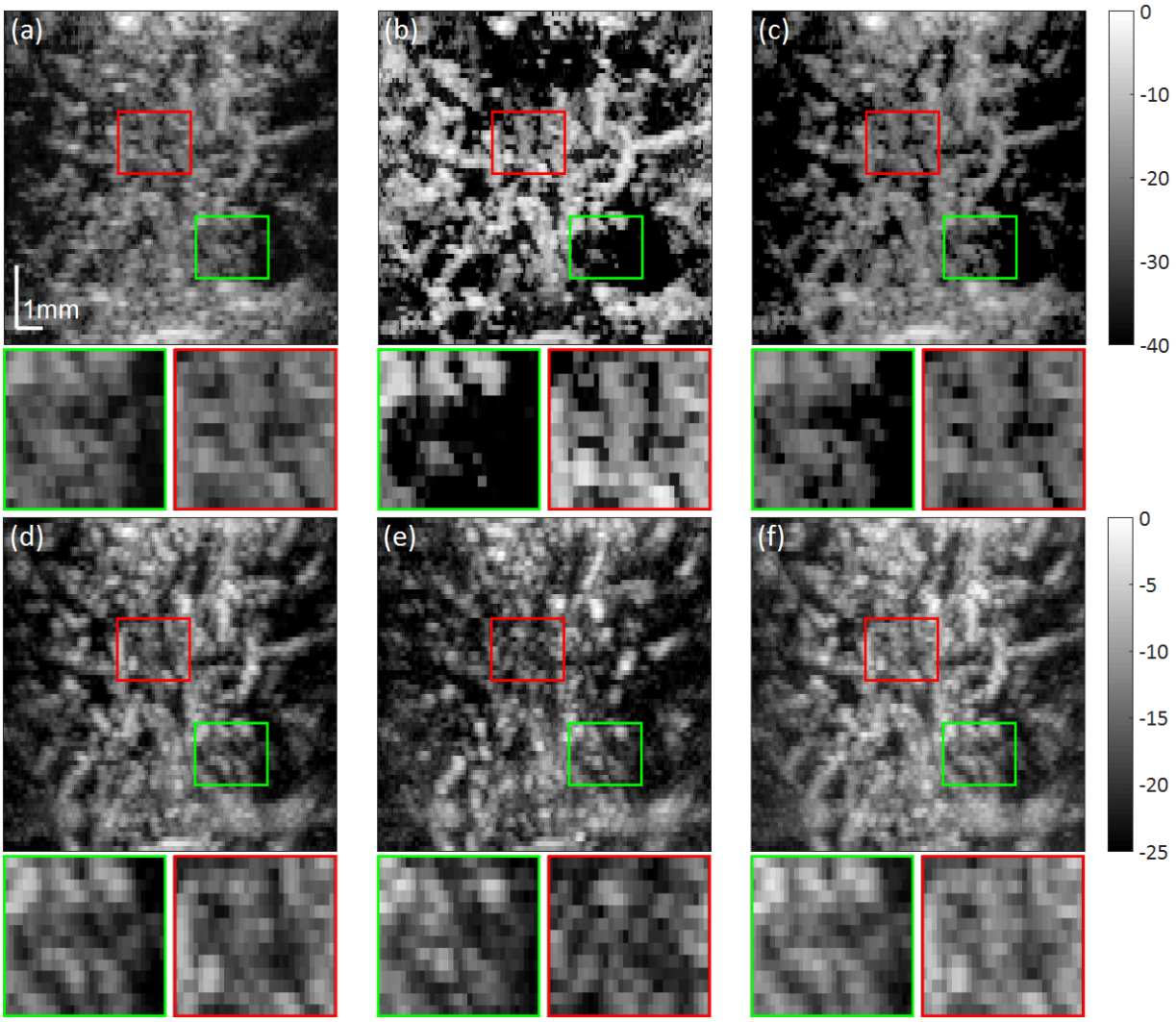


Fig. 6. Recovery of *in-vivo* CEUS signal depicting rat brain vasculature. (a) ResNet image. (b) L + S FISTA separation. (c) CORONA. (d) Wall filtering with cutoff frequency of 0.2π. (e) Wall filtering with cutoff frequency of 0.9π. (f) SVD based separation. Color bars are in dB, upper bar refers to panels (a), (b) and (c), whereas lower bar refers to panels (d), (e) and (f).

Thus, a good measure of the performance of any clutter suppression algorithm is how well quantitative parameters can be extracted from the recovered UCA signal.

To this end, in Fig. 4 we show a comparison of estimated blood flow velocities (magnitude and orientation histograms) for the different methods on the simulated dataset, compared with the ground truth velocity histograms. Flow velocities were estimated per each image in the recovered UCA sequence based on optical flow estimation using the Lukas and Kanade method (with threshold parameter 0.5) [52], as in [12], [53].

Considering the upper row of Fig. 4, it is evident that the velocity histograms calculated from the CORONA reconstruction show very good correspondence to the ground truth histogram, in both orientation and magnitude. On the other hand, all of the SVD and ResNet reconstructions show poor compatibility to the ground truth histograms, both in magnitude and orientation. This histogram mismatch implies poor velocity estimation from these reconstructions, as the result of residual tissue signal present in the UCA signal or removal of some of the UCA signal (as happened in some of the SVD reconstructions). Only CORONA obtained histograms which closely resemble those of the ground truth.

Finally, since CORONA draws its architecture from the iterative ISTA algorithm, our second aim in this section is to assess the performance of both CORONA and the FISTA algorithm by calculating the MSE of each method as a function of iteration/layer number. Each layer in CORONA can be thought of as an iteration in the iterative algorithm. To that end, in Fig. 5 we quantify the MSE over the simulated validation batch (sequence of 100 frames) as a function of layer number (CORONA) and iteration number (FISTA). For both methods, the MSEs 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) with $\alpha = 0.5$. For each layer number, we constructed an unfolded network with that number of layers, and trained it for 50 epochs on simulated data only.

Authorized licensed use limited to: Weizmann Institute of Science. Downloaded on April 04,2020 at 17:31:02 UTC from IEEE Xplore. Restrictions apply.

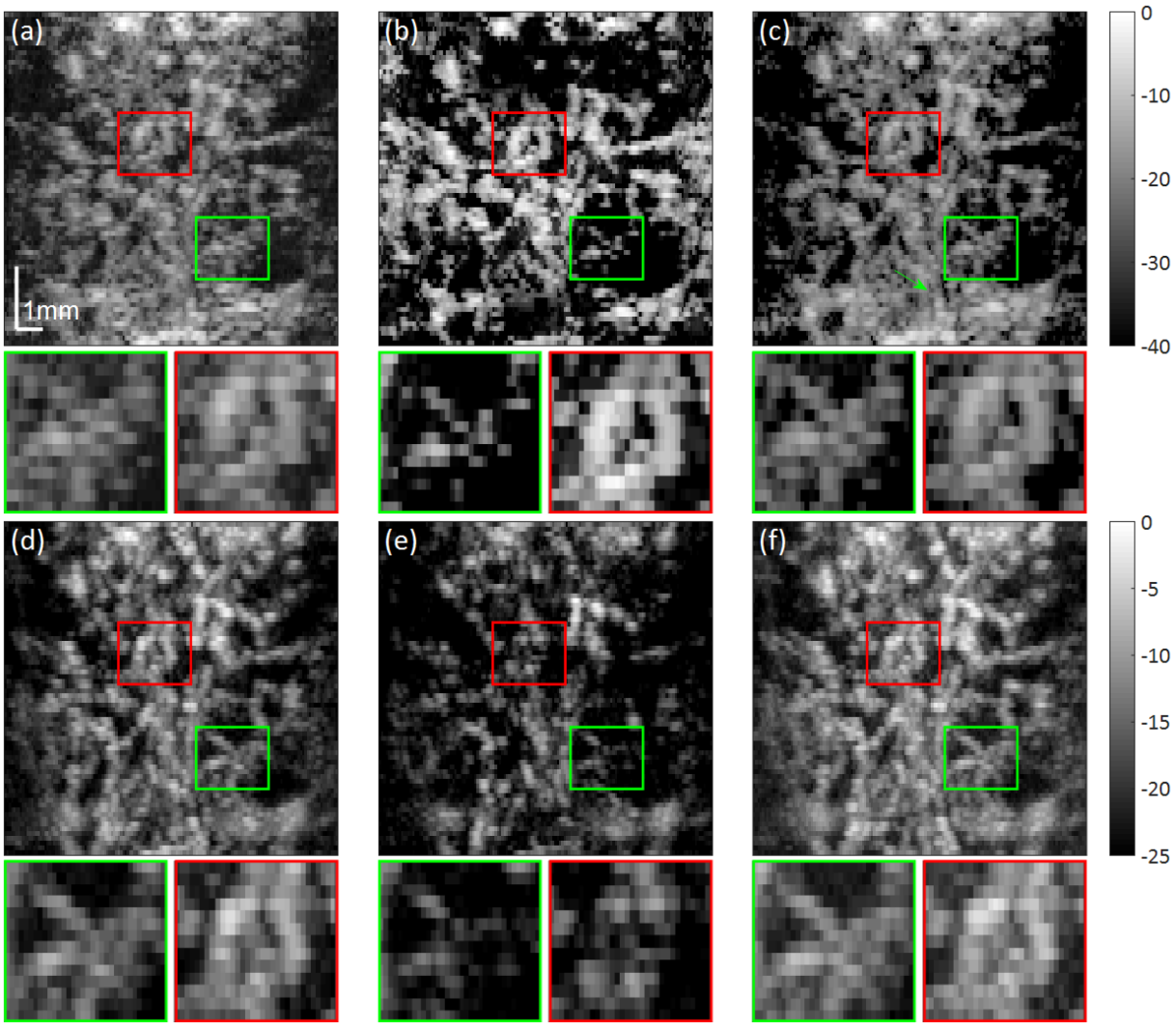


Fig. 7. Recovery of *in-vivo* CEUS signal depicting rat brain vasculature. (a) ResNet image. (b) L + S FISTA separation. (c) CORONA. (d) Wall filtering with cutoff frequency of 0.2π. (e) Wall filtering with cutoff frequency of 0.9π (f) SVD based separation. Color bars are in dB, upper bar refers to panels (a), (b) and (c), whereas lower bar refers to panels (d), (e) and (f).

Observing Fig. 5, it is clear that even when considering CORONA with only 1 layer, its performance in terms of MSE is an order of magnitude better than FISTA applied with 50 iterations. Adding more layers improves the CORONA MSE, though after 5 layers, the performance remains roughly the same. Fig. 5 also shows that a clear decreasing trend is present for the FISTA MSE, however a dramatic increase in the number of iterations is required by FISTA to achieve the same MSE values.

Additional simulation-based quantification and quantitative comparisons are given in Section II-B of the supplementary materials. Section II-C of the supplementary materials provides additional simulation results, showing the recovered UCA signal by ResNet. Although qualitatively ResNet recovers the UCA signal well, its contrast is lower than the contrast of the CORONA recovery, which presents a clearer depiction of the random vascular structure of the simulation. Moreover, ResNet does not recover the tissue signal, whereas CORONA does.

Overall, in simulations CORONA depicts more accurate separation of UCAs from the tissue signal. This is quantified by better CR and CNR values, as well as more accurate UCA velocity estimation. In particular, CORONA can recover slow moving UCAs, with velocities in the order of the tissue motion, whereas in this case, SVD fails.

B. In-Vivo Experiments

We now proceed to demonstrate the performance of CORONA on *in-vivo* scans of the second rat from 200 consecutive frames (reconstruction results for the third rat are given in Section III-D of the supplementary materials). As was described in Section III, CORONA was trained on both simulated and experimental data. In Fig. 6, panel (a) depicts the trained ResNet recovery, panel (b) shows the FISTA based separation and panel (c) shows the result of CORONA. The lower panels of Fig. 6 also compare the SVD based separation of the CEUS signal (panel (f)) on the *in-vivo* data as well

Authorized licensed use limited to: Weizmann Institute of Science. Downloaded on April 04,2020 at 17:31:02 UTC from IEEE Xplore. Restrictions apply.

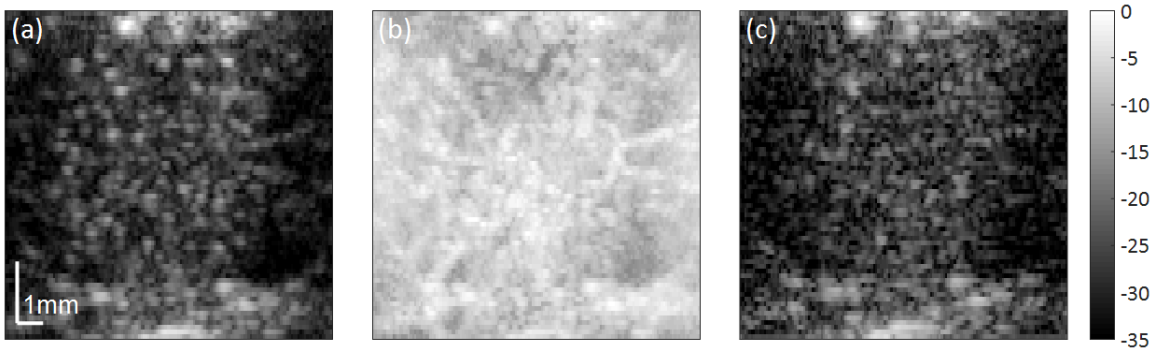


Fig. 8. Recovery of *in-vivo* tissue signal of rat brain scans. (a) SVD based separation. (b) L + S FISTA separation. (c) CORONA. Color bar is in dB.

as provide additional comparison to the commonly used wall filtering. Specifically, we use a 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 as possible, without rejecting slow moving UCAs. In panel (e), a 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. Due to dynamic range differences between the methods, and in order to fully present the reconstructed vasculature of each image in the best way possible, the FISTA, CORONA and ResNet images are displayed with a dynamic range of 40 dB, whereas the SVD and wall-filter images are displayed with a dynamic range of 25 dB.

Visually judging the panels of Fig. 6, it is evident that CORONA achieves a clearer depiction of the vasculature than ResNet. The CORONA image seems less noisy and presents a sharper depiction of the vessels' contours, as opposed to the blurier ResNet image. Although the FISTA image recovers the blood vessels, some of the vessels seem to be absent, especially in the upper portion of the image. This result is most likely due to the difficulty in tuning the regularization parameters by hand.

The lower panels show recoveries which qualitatively seem noisier than the CORONA image. Moreover, considering the wall-filter recovery in panel (e), it seems that the chosen cutoff frequency was set too high, removing slow flowing UCAs. The SVD and wall-filter image of panel (d) present a smooth depiction of the blood vessels, like CORONA.

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. In this example, the red box below the CORONA image depicts a clearer separation of the blood vessels than the red box below the SVD and wall-filter images. The green box below the FISTA image shows an incomplete image compared with the other recoveries. It also seems that both deep networks exhibit higher contrast than the other approaches.

Fig. 7 corresponds to vascular reconstruction from additional 100 frames acquired for the second rat. (similar dynamic range, and panels arrangement as in Fig. 6).

At first glance, similar conclusions can be drawn by examining the panels of Fig. 7. The CORONA image seems to exhibit the highest contrast, depicting sharper contours of the blood vessels than ResNet, whereas the FISTA image seems to miss a portion of the blood vessels (most likely the result of the difficult manual tuning of the algorithm's regularization parameters). Choosing a high cutoff frequency for the wall-filter (panel (e)) seems to remove many UCAs, resulting in a very discontinuous image. On the other hand the wall-filter image of panel (d) and the SVD recovery clearly depict the vessels, albeit with lower contrast than the CORONA image.

However, closely examining the enlarged regions in the green and red boxes reveals that a clear preference between the methods is not always possible. For example, the "spider" like shape in the green box is broken in the FISTA and wall-filter images, whereas it seems sharp and clear in the SVD image and in the wall-filter image of panel (d). In comparison to the latter, the same area seems less clear in the CORONA image, and it is very blurred in the ResNet recovery. In contrast, the red panel below the CORONA image clearly depicts a bifurcation of two vessels. This can also be observed in the SVD and wall-filter (panel (d)) images, but it seems blurrier with a less distinct depiction between the contour of the vessels. The FISTA recovery also seems sharp and clear, whereas the ResNet image seems blurred. The wall-filter image of panel (e) presents a broken and unclear image of the vessels.

Finally, Fig. 8 shows the tissue reconstruction results (over all of the 300 frames) obtained by SVD (panel (a)), FISTA (panel (b)) and CORONA (panel (c)). The FISTA recovery seems to contain some of the UCA signal, as a result of the difficulty to manually fine-tune its regularization parameters. Fig. 8 exemplifies the ability of CORONA to not only produce reliable UCA reconstructions, but also tissue imaging, which can be used as additional input to modern applications, such as tissue-based motion compensation.

V. Discussion and Conclusions

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 practiced SVD approach. We further suggested to improve

Authorized licensed use limited to: Weizmann Institute of Science. Downloaded on April 04,2020 at 17:31:02 UTC from IEEE Xplore. Restrictions apply.

both execution time and reconstructed image quality by unfolding the iterative algorithm into a deep network, referred to as CORONA. The proposed architecture utilizes convolution layers instead of FC layers and a hybrid simulation-*in-vivo* training policy. Combined, these techniques allow CORONA to achieve improved performance over its iterative counterpart, as well as over other popular deep learning 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.

We conclude by discussing several points, regarding the performance and design of deep-learning based networks. We attribute the improved performance over the commonly practiced SVD filtering, wall filtering and FISTA to two main reasons. The first, is the fact that for application on *in-vivo* data, the networks are trained based on both *in-vivo* data and simulated data. The simulated data provides the networks with an opportunity to learn from "perfect" examples, without noise and with absolute separation of UCAs and their surroundings. In Section III-A of the supplementary materials we show the effect on recovery when the network is trained with and without experimental data. 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 small. Moreover, convolutions offer spatial invariance, which allows the network to capture spatially translated UCAs.

That said, it is important to note that a clear distinction between the performance of the different methods is subjective when considering *in-vivo* data. This is due to the lack of ground-truth information. As discussed in the previous section, CORONA does exhibit higher contrast than the other methods and a smooth depiction of the vessels in most cases. However, we also presented an example in which a better depiction of vessels was attained by the SVD technique. We attribute this to the fact that although CORONA is parameter-free, it is still trained in part by patches obtained via FISTA. The parameters of FISTA are currently hand-tuned, which does not guarantee the best possible recovery (as clearly demonstrated in the *in-vivo* examples). A better selection of the FISTA regularization parameters (e.g. by using Stein's unbiased risk estimator [54], [55]) might improve the performance of CORONA. Furthermore, SVD filtering requires tuning the parameters for each specific image in an online fashion, whereas the tuning for CORONA's training set is done offline and for various images and datasets. The in-depth simulation analysis presented in the paper validates the improvement in contrast obtained by CORONA, as well as its benefit in separating between the tissue and slow moving UCAs. In this case, SVD might fail as the velocity spectrum of the UCAs and tissue might overlap. In contrast, CORONA also exploits the sparse nature of the UCAs, thus achieves separation even for slow moving UCAs.

Focusing on patch-based training (Section II-D) 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 is less likely to bias the network towards specific blood vessel structures and enables the network to generalize better. Furthermore, as FISTA and CORONA employ SVD operations, processing the data in small batches improves execution time [27], [56].

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 [37]. 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 [37] 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 retrained 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. 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_1. Since typically these parameters are chosen empirically, a wrong choice of λ_1 may lead to suboptimal reconstruction results of the RPCA problem, which are then used as training data for the fixed complexity learned algorithm. 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.

Currently CORONA and ResNet offer a trade-off between them. By relying on convolutions, CORONA is trained with a considerable lower number of parameters (314 for 1 layer, 1796 for 10 layers) than ResNet (25378). CORONA outperforms ResNet in both visual quality and quantifiable metrics, as presented in Section IV. However, its training and execution times are slower (see Section III-E of the supplementary materials). This performance-runtime trade-off is attributed to the fact that CORONA relies on SVD decomposition in each layer, which is a relatively computationally demanding operation. However, it allows the network to learn the rank of the low-rank matrix, without the need to upper bound it and restrict the architecture of the network. Incorporation of fast approximations for SVD computations, such as truncated or random SVD [56]–[59], can potentially expedite the network's

Authorized licensed use limited to: Weizmann Institute of Science. Downloaded on April 04,2020 at 17:31:02 UTC from IEEE Xplore. Restrictions apply.

performance and achieve faster execution than ResNet. It is also important to keep in mind that ResNet does not recover the tissue signal, but 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.

On a final note, 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.

Acknowledgment

The authors would like to thank Mr. De Ma and Dr. Zhifei Dai from the Department of Biomedical Engineering at Peking University for their help in performing the *in-vivo* experiments.

References

[1] A. Fenster and J. C. Lacefield, *Ultrasound Imaging and Therapy*. New York, NY, USA: Taylor & Francis, 2015.

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

[3] T. Opacic *et al.*, "Motion model ultrasound localization microscopy for preclinical and clinical multiparametric tumor characterization," *Nature Commun.*, vol. 9, no. 1, p. 1527, 2018.

[4] N. de Jong, F. T. Cate, C. Lancée, J. R. T. C. 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," *Eur. Radiol. Suppl.*, vol. 17, no. 6, pp. 89–98, 2007.

[6] J. M. Hudson *et al.*, "Dynamic contrast enhanced ultrasound for therapy monitoring," *Eur. J. Radiol.*, vol. 84, no. 9, pp. 1650–1657, Sep. 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 Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 61, no. 12, pp. 1988–2000, Dec. 2014.

[8] C. Tremblay-Darveau, R. Williams, L. Milot, M. Bruce, and P. N. Burns, "Visualizing the tumor microvasculature with a nonlinear plane-wave Doppler imaging scheme based on amplitude modulation," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 699–709, Feb. 2016.

[9] A. Bar-Zion, O. Solomon, C. Tremblay-Darveau, D. Adam, and Y. C. Eldar, "SUSHI: Sparsity-based ultrasound super-resolution hemodynamic imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 65, no. 12, pp. 2365–2380, Dec. 2018.

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

[11] R. J. G. van Sloun *et al.*, "Super-resolution ultrasound localization microscopy through deep learning," 2018, *arXiv:1804.07661*. [Online]. Available: https://arxiv.org/abs/1804.07661

[12] O. Solomon, R. J. G. van Sloun, H. Wijkstra, M. Mischi, and Y. C. Eldar, "Exploiting flow dynamics for super-resolution in contrast-enhanced ultrasound," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, to be published.

[13] C. Errico *et al.*, "Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging," *Nature*, vol. 527, no. 7579, pp. 499–502, Nov. 2015.

[14] K. Christensen-Jeffries, R. J. Browning, M. X. Tang, C. Dunsby, and R. J. Eckersley, "*In vivo* acoustic super-resolution and super-resolved velocity mapping using microbubbles," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 433–440, Feb. 2015.

[15] S. Bjærum, H. Torp, and K. Kristoffersen, "Clutter filter design for ultrasound color flow imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 49, no. 2, pp. 204–216, Feb. 2002.

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

[17] Y. M. Yoo, R. Managuli, and Y. Kim, "Adaptive clutter filtering for ultrasound color flow imaging," *Ultrasound Med. Biol.*, vol. 29, no. 9, pp. 1311–1320, Sep. 2003.

[18] P. J. A. Frinking, A. Bouakaz, J. Kirkhorn, F. J. Ten Cate, and N. de Jong, "Ultrasound contrast imaging: Current and new potential methods," *Ultrasound Med. Biol.*, vol. 26, no. 6, pp. 965–975, Jul. 2000.

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

[20] A. C. H. Yu and L. Lovstakken, "Eigen-based clutter filter design for ultrasound color flow imaging: A review," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 57, no. 5, pp. 1096–1111, May 2010.

[21] F. W. Mauldin, F. Viola, and W. F. Walker, "Complex principal components for robust motion estimation," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 57, no. 11, pp. 2437–2449, Nov. 2010.

[22] F. W. Mauldin, Jr., 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 Trans. Med. Imag.*, vol. 30, no. 11, pp. 1951–1964, Nov. 2011.

[23] C. M. Gallippi, K. R. Nightingale, and G. E. Trahey, "BSS-based filtering of physiological and ARFI-induced tissue and blood motion," *Ultrasound Med. Biol.*, vol. 29, no. 11, pp. 1583–1592, 2003.

[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 Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 53, no. 9, pp. 1597–1608, Sep. 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 Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 49, no. 10, pp. 1384–1399, Oct. 2002.

[26] C. Demené *et al.*, "Spatiotemporal clutter filtering of ultrafast ultrasound data highly increases Doppler and fultrasound sensitivity," *IEEE Trans. Med. Imag.*, vol. 34, no. 11, pp. 2271–2285, Nov. 2015.

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

[28] A. J. Chee and C. Alfred, "Receiver-operating characteristic analysis of eigen-based clutter filters for ultrasound color flow imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 65, no. 3, pp. 390–399, Mar. 2018.

[29] 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.

[30] 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, Jan. 2016.

[31] J. Baranger, B. Arnal, F. Perren, O. Baud, M. Tanter, and C. Demené, "Adaptive spatiotemporal SVD clutter filtering for ultrafast Doppler imaging using similarity of spatial singular vectors," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1574–1586, Jul. 2018.

[32] R. Otazo, E. J. Candès, and D. K. Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components," *Magn. Reson. Med.*, vol. 73, no. 3, pp. 1125–1136, 2014.

[33] 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 *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 1080–1084.

[34] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, p. 11, May 2011.

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

Authorized licensed use limited to: Weizmann Institute of Science. Downloaded on April 04,2020 at 17:31:02 UTC from IEEE Xplore. Restrictions apply.

[36] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 399–406.
[37] P. Sprechmann, A. M. Bronstein, and G. Sapiro, "Learning efficient sparse and low rank models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1821–1833, Sep. 2015.
[38] H. Sreter and R. Giryes, "Learned convolutional sparse coding," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 2191–2195.
[39] R. Giryes, Y. C. Eldar, A. M. Bronstein, and G. Sapiro, "Tradeoffs between convergence speed and reconstruction accuracy in inverse problems," *IEEE Trans. Signal Process.*, vol. 66, pp. 1676–1690, Apr. 2018.
[40] R. Giryes, Y. C. Eldar, A. M. Bronstein, and G. Sapiro, "The learned inexact project gradient descent algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 6767–6771.
[41] N. Samuel, T. Diskin, and A. Wiesel, "Learning to detect," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2554–2564, May 2019.
[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[43] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, Jul. 2017, pp. 1–5.
[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 Trans. Med. Imag.*, vol. 36, no. 1, pp. 169–180, Jan. 2017.
[45] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, p. 717, Dec. 2009.
[46] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[47] D. P. Palomar and Y. C. Eldar, *Convex Optimization in Signal Processing and Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[48] J.-J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bull. Sociol. Math. France*, vol. 93, pp. 273–299, 1965.
[49] Z. Tan, Y. C. Eldar, A. Beck, and A. Nehorai, "Smoothing and decomposition for analysis sparse recovery," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1762–1774, Apr. 2014.
[50] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
[51] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.
[52] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 7th Int. Joint Conf. Artif. Intell.*, 1981, pp. 674–679.
[53] B. D. de Senneville *et al.*, "Development of a fluid dynamic model for quantitative contrast-enhanced ultrasound imaging," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 372–383, Feb. 2018.
[54] Y. C. Eldar, "Generalized SURE for exponential families: Applications to regularization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 471–481, Feb. 2009.
[55] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 9, no. 6, pp. 1135–1151, 1981.
[56] P. Song *et al.*, "Accelerated singular value-based ultrasound blood flow clutter filtering with randomized singular value decomposition and randomized spatial downsampling," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 64, no. 4, pp. 706–716, Apr. 2017.
[57] E. Liberty, F. Woolfe, P.-G. Martinsson, V. Rokhlin, and M. Tygert, "Randomized algorithms for the low-rank approximation of matrices," *Proc. Nat. Acad. Sci. USA*, vol. 104, pp. 20167–20172, 2007.
[58] N. Halko, P. G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Rev.*, vol. 53, no. 2, pp. 217–288, 2011.
[59] P.-G. Martinsson and S. Voronin, "A randomized blocked algorithm for efficiently computing rank-revealing factorizations of matrices," *SIAM J. Sci. Comput.*, vol. 38, no. 5, pp. S485–S507, 2016.

Authorized licensed use limited to: Weizmann Institute of Science. Downloaded on April 04,2020 at 17:31:02 UTC from IEEE Xplore. Restrictions apply.