Deep-fUS: A Deep Learning Platform for Functional Ultrasound Imaging of the Brain Using Sparse Data

Tommaso Di Ianni [Member, IEEE],

Raag D. Airan

Neuroradiology Division, Department of Radiology, Stanford University School of Medicine, Stanford, CA 94305 USA.

Abstract

Functional ultrasound (fUS) is a rapidly emerging modality that enables whole-brain imaging of neural activity in awake and mobile rodents. To achieve sufficient blood flow sensitivity in the brain microvasculature, fUS relies on long ultrasound data acquisitions at high frame rates, posing high demands on the sampling and processing hardware. Here we develop an image reconstruction method based on deep learning that significantly reduces the amount of data necessary while retaining imaging performance. We trained convolutional neural networks to learn the power Doppler reconstruction function from sparse sequences of ultrasound data with compression factors of up to 95%. High-quality images from *in vivo* acquisitions in rats were used for training and performance evaluation. We demonstrate that time series of power Doppler images can be reconstructed with sufficient accuracy to detect the small changes in cerebral blood volume (~10%) characteristic of task-evoked cortical activation, even though the network was not formally trained to reconstruct such image series. The proposed platform may facilitate the development of this neuroimaging modality in any setting where dedicated hardware is not available or in clinical scanners.

Index Terms—

Functional ultrasound imaging; deep learning; convolutional neural networks; neuroimaging; Doppler ultrasound

I. Introduction

Functional ultrasound (fUS) is an innovative imaging modality that creates brain-wide neural activity maps at micrometer and millisecond-scale resolution by tracking temporal

cerebral blood volume (CBV) changes in the brain microvasculature [1]. Similar to blood oxygen level dependent functional magnetic resonance imaging (BOLD fMRI), the detected CBV signals provide an indirect measurement of local spiking activity via neurovascular coupling [2]. However, fUS yields higher spatiotemporal resolution than fMRI and uses more affordable and portable equipment, opening the possibility for functional neuroimaging performed directly at the bedside [3]–[6]. In addition, fUS is more sensitive than fMRI to optogenetically driven stimuli and provides larger effect size, potentially facilitating scientific investigations in animal models [7].

Preclinically, fUS enables imaging of neural activity in awake and freely behaving rodents and reduces the confounding factors introduced by anesthesia/sedation or physical restraint [8], [9]. Furthermore, fUS has proven useful for imaging resting state and task-evoked functional connectivity in the rat and mouse brain [2], [10], [11] and for mapping neural activation in primates during cognitive tasks and visual stimulation [12]–[14]. In humans, fUS has been used intraoperatively for image-monitored brain tumor removal surgeries [4], [5], and in neonates to visualize epileptic activity and measure functional connectivity through the anterior fontanel window [3], [6].

To detect hemodynamic changes in the brain microvascular network, fUS relies on highly sensitive power Doppler sequences based on the use of plane wave emissions. Unfocused ultrasound waves insonify the entire field of view, and the received radiofrequency (RF) data from tilted plane waves are re-focused (or beamformed) and coherently compounded to increase resolution and depth of penetration. This strategy makes it possible to continuously acquire long sequences of ultrasound data at high frame rates. The obtained compound Doppler signals are then processed to filter out the strong, undesired clutter originating from the tissue, and are squared and time-integrated to create power Doppler images with pixel amplitude proportional to the CBV (Fig. 1a).

The length of the acquisition sequence is critical to effectively discriminate the weak signals scattered by red blood cells circulating in the blood stream from the strong clutter originating in the surrounding tissue. When long observation windows are used, efficient clutter filtration can be achieved in both large and small vessels by using temporal and singular-value decomposition (SVD) filters [9], [15], [16]. Conversely, clutter filtration becomes challenging with shorter acquisitions, in particular in the smaller vessels where the blood-signal-to-clutter ratio is reduced and the low-frequency Doppler spectral components overlap with the tissue spectrum. As a result, fUS imaging implementations use hundreds of compound frames (typically 200 to 400) to create a single power Doppler image.

The need to acquire and process large ultrasound datasets poses high demands on the hardware platform in terms of storage capacity and computational power, with data throughputs on the order of 250 MSa/image (see detailed calculations in the Appendix). These requirements make real-time fUS imaging challenging even in graphics processing unit (GPU) implementations, and these considerations are yet more relevant for volumetric fUS sequences [17]–[19]. Therefore, it is highly desirable to achieve state-of-the-art (SoA) fUS imaging performance with shorter ultrasound acquisitions, as this may effectively improve access to this imaging modality and expedite its clinical translation.

Here we propose a deep learning platform to reconstruct power Doppler images from temporally and spatially sparse compound datasets. We implemented convolutional neural networks (CNNs) based on an encoder-decoder architecture (U-Net) [20]. Variants of this model have been used for biomedical image reconstruction in applications spanning compressed sensing MRI [21], sparse-projection photoacoustic imaging [22], and sparse X-ray computed tomography [23]. We modified the U-Net by (*i*) adding an input layer of 3-D convolutional filters that extract spatiotemporal features from the 3-D input data; (*ii*) using residual blocks that implement shortcut connections between the input and output features in each layer; and (*iii*) using a custom loss function.

Prior CNN applications in medical ultrasound imaging include contrast improvement [24] and image de-speckling [25], ultrasound contrast agent localization and tracking [26], [27], and under-sampled and adaptive beamforming [28]–[31]. To our knowledge, our platform is the first attempt to learn a reconstruction mapping between the sparse sequence of compound ultrasound data and the power Doppler output image, without requiring any prior model-based information (Fig. 1b, c). We trained the networks on high-quality power Doppler images from *in vivo* acquisitions in rats and using a custom loss function.

II. Methods

A. Deep-fUS Networks

We modified a U-Net and trained it to perform the power Doppler reconstruction task. This fully convolutional neural network is based on an encoder/decoder architecture. The encoder progressively down-samples the input data and learns high-level features that are propagated to the next stages. The decoder uses up-sampling operators to increase the resolution of the encoder features and to consecutively restore the input resolution at the output stage. Skip connections between the encoding and decoding paths allow retention of context information, which is propagated to the symmetric up-sampling layers.

Software for all the Deep-fUS networks, trained models, and test data sets are available at https://github.com/todiian/deep-fus.

1) 3D-Res-UNet: In the 3D-Res-UNet model, we modified the U-Net by adding an input layer of 4 3-D convolutional filters followed by rectified linear unit (ReLU) activations. This layer extracts spatiotemporal features from the 3-D input structure. In addition, we replaced convolutional layers with residual blocks composed of two cascaded Conv/ReLU/Dropout layers and included shortcut connections between the input and output features. Residual blocks were arranged in a 5-layer encoder followed by a 4-layer decoder and implement 3×3 convolutions followed by ReLU activations and a dropout layer to improve regularization. We used 1×1 convolutions at the input of each layer to equalize the number of input and output features of each residual block. In the encoder path, down-sampling is performed by a 2×2 max pooling operator that halves the resolution in both the image dimensions. In the decoder, 2×2 transposed convolutions with ReLU activations are used as up-sampling operators. The number of channels is progressively increased in the encoder (32, 64, 128, 256, and 512 filters) and then decreased in the decoder (256, 128, 64, and 32 filters). The output layer is a single-channel 1×1 convolution block. The stride is equal to 1 in

all the convolutional layers and 2 in the max pooling and transposed convolution blocks. This network has a total of 9,788,421 trainable parameters (Table II). The size of the filter kernels at the input stage and the dropout rate were considered as hyperparameters and were optimized via Bayesian optimization. All the convolutional kernels were initialized using the He initialization [32].

2) 3D-UNet, UNet, and PP-UNet: In addition to the 3D-Res-UNet, we trained and optimized three networks. The 3D-UNet uses simple convolutional blocks in place of residual blocks. Specifically, each layer is composed of 2 consecutive 3×3 convolution blocks, each followed by ReLU activations and dropout for network regularization. The output layer is a single-channel 1×1 convolution block. The stride is equal to 1 in all the convolutional layers and 2 in the max pooling and transposed convolution blocks. The size of the filter kernels in the first layer and the dropout rate were considered as hyperparameters and were optimized using Bayesian optimization.

The UNet is analogous to the 3D-UNet except for the absence of the 3-D convolutional filters at the input. These two networks were independently trained and optimized to separately analyze the effects on the reconstruction performance of the input 3-D convolutional filters and of the residual shortcut connections. In addition, we trained and optimized a network with the same characteristics as the above U-Net to perform the post-processing of power Doppler images that were generated by conventional processing of sparse compound sequences. We refer to this network as postprocessing (PP)-UNet.

All the convolutional kernels were initialized using the He initialization [32].

B. Datasets

We trained the networks to learn a function $y = f(x)$ that maps the input sequence x of compound frames of $N_x \times N_y$ pixels to the output power Doppler image y of dimensions $N_x \times N_y$ (Fig. 1b). In all our experiments, we used images of 96×96 pixels, and we standardized the input compound datasets. We chose to base the processing on beamformed data instead of sensor RF data to minimize data throughput and storage (see Appendix). SoA images were obtained from *in vivo* acquisitions of coronal and sagittal slices of the rat brain reconstructed by state-of-the-art power Doppler processing using 250 complex compound frames. To improve the network regularization, we performed random cropping when more than 96 pixels were available in any image dimension, and a random horizontal flipping was applied with a probability of 50%.

The training, validation, and test sets were created by selecting pairs of compound data and power Doppler images from a total of 58 *in vivo* acquisitions of coronal brain slices recorded in n = 15 rats between 2.7 mm anterior and 7.0 mm posterior to bregma [33]. In addition to the variability given by the different subjects and anteroposterior position, each pair captured distinct hemodynamic states and levels of anesthesia. We used 740 pairs from 50 data acquisitions for training, 40 pairs from 4 acquisitions for validation, and 40 pairs from 4 acquisitions for testing. Furthermore, in n = 1 rat we recorded data in sagittal view to demonstrate the generalization capabilities of the method.

We performed under-sampling of the compound sequences in the temporal domain by selecting the first k frames in each sequence (Fig. 1b). We retained only the real part of the beamformed data. For the experiments in Fig. 6, we also under-sampled the compound frames in the image domain by selecting sub-samples of pixels with a ratio $m = N_{Ret}/N_{Tot}$, with N_{Ret} the number of retained pixels and $N_{Tot} = 96^2$ the number of total image pixels.

We calculated the final compression factor as

$$CF = \left(1 - \frac{k}{250}\frac{1}{2}m\right) * 100\%, \tag{1}$$

where the factor of 1/2 accounts for the missing imaginary part.

C. Training and Hyperparameter Optimization

At each iteration, the networks predict a new estimate $\hat{y}_i$, and the parameters are learned using the Adam optimizer with default settings $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-7}$ [34], [35] to minimize the loss function

$$L(y, \hat{y}) = \lambda L_{SSIM}(y, \hat{y}) + (1 - \lambda)L_{MAE}(y, \hat{y}) \tag{2}$$

with

$$L_{MAE}(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\|y_i - \hat{y}_i\|_1}{N} \tag{3}$$

$$L_{SSIM}(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n} 1 - SSIM(y_i, \hat{y}_i). \tag{4}$$

In the above equations, y denotes the SoA training images, $\|\cdot\|_1$ the l_1 norm, N the number of image pixels, and n the number of examples. The structural dissimilarity index metric loss L_{SSIM} is a perceptual loss based on the structural similarity index metric (SSIM), which integrates luminance, contrast, and structural information [36]. A kernel of 3×3 pixels was used for the SSIM calculation. We considered the learning rate and the parameter λ as hyperparameters, and their optimal value was determined via Bayesian optimization.

We based our quantitative performance analysis on the SSIM of the reconstructed images versus the respective SoA images, the normalized mean squared error NMSE $= \|y_i - \hat{y}_i\|_2/\|y_i\|_2$, with $\|\cdot\|_2$ the l_2 norm, and on the peak signal-to-noise ratio (PSNR). We implemented the networks in Python using TensorFlow 2.1 with Keras API. The networks were trained on a single NVIDIA Titan RTX GPU with 24 GB of RAM mounted on a Dell Precision Tower workstation (20-core Intel Xeon 3.3 GHz; 32 GB of RAM). The mini-batch size was set to 1 in all the experiments.

For each network, we first optimized the hyperparameters using the Bayesian optimization routine in the Keras Tuner library. We ran 15 optimization trials using the sparse dataset

with CF 75%. The optimization routine was instructed to maximize the validation SSIM. Each trial trained the reconstruction CNNs for 2500 epochs and selected the model with the best performance. The results of the optimal hyperparameter search for all the networks are reported in Table I. Then, we trained the CNNs with the respective optimal hyperparameters using CFs of 80%, 85%, 90%, and 95%. We trained the 3D-Res-UNet and 3D-UNet networks for 1500 epochs (we noted that these CNNs converged more quickly during optimization), the U-Net for 2500 epochs, and the PP-UNet for 500 epochs. In all trainings, we saved the model with the best validation SSIM.

D. Ultrasound System and Data Acquisition

For ultrasound data acquisition, we used two 128-element linear array transducers (L22–14vX and L22–14vLF; Verasonics Inc.) operating at a 15-MHz center frequency with a Vantage 256 research scanner (Verasonics Inc.). The probes are geometrically identical apart from the focus in the elevation plane; the L22–14vX is focused at a distance of 8 mm, and the L22–14vLF is focused at 20 mm. For exact positioning relative to the skull landmarks, the imaging probe was housed in a custom 3-D printed holder mounted on a motorized positioning system. Ultrasound gel was used for acoustic coupling. We used tilted plane waves at angle ($-6°$, $-3°$, $0°$, $3°$, 6) emitted with a pulse repetition frequency of 19 kHz. Two consecutively emitted plane waves were averaged for each angle to increase the signal-to-noise ratio, giving a total of 10 emissions per compound frame. We acquired data for 250 compound frames at a rate of 1 kHz (i.e., acquiring a new sequence of compound frames takes 250 ms), and the data for each compound sequence ($250 \cdot 10$ emissions) were transferred in batch to the host computer. Compound frames were created by beamforming the received sensor RF data in a regular grid of pixels of 100 μm $\times$ 100 μm in an NVIDIA Titan RTX GPU using a GPU beamformer [37]. Ultrasound data were acquired asynchronously and continuously, i.e., a new sequence of frames was acquired during processing of the previous sequence and held in the scanner buffer until the host computer was available. The compound frames were saved on the host machine for offline processing. The final power Doppler frame rate was 0.6 frames/s.

E. Conventional Power Doppler Processing

Sequences of compound ultrasound frames were processed in MATLAB (MathWorks, Inc.) for clutter filtration and power Doppler computation. We used a 5th-order temporal high-pass Butterworth filter with a cutoff frequency of 40 Hz cascaded with an SVD filter that eliminates the first singular value [9]. In the Doppler space, frequencies are linearly proportional to the velocity of the scatterers from which the Doppler signal originated. Therefore, it is expected that signals emanating from the slowly moving tissue surrounding the blood vessels (clutter) are positioned at around 0 Hz, and this assumption justifies the use of a temporal high-pass filter. Singular value decomposition filters aim to eliminate highly coherent signal components and assume that, while blood signals are highly incoherent due to the time-varying stochastic distribution of the moving scatterers (red blood cells), tissue signal maintains a high degree of correlation over time. At each pixel location (x, y), the intensity of the filtered signal was then calculated to find the power Doppler value $I(x, y) = \int s^2(x, y, t)dt$ (Fig. 1a). For the SoA processing (250 complex compound frames), the entire time window of 250 ms was integrated.

F. Animal Preparation and Imaging Experiments

The experimental protocol for the animal study was approved by the Institutional Animal Care and Use Committee at Stanford University. The university's animal care and use program and facilities are AAALAC International accredited, PHS-assured, and USDA licensed. Long Evans and Sprague Dawley rats (Charles River; n = 16; age 10–14 weeks; weight 260–400 g) were used in this study. We prepared the animals by performing a bilateral surgical craniotomy and chronic prosthesis implant using previously published protocols [8]. Briefly, animals were anesthetized with 3.5% isoflurane in oxygen and anesthesia was maintained with 1.5% isoflurane. Rats were placed in a stereotaxic frame during surgery for head fixation and orientation. Body temperature was monitored by a rectal probe and maintained at 36.5 °C using a warming pad (RightTemp Jr.; Kent Scientific). A pulse oximeter was used to monitor heart rate and arterial oxygen saturation (MouseStat Jr.; Kent Scientific). We administered an anti-inflammatory agent to prevent brain swelling and inflammation (1 mg/kg dexamethasone intraperitoneally). After a skin incision was performed, parietal and frontal skull bone fragments (AP +4 to −9 mm; ML ±6 mm) were cut using a handheld high-speed drill with a 0.7 mm drill bit (Fine Science Tools). We gently removed the bone flaps, paying special attention to avoid any damage to the dura mater. We used dental cement (Tetric EvoFlow; Ivoclar Vivadent) to seal a 125 μm thick polymethylpentene prosthesis covering the entire craniotomy. The bone was pre-treated with a bonding agent (iBOND Total Etch; Kulzer). The space between the dura mater and the polymer prosthesis was filled with 0.9% sterile saline. Animals were then allowed to recover for 1 week before the first imaging session.

During the imaging sessions, animals were either anesthetized and kept under anesthesia with 1.5% isoflurane while placed in a stereotaxic frame or were lightly sedated with 0.5% isoflurane and kept in a restraining apparatus [38]. The restrained imaging protocol was also used in the lightly sedated fUS experiment of Fig. 7.

G. Visual Stimulation Protocol and Functional Activity Maps

To evaluate whether the Deep-fUS approach provides sufficient accuracy in the reconstruction of time series of power Doppler images in a functional neuroimaging application, we imaged visual task-evoked brain activation in rats exposed to binocular green light stimulation. Rats were anesthetized, placed in a stereotaxic frame, and kept in a dark chamber for at least 30 min prior to the visual stimulation session for dark adaptation. Six bilateral visual stimuli were delivered using two green light LEDs driven by a custom circuit. We controlled the stimulus pattern through a microcontroller board (Arduino Uno) connected to MATLAB via the serial port and interfaced with the Verasonics scanner for synchronization with the imaging sequence. For each light stimulus, the LEDs were flashed for 30 s at a frequency of 3 Hz. Each stimulus was followed by a >30 s pause in a pseudo-random fashion. We did not average the resulting CBV traces over multiple trials. This stimulation protocol was shown to maximize visual cortex response in prior fUS imaging studies [39].

The temporal CBV signals were filtered using a 6th-order median filter and a 2nd-order Butterworth highpass filter with cutoff frequency 0.006 Hz to remove the DC offset.

Functional activation maps were then created by calculating the Pearson's correlation coefficient *r* between the temporal power Doppler signal and the stimulus pattern [1], [9]. We used a Fisher's transformation to calculate the *z* scores as

$$z = \frac{1}{2}\sqrt{N-3}\, ln\frac{1+r}{1-r} \tag{5}$$

with *N* = 250 the number of temporal samples for a total time of 416 s. Each pixel was considered significant for *z* > 4.74, corresponding to *P* < 0.01 in a one-tailed *t*-test after Bonferroni correction. We used this threshold to create binary activation maps that only show the significant pixels.

III. Results

The received sensor RF data from 10 plane wave emissions were beamformed in a regular grid of 96 × 96 pixels with a spatial resolution of 100 μm × 100 μm to create compound frames. Sequences of compound frames were then processed to compute the power Doppler images. The conventional processing achieves a satisfactory level of detail in coronal brain images reconstructed from 250 complex compound frames (Fig. 2b). The resulting SoA images were used for the CNN training and as a reference for evaluating the reconstruction performance. To test the power Doppler reconstruction with under-sampled data, we retrospectively created sparse data sequences by selecting subsamples of *k* compound frames from each sequence, with CF of 75% (*k* = 125), 80% (*k* = 100), 85% (*k* = 75), 90% (*k* = 50) and 95% (*k* = 25). Power Doppler images reconstructed by the conventional processing with under-sampled data appear increasingly noisy due to the reduced blood flow sensitivity (Fig. 2c).

A. Deep-fUS Power Doppler Reconstruction

The Deep-fUS networks blindly solve a reconstruction problem to directly extract the power Doppler values from a sequence of compound frames. The networks take in input a sparse compound sequence and output the corresponding power Doppler image (Fig. 1b). The results of the Bayesian hyperparameter optimization are reported in Table I.

The Deep-fUS reconstruction restored the reference imaging performance and was able to create maps of the rat brain microvasculature from sparse data with CF of up to 95% (Fig. 2–4). Our CNNs produced a considerable improvement in the under-sampled power Doppler reconstruction when compared to the conventional processing, as confirmed by the quantitative metrics in Table II. While all the trained CNNs performed significantly better than the conventional processing with sparse data, the 3D-Res-UNet achieved overall superior reconstruction performance, with maximum SSIM of 0.92, PSNR of 30.29 dB, and minimum NMSE of 0.04. Introducing the 3-D convolutional input layer resulted in a maximum SSIM improvement of 0.07 (CF 75%), PSNR improvement of 2.16 dB (CF 80%), and NMSE reduction of 0.08 (CF 95%). The residual connections were responsible for a further maximum SSIM increase of 0.01, PSNR increase of 0.75 dB, and NMSE reduction of 0.02, all in the CF 75% case. As further discussed in Sec. III–B, the 3D-Res-UNet also

provided overall better performance in the computation of functional activation maps, for which we report the mean absolute error (MAE) in Table II.

Fig. 2 displays a representative coronal slice of the rat brain reconstructed by the 3D-Res-UNet and by the conventional approach, and scatter plots in Fig. 3a highlight that reconstruction errors are more prominent in correspondence of the lower power Doppler values, particularly in the case of conventional processing. To further confirm the generalization capabilities of our approach, we used the network trained on coronal brain images to reconstruct a sagittal slice, as displayed in Fig. 4. The mean prediction time for the 3D-Res-UNet calculated on the test set was between 4.4 and 13.5 ms/image (Table III). Movies showing side-by-side comparisons of the conventional and Deep-fUS reconstruction with CF 75%, 85%, and 95%, are available in Video S1–3.

The PP-UNet provides imaging performance comparable to the 3D-Res-UNet (Table II) but adds a run-time overhead of 212 ms/image (averaged over all the CFs) for the conventional processing of the sparse power Doppler images, which are then provided as input to the neural network for post processing (Table III). It is also worth noting that this approach, like the conventional method, is inherently dependent on the design of the tissue clutter filter, and a wide range of parameters have been used in the fUS literature for the temporal high-pass and SVD filters. Therefore, a data-driven method that does not require prior model-based knowledge would be highly advantageous and potentially simplify fUS implementations. Interestingly, we noted that the learned convolutional filters in the input layer of the 3D-Res-UNet implement highpass transfer functions with strong rejection of the 0-Hz component (Fig. 1d). These filters appear to mimic the filters used in the conventional processing but are learned directly from the data during training.

B. Task-Evoked Functional Activity Imaging

Power Doppler images of the brain vasculature only provide a screenshot of the CBV signal at a given time. In fUS neuroimaging applications, the functional information is extracted from the temporal evolution of the CBV signals in the form of temporal correlations with a stimulus (i.e., functional activation) or between brain regions of interest (i.e., functional connectivity). We thus sought to test the performance of the Deep-fUS approach in a visual-evoked functional activation application.

The SoA activation map is shown for reference in Fig. 5b. This was created by using conventionally reconstructed power Doppler images using the full compound dataset and shows significant bilateral activation of the primary and secondary visual cortices (V1/2) and superior colliculus (SC). In Fig. 5c we show the activation maps created using power Doppler time series reconstructed by Deep-fUS (3D-Res-UNet) using sparse data with CF between 75% and 95%. Although the quality of the activation maps degraded with increasing data sparsity, significant V1/2 and SC activation could be detected with a CF up to 95%. The network generalized well to the reconstruction of time series of power Doppler images and was able to accurately reproduce the small changes in relative CBV signal (~10%; Fig. 5e–f) characteristic of visual-evoked cortical activation. It is worth pointing out that the network did not formally learn to perform the reconstruction of such image

sequences, as it was trained on single images, and had no direct knowledge of the functional information content of the sequence.

Notably, Deep-fUS with CF 95% performed better than the conventional approach with CF 75%. With shorter data sequences (i.e., higher CF), the conventional processing provides increasingly noisy CBV temporal signals that result in lower and non-significant correlations with the stimulus (Fig. 5e–f). This is also confirmed by the quantitative error metric (MAE) in Table II. Introducing the 3-D convolutional input layer and residual connections reduced the activation maps MAE in all the cases except for CF 90%, as compared to the simple U-Net. This may be due to the short temporal signals that make it more challenging to train the 3-D filters.

To investigate an alternative approach to data compression, we created temporally and spatially under-sampled sequences by retaining only a subset of compound samples in each frame, with a spatial under-sampling ratio $m = 1/2$ and $m = 1/4$. The spatial samples were selected as shown in the bottom plots of Fig. 6. We used $k = 50$ and $k = 100$ in the two cases to equalize the CF to 95%. This approach improved the quality of the functional activation maps compared to the case with temporal under-sampling only (Fig. 6), and the resulting activation map MAE was 0.1406 for $m = 1/2$ and 0.1279 for $m = 1/4$. These results suggest that spatial sparsity may be a viable option to further increase data compression while retaining the advantages of longer acquisitions. On the other hand, this approach does not benefit from the decreased sensitivity to motion artifacts discussed in the next section.

A movie of the Deep-fUS power Doppler series and relative CBV variation from the visual-evoked experiment with temporally and spatially sparse data is provided in Video S-4.

C. Motion Artifact Reduction

To determine whether shorter acquisition sequences reduce the occurrence of motion artifacts, we used Deep-fUS to reconstruct a time series of power Doppler images acquired in a lightly sedated and restrained animal. We computed the SSIM of each image in the series versus a baseline calculated as the median of all images in the acquisition, then we applied a SSIM threshold to filter out the images that showed significant degradation, possibly due to animal motion (Fig. 7). In the case of SoA processing using the full compound sequence, 8.2% of power Doppler images were discarded by the filter (Fig. 7a). Image scrubbing was reduced to between 4.5% (with CF 75%) and 2.1% (with CF 95%), giving a maximum scrubbing reduction of 74%. Fig. 7c displays a representative SoA power Doppler image that was discarded by the SSIM filter. Motion artifacts were resolved in the same image processed by Deep-fUS (Fig. 7d). We provide a side-by-side comparison of the SoA and Deep-fUS image scrubbing in the lightly sedated fUS experiment in Video S-5.

IV. Discussion

Deep learning and CNNs are drawing increasing attention for the reconstruction and processing of biomedical images with sparse data [40]–[42]. In medical ultrasound, several strategies have been proposed to restore high image quality while reducing data sampling, transmission, and processing [28], [29], [31], [43]. With the exception of a single

preliminary study reporting deep learning of color Doppler images [44], however, CNNs have not been applied as extensively to ultrasound imaging of blood flows. We proposed a deep learning platform for the direct reconstruction of power Doppler images from a 3-D space of sparse compound ultrasound data (Fig. 1). Our proposed approach largely enhanced imaging performance compared to the conventional method with compression factors up to 95%, as clearly indicated by the presented quantitative metrics (Fig. 3). We demonstrate that our method is able to reconstruct time series of power Doppler images with sufficient accuracy to compute functional activation maps in a task-evoked neuroimaging application (Fig. 5 and 6). Although it was not formally trained for such reconstruction task, the network generalized well and accurately reproduced changes in relative CBV signals on the order of 10%. Additionally, we show that by minimizing the length of the acquisition sequence, our network allows greater robustness to motion artifacts in an experiment with a lightly sedated animal and is less sensitive to image scrubbing (Fig. 7).

The first advantage of using sparse sequences is the net reduction in data acquisition, storage, and processing resources, which may effectively simplify fUS development. To provide an indication of a representative scenario, in our implementation with 250 compound images each made of 10 plane waves, we acquire and beamform 2500 full B-mode-equivalent frames to compute a single power Doppler image. The acquisition lasts 250 ms, and RF data from each acquisition are sampled and stored on the scanner. The first bottleneck is the direct memory access (DMA) for data transfer to the host machine via PCIe bus. Data are then beamformed at runtime in a GPU to reduce storage requirements (see calculation in the Appendix) and saved on a solid-state drive. Processing (on the host machine) and data acquisition (on the scanner) happen in parallel. The power Doppler computation is performed offline, as SVD filters are also computationally demanding. Our resulting frame rate excluding power Doppler processing is 0.6 Hz. However, it is important to note that our field of view is limited, as we are imaging the rat brain in coronal view (~10 mm × 10 mm), but these considerations become yet more significant in situations where larger regions are imaged, as in larger animals or humans.

The suggested approach may facilitate the development of fUS neuroimaging in any setting where dedicated hardware is not available or in clinical scanners, making this technology more affordable and opening the way to new potential applications based on this imaging modality. Additionally, sparse sequences may prove beneficial in experimental situations where fUS acquisitions need to be interleaved with long therapeutic ultrasound pulses, such as in the monitoring of focused ultrasound neurointerventions [45], [46]. Although in this study we retrospectively under-sampled the compound data, we clearly demonstrate that the network may considerably reduce the beamforming complexity and eliminate the need for computationally demanding filters [15], [16]. Additionally, the network has the potential to increase the imaging frame rate and to facilitate the implementation of volumetric fUS imaging using swept linear arrays [6]. The platform and conceptual framework that we propose may be adapted to other high-frame-rate Doppler ultrasound imaging modalities, including vector flow imaging, to expedite their deployment in portable ultrasound systems [47], [48].

To create sparse sequences, we selected subsets of compound images in the initial portion of the original sequence. This approach has the advantage of yielding shorter temporal acquisition windows and reduces the occurrence of motion artifacts and data scrubbing, which are inevitable in mobile rodent and handheld fUS imaging applications [5], [9], [49]. Alternative sampling approaches, including interleaved or randomly distributed frames, may be integrated with our proposed platform by training and optimizing the networks with the desired sampling method. We expect that longer observation times will facilitate the power Doppler reconstruction, as displayed in Fig. 6, but will make the sequence more sensitive to motion artifacts, as discussed in Sec. III–C. Adaptive sampling based on a motion measure could be an exciting future development for this power Doppler reconstruction approach.

We acknowledge that variants of the U-Net have been previously applied to different biomedical imaging modalities. However, most of the literature is focused on removing artifacts from sub-optimally reconstructed images. We were specifically interested in demonstrating a data-driven reconstruction method that, once trained, requires no prior model-based knowledge of the image formation process nor requires hand-picked parameters. We decided to base our implementation on the U-Net as we hypothesized that its encoder-decoder architecture would fit the nature of our data. A critical step in the power Doppler reconstruction process is the filtration of the strong clutter signal originating from the moving tissue. In the 3-D space formed by the image plane and Doppler time, the clutter signal is slowly varying in the temporal domain and highly correlated in the spatial domain, therefore it is crucial to account for both spatially and temporally varying features in the reconstruction process. By progressively expanding the spatial field of view in the encoder layers and with the input filters performing temporal convolutions, our network extracts spatiotemporal features from severely under-sampled input datasets. By using more sophisticated networks, interesting applications may be developed in the future based on the current work. Unsupervised algorithms may be designed to train variants of generative adversarial networks (for example, CycleGAN [50]) on 2-D power Doppler images for the reconstruction of volumetric fUS data acquired with sparse physical apertures. Considering the cost, complexity, and bulkiness of 3-D ultrasound systems, such advances may greatly facilitate 4-D fUS imaging applications.

A main limitation of using ultrasound for brain imaging is the presence of the skull, which is highly absorbing at the imaging frequencies. This has limited clinical fUS to intraoperative applications or to scenarios with natural skull openings, such as the neonatal anterior fontanel window. These limitations may be partly overcome by using contrast agents to enhance the ultrasound signal [51]–[53].

Finally, further validation of our reconstruction approach is recommended, as changes in the functional activation maps at higher compression factors may distort the interpretation of clinical or scientific imaging data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

Appendix

Calculation of RF and Beamformed Data Throughputs

Considering compound frames composed of 10 plane wave emissions acquired with an array of 128 elements, sampled with a sampling frequency of 60 MHz ($4\times$ the pulse center frequency of 15 MHz), and covering a depth of 9.6 mm, the data throughput for each compound frame is

$$\frac{2*9.6 10^{-3}}{1.54} 10^3 * 60 10^6 * 128 * 10 = 957,510 \tag{6}$$

RF samples. Eq. (1) considers a speed of sound of 1540 m s^{-1}, and the factor 2 accounts for the pulse-echo time-of-flight. If 250 compound frames are used to compute a power Doppler image, the total number of RF samples to be transmitted from the scanner to the host computer is

$$957,510 * 250 = 239,377,500 \tag{7}$$

per image. If 96×96 pixels are beamformed in each compound frame, the beamforming load is

$$96 * 96 * 2 * 250 = 4,608,000 \tag{8}$$

pixels per power Doppler image. The factor of 2 in (3) accounts for the beamforming of complex (real and imaginary) samples.

By saving beamforming data instead of RF samples, memory usage can be reduced by

$$\frac{957,510}{96 * 96 * 2} = 52 \tag{9}$$

times, assuming equivalent data precision.

Our CNN processing approach reduces the hardware requirements for data acquisition, as well as the storage and beamforming requirements. For example, in the case of 80% compression factor with $k = 100$ real compound frames per power Doppler image and no spatial under-sampling ($m = 1$), the number of RF samples is reduced to

$$957.510 * 100 = 95,751,000 \tag{10}$$

(60% reduction), while the number of beamformed pixels to be processed and stored in memory is reduced to

$$96 * 96 * 100 = 921,600 \tag{11}$$

(80% reduction). With $k = 100$ and $m = 1/4$ (Fig. 6), the number of RF samples is unchanged while the number of beamformed pixels becomes

$$\frac{1}{4} * 96 * 96 * 100 = 230,400 \tag{12}$$

(95% reduction). For the sake of simplicity, in the main text we only refer to the beamformed data reduction to compute the data compression factors.

References

[1]. Macé E, Montaldo G, Cohen I, Baulac M, Fink M, and Tanter M, "Functional ultrasound imaging of the brain," Nature Methods, vol. 8, no. 8, pp. 662–664, 2011. [PubMed: 21725300]

[2]. Macé É et al. , "Whole-brain functional ultrasound imaging reveals brain modules for visuomotor integration," Neuron, vol. 100, no. 5, pp. 1241–1251, 2018. [PubMed: 30521779]

[3]. Demene C et al. , "Functional ultrasound imaging of brain activity in human newborns," Sci. Transl. Med, vol. 9, no. 411, Oct. 2017, Art. no. eaah6756. [PubMed: 29021168]

[4]. Imbault M, Chauvet D, Gennisson J-L, Capelle L, and Tanter M, "Intraoperative functional ultrasound imaging of human brain activity," Sci. Rep, vol. 7, no. 1, pp. 1–7, Dec. 2017. [PubMed: 28127051]

[5]. Soloukey S et al. , "Functional ultrasound (fUS) during awake brain surgery: The clinical potential of intra-operative functional and vascular brain mapping," Frontiers Neurosci., vol. 13, pp. 1–14, Jan. 2020.

[6]. Baranger J et al. , "Bedside functional monitoring of the dynamic brain connectivity in human neonates," Nature Commun, vol. 12, no. 1, pp. 1–10, Dec. 2021. [PubMed: 33397941]

[7]. Edelman BJ, Ielacqua GD, Chan RW, Asaad M, Choy M, and Lee JH, "High-sensitivity detection of optogenetically-induced neural activity with functional ultrasound imaging," NeuroImage, vol. 242, Nov. 2021, Art. no. 118434. [PubMed: 34333106]

[8]. Sieu L-A et al. , "EEG and functional ultrasound imaging in mobile rats," Nature Methods, vol. 12, no. 9, pp. 831–834, Sep. 2015. [PubMed: 26237228]

[9]. Urban A, Dussaux C, Martel G, Brunner C, Mace E, and Montaldo G, "Real-time imaging of brain activity in freely moving rats using functional ultrasound," Nature Methods, vol. 12, no. 9, pp. 873–878, Sep. 2015. [PubMed: 26192084]

[10]. Osmanski B-F, Pezet S, Ricobaraza A, Lenkei Z, and Tanter M, "Functional ultrasound imaging of intrinsic connectivity in the living rat brain with high spatiotemporal resolution," Nature Commun, vol. 5, no. 1, pp. 1–14, Dec. 2014.

[11]. Ferrier J, Tiran E, Deffieux T, Tanter M, and Lenkei Z, "Functional imaging evidence for task-induced deactivation and disconnection of a major default mode network hub in the mouse brain," Proc. Nat. Acad. Sci. USA, vol. 117, no. 26, pp. 15270–15280, Jun. 2020. [PubMed: 32541017]

[12]. Blaize K et al. , "Functional ultrasound imaging of deep visual cortex in awake nonhuman primates," Proc. Nat. Acad. Sci. USA, vol. 117, no. 25, pp. 14453–14463, Jun. 2020. [PubMed: 32513717]

[13]. Dizeux A et al. , "Functional ultrasound imaging of the brain reveals propagation of task-related brain activity in behaving primates," Nature Commun, vol. 10, no. 1, pp. 1–9, Dec. 2019. [PubMed: 30602773]

[14]. Norman SL et al. , "Single-trial decoding of movement intentions using functional ultrasound neuroimaging," Neuron, vol. 109, no. 9, pp. 1554–1566, May 2021. [PubMed: 33756104]

[15]. Demené C 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.

[16]. Baranger J, Arnal B, Perren F, Baud O, Tanter M, and Demené C, "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.

[17]. Rabut C et al. , "4D functional ultrasound imaging of whole-brain activity in rodents," Nature Methods, vol. 16, no. 10, pp. 994–997, Oct. 2019. [PubMed: 31548704]

[18]. Sauvage J et al. , "4D functional imaging of the rat brain using a large aperture row-column array," IEEE Trans. Med. Imag, vol. 39, no. 6, pp. 1884–1893, Jun. 2020.

[19]. Brunner C et al. , "A platform for brain-wide volumetric functional ultrasound imaging and analysis of circuit dynamics in awake mice," Neuron, vol. 108, no. 5, pp. 861–875, 2020. [PubMed: 33080230]

[20]. Ronneberger O, Fischer P, and Brox T, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 9351, 2015, pp. 234–241.

[21]. Yang G et al. , "DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," IEEE Trans. Med. Imag, vol. 37, no. 6, pp. 1310–1321, Jun. 2017.

[22]. Davoudi N, Deán-Ben XL, and Razansky D, "Deep learning optoacoustic tomography with sparse data," Nature Mach. Intell, vol. 1, no. 10, pp. 453–460, Oct. 2019.

[23]. Jin KH, Mccann MT, Froustey E, and Unser M, "Deep convolutional neural network for inverse problems in imaging," IEEE Trans. Image Process, vol. 26, no. 9, pp. 4509–4522, Sep. 2017. [PubMed: 28641250]

[24]. Luchies AC and Byram BC, "Deep neural networks for ultrasound beamforming," IEEE Trans. Med. Imag, vol. 37, no. 9, pp. 2010–2021, Sep. 2018.

[25]. Hyun D, Brickson LL, Looby KT, and Dahl JJ, "Beamforming and speckle reduction using neural networks," IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 66, no. 5, pp. 898–910, May 2019. [PubMed: 30869612]

[26]. Hyun D, Abou-Elkacem L, Bam R, Brickson LL, Herickhoff CD, and Dahl JJ, "Nondestructive detection of targeted microbubbles using dual-mode data and deep learning for real-time ultrasound molecular imaging," IEEE Trans. Med. Imag, vol. 39, no. 10, pp. 3079–3088, Oct. 2020.

[27]. Youn J, Ommen ML, Stuart MB, Thomsen EV, Larsen NB, and Jensen JA, "Detection and localization of ultrasound scatterers using convolutional neural networks," IEEE Trans. Med. Imag, vol. 39, no. 12, pp. 3855–3867, Dec. 2020.

[28]. Gasse M, Millioz F, Roux E, Garcia D, Liebgott H, and Friboulet D, "High-quality plane wave compounding using convolutional neural networks," IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 64, no. 10, pp. 1637–1639, Oct. 2017. [PubMed: 28792894]

[29]. Yoon YH, Khan S, Huh J, and Ye JC, "Efficient B-mode ultrasound image reconstruction from sub-sampled RF data using deep learning," IEEE Trans. Med. Imag, vol. 38, no. 2, pp. 325–336, Feb. 2019.

[30]. Luijten B et al. , "Adaptive ultrasound beamforming using deep learning," IEEE Trans. Med. Imag, vol. 39, no. 12, pp. 3967–3978, Dec. 2020.

[31]. Nair AA, Washington KN, Tran TD, Reiter A, and Bell MAL, "Deep learning to obtain simultaneous image and segmentation outputs from a single input of raw ultrasound channel data," IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 67, no. 12, pp. 2493–2509, Dec. 2020. [PubMed: 32396084]

[32]. He K, Zhang X, Ren S, and Sun J, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[33]. Paxinos G and Watson C, The Rat Brain in Stereotaxic Coordinates: Hard Cover Edition. New York, NY, USA: Academic, 1998.

[34]. Adam Optimizer; Keras API. Accessed: Jan. 7, 2022. [Online]. Available: https://keras.io/api/optimizers/adam/

[35]. Kingma DP and Ba JL, "Adam: A method for stochastic optimization," in Proc. 3rd Int. Conf. Learn. Represent. (ICLR), 2015, pp. 1–15.

[36]. Wang Z, Bovik AC, Sheikh HR, and Simoncelli EP, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process, vol. 13, no. 4, pp. 600–612, Apr. 2004. [PubMed: 15376593]

[37]. Hyun D, Trahey GE, and Dahl JJ, "Real-time high-framerate *in vivo* cardiac SLSC imaging with a GPU-based beamformer," in Proc. IEEE Int. Ultrason. Symp. (IUS), Oct. 2015, pp. 1–4.

[38]. Stenroos P et al. , "Awake rat brain functional magnetic resonance imaging using standard radio frequency coils and a 3D printed restraint kit," Frontiers Neurosci, vol. 12, p. 548, Aug. 2018.

[39]. Gesnik M et al. , "3D functional ultrasound imaging of the cerebral visual system in rodents," NeuroImage, vol. 149, pp. 267–274, Apr. 2017. [PubMed: 28167348]

[40]. Zhu B, Liu JZ, Cauley SF, Rosen BR, and Rosen MS, "Image reconstruction by domain-transform manifold learning," Nature, vol. 555, no. 7697, pp. 487–492, Mar. 2018. [PubMed: 29565357]

[41]. Vishnevskiy V, Walheim J, and Kozerke S, "Deep variational network for rapid 4D flow MRI reconstruction," Nature Mach. Intell, vol. 2, no. 4, pp. 228–235, Apr. 2020.

[42]. Shen L, Zhao W, and Xing L, "Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning," Nature Biomed. Eng, vol. 3, no. 11, pp. 880–888, Nov. 2019. [PubMed: 31659306]

[43]. Wiacek A, González E, and Bell MAL, "CohereNet: A deep learning architecture for ultrasound spatial correlation estimation and coherence-based beamforming," IEEE Trans. Ultrason. Ferroelectr., Freq. Control, vol. 67, no. 12, pp. 2574–2583, Dec. 2020. [PubMed: 32203018]

[44]. Huijben IAM, Veeling BS, Janse K, Mischi M, and van Sloun RJG, "Learning sub-sampling and signal recovery with applications in ultrasound imaging," IEEE Trans. Med. Imag, vol. 39, no. 12, pp. 3955–3966, Dec. 2020.

[45]. Wang JB et al. , "Focused ultrasound for noninvasive, focal pharmacologic neurointervention," Frontiers Neurosci, vol. 14, p. 675, Jul. 2020.

[46]. Wang JB, Aryal M, Zhong Q, Vyas DB, and Airan RD, "Noninvasive ultrasonic drug uncaging maps whole-brain functional networks," Neuron, vol. 100, no. 3, pp. 728–738, 2018. [PubMed: 30408444]

[47]. Ianni TD et al. , "A vector flow imaging method for portable ultrasound using synthetic aperture sequential beamforming," IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 64, no. 11, pp. 1655–1665, Nov. 2017. [PubMed: 28841555]

[48]. Di Ianni T, Kjeldsen TK, Hoyos CAV, Mosegaard J, and Jensen JA, "Real-time implementation of synthetic aperture vector flow imaging on a consumer-level tablet," in Proc. IEEE Int. Ultrason. Symp. (IUS), Sep. 2017, pp. 1–4.

[49]. Rabut C et al. , "Pharmaco-fUS: Quantification of pharmacologically-induced dynamic changes in brain perfusion and connectivity by functional ultrasound imaging in awake mice," NeuroImage, vol. 222, Nov. 2020, Art. no. 117231. [PubMed: 32795659]

[50]. Ma Z, Wang F, Wang W, Zhong Y, and Dai H, "Deep learning for *in vivo* near-infrared imaging," Proc. Nat. Acad. Sci. USA, vol. 118, no. 1, pp. 1–8, 2021.

[51]. Errico C, Osmanski B-F, Pezet S, Couture O, Lenkei Z, and Tanter M, "Transcranial functional ultrasound imaging of the brain using microbubble-enhanced ultrasensitive Doppler," Neuroimage, vol. 124, pp. 752–761, Jan. 2016. [PubMed: 26416649]

[52]. Maresca D et al. , "Acoustic biomolecules enhance hemodynamic functional ultrasound imaging of neural activity," NeuroImage, vol. 209, Apr. 2020, Art. no. 116467. [PubMed: 31846757]

[53]. Demené C et al. , "Transcranial ultrafast ultrasound localization microscopy of brain vasculature in patients," Nature Biomed. Eng, vol. 5, no. 3, pp. 219–228, Mar. 2021. [PubMed: 33723412]
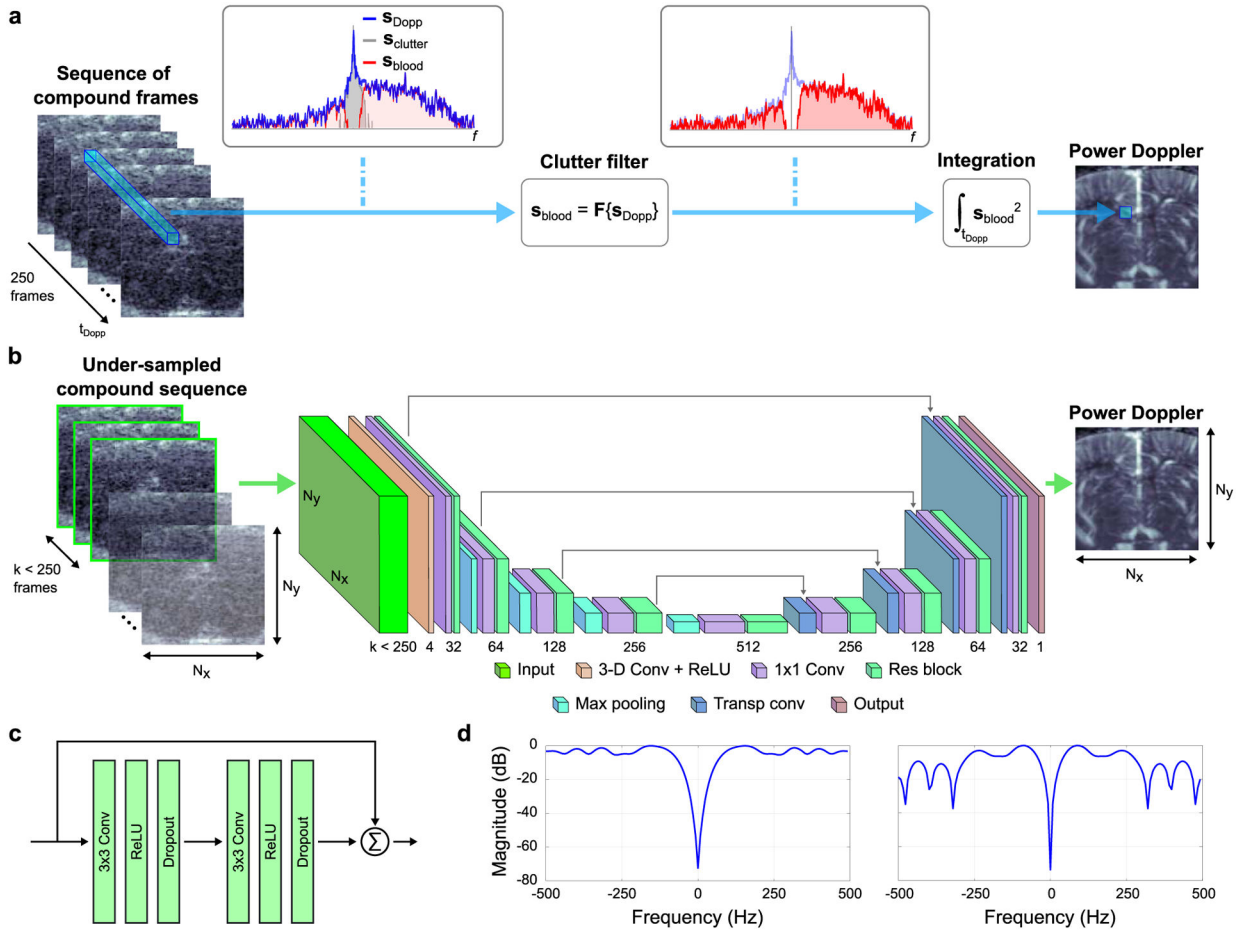
Fig. 1.
a, In the state-of-the-art processing, a power Doppler image is created from a sequence of 250 compound ultrasound frames. In each pixel, the temporal signal s_{Dopp} sampled in the Doppler time t_{Dopp}, is passed through a bank of filters F to remove the tissue clutter component $s_{clutter}$. The retained blood signal s_{blood} is squared and time-integrated to compute the power Doppler pixel value proportional to cerebral blood volume. **b**, The Deep-fUS 3D-Res-UNet architecture uses a modified U-Net network consisting of residual blocks arranged in a 5-layer encoder followed by a decoder. An input 3-D convolutional layer extracts spatiotemporal features from the 3-D input structure. The input data is an under-sampled compound sequence created by selecting the first k frames of $N_x \times N_y$ pixels (selected frames displayed with a green border). The network outputs $N_x \times N_y$ power Doppler images. **c**, Residual blocks composed of two cascaded Conv/ReLU/Dropout layers implement a shortcut connection between the input and output. **d**, Representative transfer functions of the input 3-D convolutional filters learned by the network. These were computed by performing a fast Fourier transform of the filter kernels averaged in the 3×3 spatial domain. The cutoff frequencies (−3 dB) for the two filters are 95 Hz (left) and 58 Hz (right).
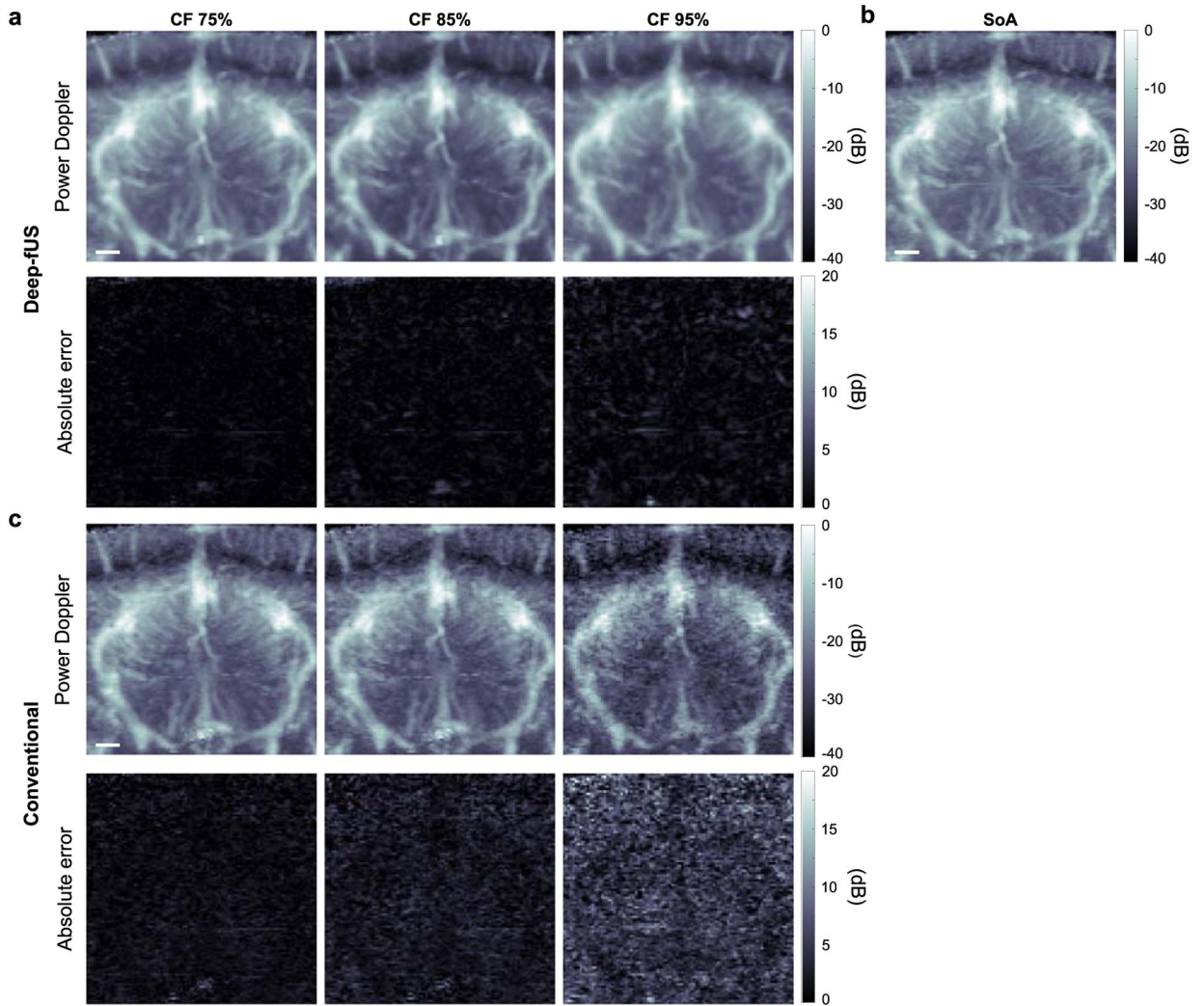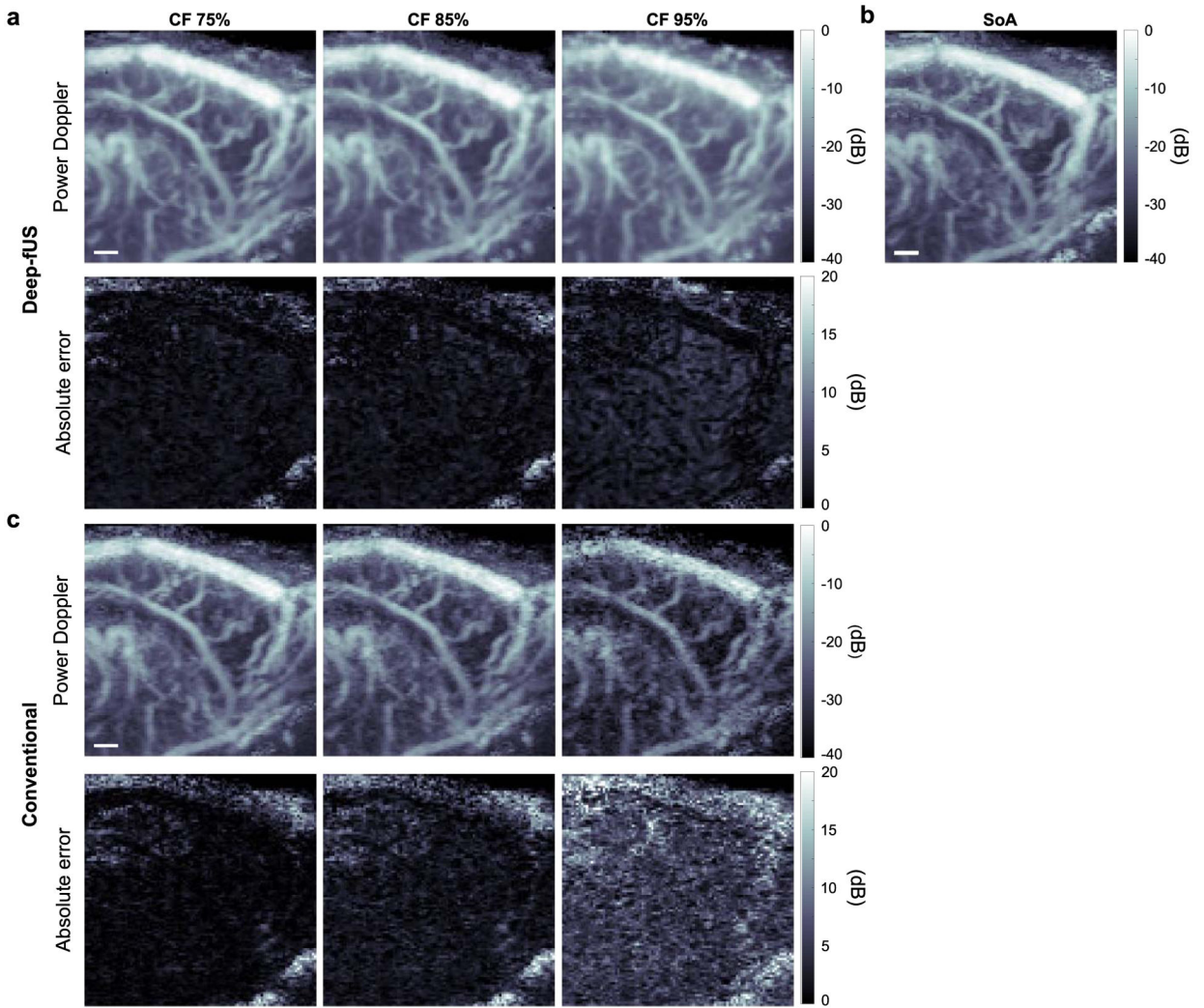
Fig. 2.

a, Representative power Doppler image of a coronal slice of the rat brain reconstructed by Deep-fUS (3D-Res-UNet) from under-sampled sequences with compression factor of (CF) 75%, 85%, and 95% (Top) and absolute error images calculated against the state-of-the-art (SoA) image (Bottom). **b**, SoA image reconstructed by the conventional processing using 250 complex compound frames. **c**, Power Doppler images reconstructed with the conventional processing using under-sampled compound data (Top) and respective absolute error images (Bottom). Scale bar in **a**, **b**, **c**: 1 mm.
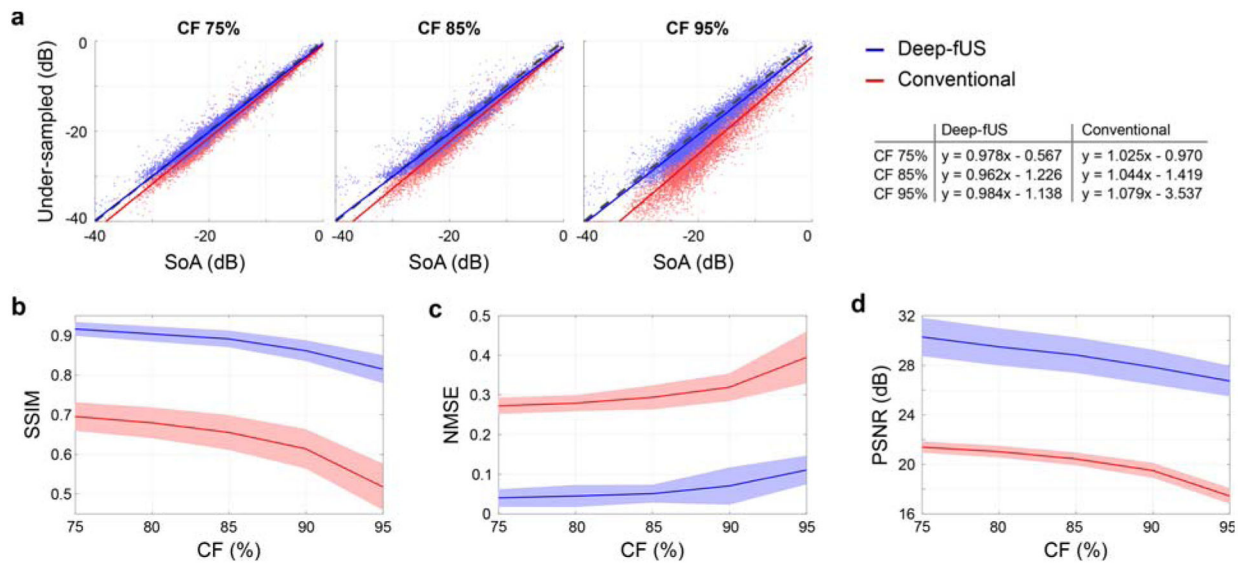
Fig. 3.
a, Scatter plots of the power Doppler pixel amplitudes and linear regression analysis (y = b₁x + b₂). **b-d**, Structural similarity index metric (SSIM), normalized mean squared error (NMSE), and peak signal-to-noise ratio (PSNR) of power Doppler images reconstructed by Deep-fUS (3D-Res-UNet; blue) and by the conventional approach (red). The quantitative metrics were calculated against the respective SoA reference images. Results are reported as mean (solid line) and standard deviation (shaded area) calculated over the test set.

Fig. 4.

a, Representative power Doppler image of a sagittal slice of the rat brain reconstructed by Deep-fUS (3D-Res-UNet) from under-sampled sequences with compression factor (CF) 75%, 85%, and 95% (Top) and absolute error images calculated against the state-of-the-art (SoA) image (Bottom). **b**, SoA image reconstructed by the conventional processing. **c**, Power Doppler images reconstructed with the conventional processing using under-sampled compound data (Top) and respective absolute error images (Bottom). Scale bar in **a**, **b**, **c**: 1 mm.
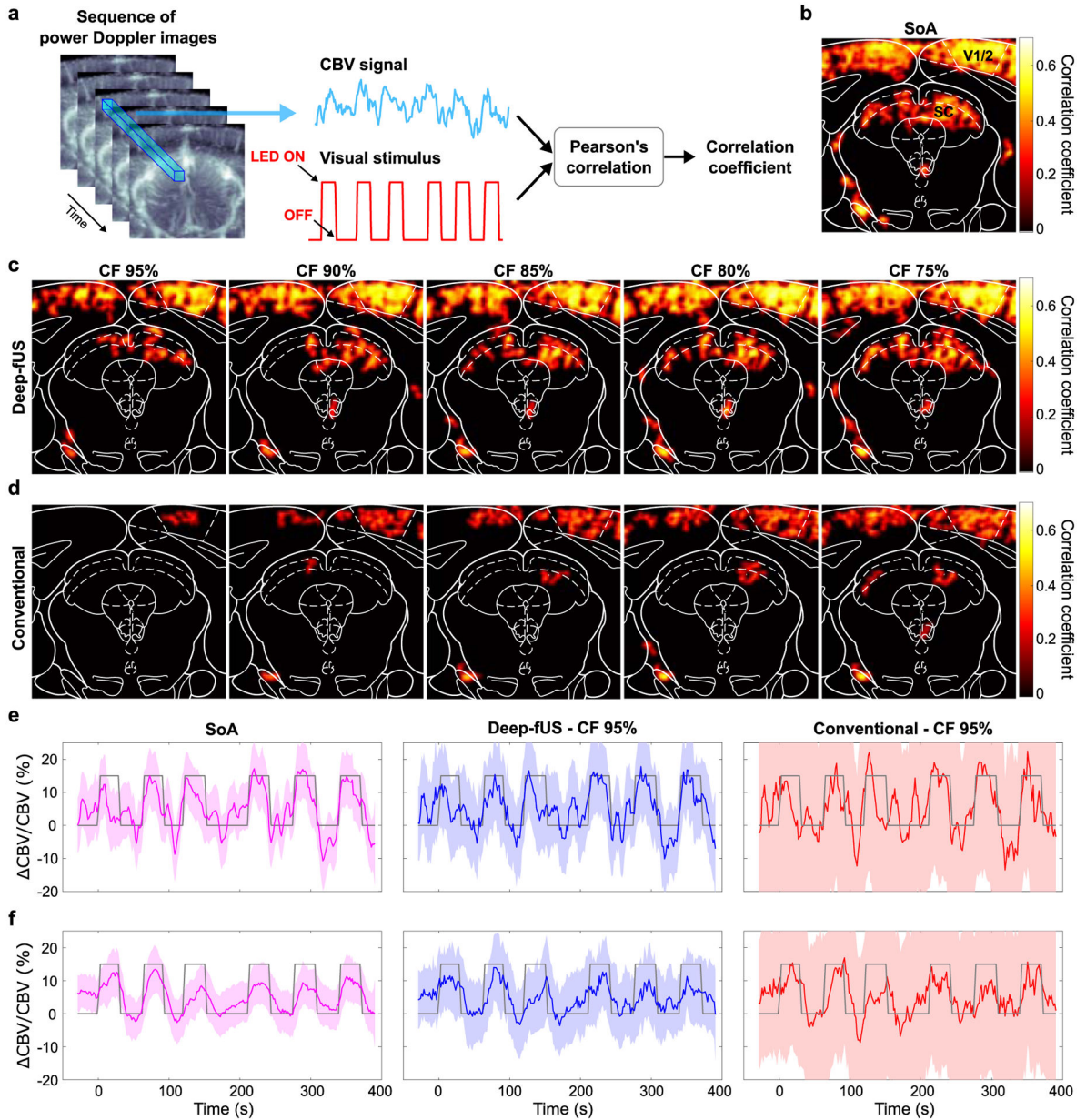
Fig. 5.

a, Time series of power Doppler images were recorded continuously during a visual stimulation task. The resulting cerebral blood volume (CBV) signals were correlated with the stimulus pattern. The visual stimulus consisted of 6 light stimuli, each with an ON time of 30 s, distributed in a pseudo-random fashion. **b**, State-of-the-art (SoA) activation map computed using power Doppler images reconstructed by the conventional approach using 250 complex compound frames. Statistically significant pixels (P < 0.01; Bonferroni corrected) are shown in the heat map. The white contour displays the slice at bregma −7.0 mm from the Paxinos brain atlas. The activation map shows significant bilateral activation of the rat primary and secondary visual cortices (V1/2) and superior colliculus (SC). **c**, Activation maps computed using power Doppler images reconstructed by Deep-fUS

with compression factor (CF) between 75% and 95%. **d**, Activation maps computed using power Doppler images reconstructed by conventional processing with CF between 75% and 95%. **e**, Relative CBV signals in the statistically significant pixels of the SoA map in V1/2 for the SoA data (magenta), and Deep-fUS (blue) and conventional processing (red) with CF of 95%. Results are reported as mean (solid line) and standard deviation (shaded area). The stimulus pattern is displayed in gray. **f**, Relative CBV signals in the statistically significant pixels of the SoA map in SC for the SoA data (magenta), and Deep-fUS (blue) and conventional processing (red) with CF of 95%.
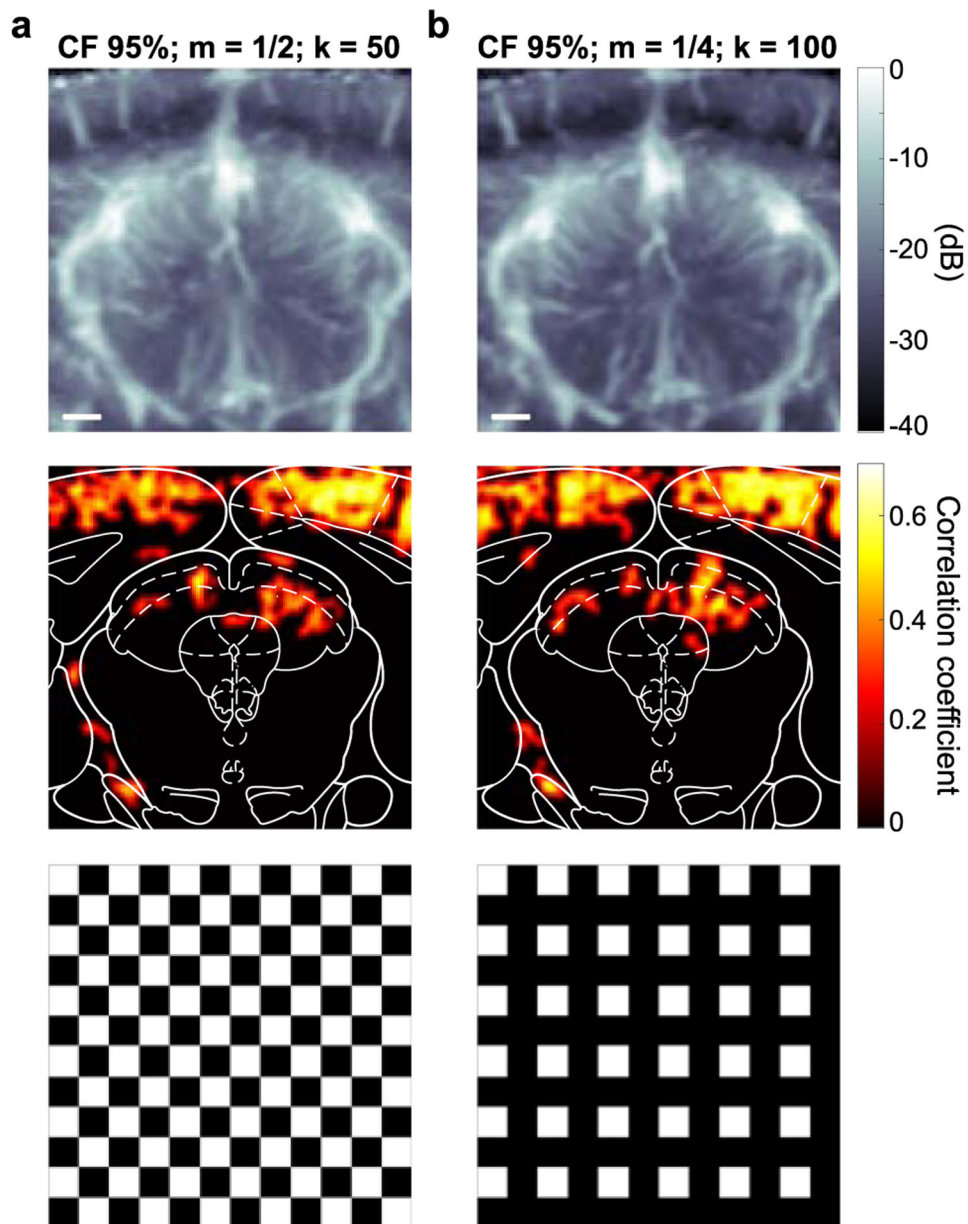
Fig. 6.
Representative power Doppler test images (top) and activation maps (middle) computed with spatially under-sampled sequences with spatial sampling ratio m = 1/2 (**a**) and m = 1/4 (**b**). To equalize the compression factor (CF) to 95%, k = 50 and k = 100 compound frames were used in the two cases. The spatial sampling maps are displayed in the bottom plots. The black and white pixels show discarded and retained pixels, respectively. The maps are shown in a sub-grid of 12×12 pixels.
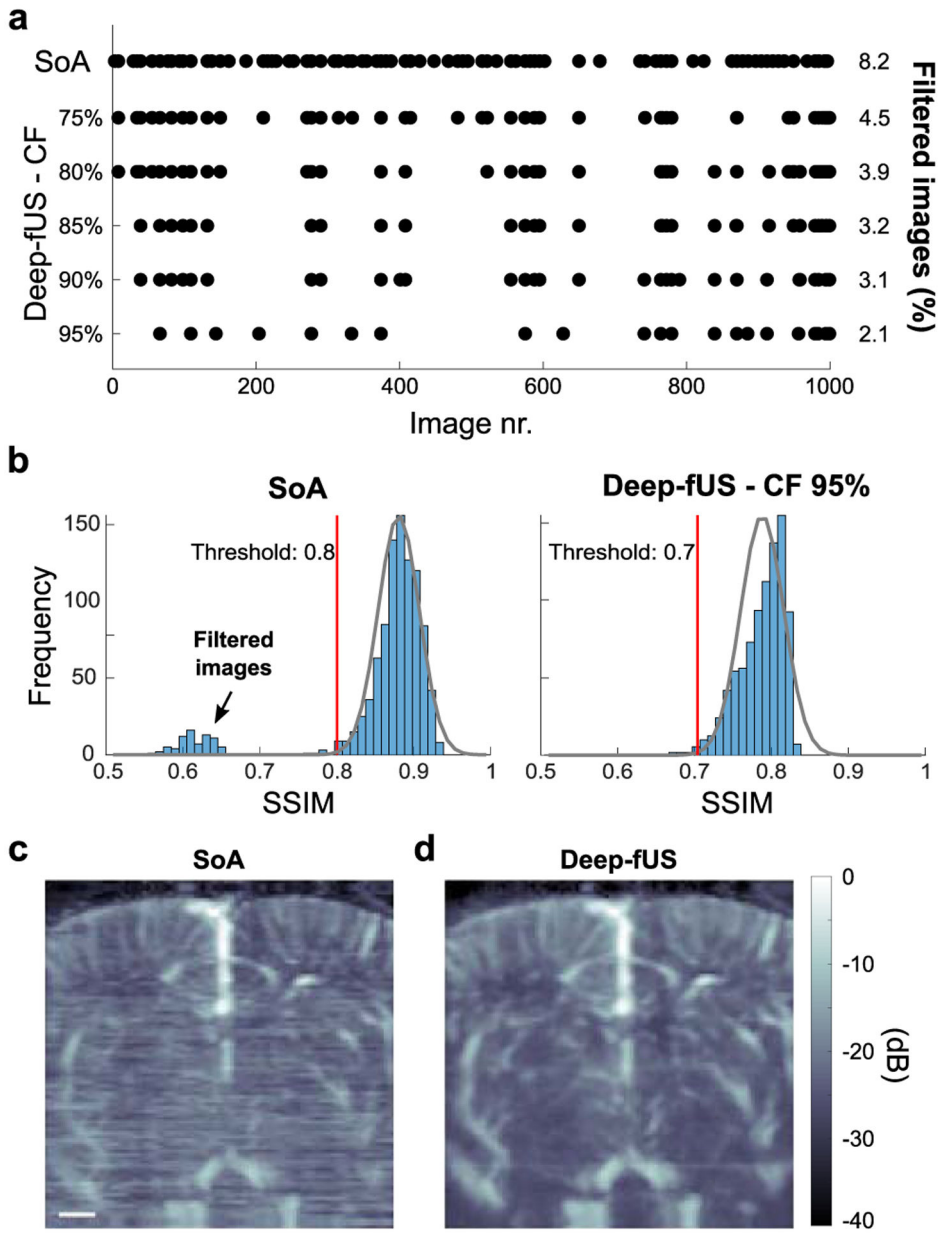
Fig. 7.

a, A series of 1000 power Doppler images was filtered based on a structural similarity index metric (SSIM) filter. Black dots display the discarded images in the series. **b**, We defined a threshold (red) to remove the images with an SSIM value lower than 3 standard deviations from the baseline. **c, d**, Representative power Doppler coronal images in a case of significant degradation in the conventional reconstruction (**c**) that was completely resolved with under-sampled processing (**d**). Scale bar: 1 mm.

TABLE I

Results of Bayesian Optimization

Network	Hyperparameter	Optimized value
3D-Res-UNet	Conv3D, $k_{1,2}$	3
	Conv3D, k_3	16
	Learning rate	5.5×10^{-4}
	Dropout rate	0.2
	Lambda	0.1
3D-UNet	Conv3D, $k_{1,2}$	1
	Conv3D, k_3	16
	Learning rate	1.1×10^{-4}
	Dropout rate	0.1
	Lambda	0.9
UNet	Learning rate	7.4×10^{-5}
	Dropout rate	0.2
	Lambda	0.8
PP-UNet	Learning rate	7.4×10^{-4}
	Dropout rate	0.1
	Lambda	0.2

TABLE II

Quantitative Performance Metrics

Network	CF	SSIM	PSNR (dB)	NMSE	Activation Map MAE
3D-Res-UNet (reconstruction)	75%	0.9166 ± 0.0174	30.2851 ± 1.5516	0.0399 ± 0.0222	0.1019
	80%	0.9041 ± 0.0192	29.4798 ± 1.5038	0.0448 ± 0.0278	0.1038
	85%	0.8915 ± 0.0211	28.8217 ± 1.4282	0.0506 ± 0.0224	0.1193
	90%	0.8619 ± 0.0264	27.8481 ± 1.4058	0.0701 ± 0.0472	0.1274
	95%	0.8154 ± 0.0353	26.7270 ± 1.2469	0.1106 ± 0.0364	0.1724
3D-UNet (reconstruction)	75%	0.9046 ± 0.0189	29.5322 ± 1.4163	0.0647 ± 0.0259	0.1161
	80%	0.8981 ± 0.0201	29.4254 ± 1.2746	0.0652 ± 0.0335	0.1278
	85%	0.8874 ± 0.0202	29.0893 ± 1.2615	0.0728 ± 0.0308	0.1264
	90%	0.86 ± 0.0238	28.2881 ± 1.2772	0.0913 ± 0.0464	0.1597
	95%	0.8139 ± 0.0373	27.1319 ± 0.7524	0.1177 ± 0.0495	0.1858
UNet (reconstruction)	75%	0.8394 ± 0.0342	27.6412 ± 0.7867	0.1222 ± 0.0488	0.1362
	80%	0.8348 ± 0.0323	27.2672 ± 0.7142	0.1375 ± 0.0456	0.1517
	85%	0.8384 ± 0.0325	27.3371 ± 0.8499	0.1399 ± 0.0385	0.1355
	90%	0.8237 ± 0.0337	26.7846 ± 0.8862	0.1398 ± 0.0467	0.1376
	95%	0.7927 ± 0.0444	26.1352 ± 0.8169	0.1946 ± 0.0550	0.1674
PP-UNet (post-processing)	75%	0.9269 ± 0.0153	31.1037 ± 1.2606	0.0359 ± 0.0239	0.1020
	80%	0.9165 ± 0.0179	30.6757 ± 1.4619	0.0396 ± 0.0224	0.1100
	85%	0.902 ± 0.0212	30.2345 ± 1.1474	0.0459 ± 0.0242	0.1245
	90%	0.876 ± 0.0233	29.1286 ± 1.3275	0.061 ± 0.0323	0.1499
	95%	0.8226 ± 0.0376	27.3628 ± 0.9142	0.102 ± 0.0537	0.1691
Conventional	75%	0.6955 ± 0.0362	21.3873 ± 0.4584	0.2726 ± 0.0208	0.1602
	80%	0.6799 ± 0.0387	21.0242 ± 0.4763	0.2793 ± 0.0202	0.1808
	85%	0.6553 ± 0.0438	20.4423 ± 0.5235	0.2938 ± 0.0303	0.2061
	90%	0.6142 ± 0.0497	19.5135 ± 0.6068	0.3193 ± 0.0349	0.2440
	95%	0.5181 ± 0.0585	17.4375 ± 0.6265	0.3947 ± 0.0652	0.2942

All the metrics for the different models were calculated versus the respective reference images. CF: compression factor; SSIM: structural similarity Index metric; PSNR: peak signal-to-noise ratio; NMSE: normalized mean squared error; MAE: mean absolute error.

TABLE III

Deep-fUS Networks Parameters and Processing Times

Network		3D-Res-UNet	3D-UNet	UNet	PP-UNet	Conventional
N. Layers		5 + 1	5 + 1	5	5	-
N. trainable parameters		9,788,421	8,773,701	8,665,633	8,629,921	-
Training epochs		1500	1500	2500	500	-
Reconstruction time (ms/image)	CF 75%	13.5	7.12	5.28	2.06 + 255.7	255.7
	CF 80%	11.1	5.67	4.6	2.04 + 220.3	220.3
	CF 85%	9.2	4.77	3.88	2.09 + 207.2	207.2
	CF 90%	6.7	3.89	3.34	2.08 + 194.8	194.8
	CF 95%	4.4	2.95	2.66	2.04 + 183.4	183.4