Deep-fUS: functional ultrasound imaging of the brain using deep learning and sparse data

Tommaso Di Ianni[*], Raag D. Airan

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

[*]: Correspondence should be addressed to T.D.I. (todiian@stanford.edu)

Abstract

Functional ultrasound imaging is rapidly establishing itself as a state-of-the-art neuroimaging modality owing to its ability to image neural activation in awake and mobile small animals, its relatively low cost, and its unequaled portability. To achieve high blood flow sensitivity in the brain microvasculature, functional ultrasound relies on long sequences of ultrasound data acquired at high frame rates, which pose high demands on the hardware architecture and effectively limit the practical utility and clinical translation of this imaging modality. Here we propose Deep-fUS, a deep learning approach that aims to significantly reduce the amount of ultrasound data necessary, while retaining the imaging performance. We trained a neural network to learn the power Doppler reconstruction function from sparse sequences of compound ultrasound data, using ground truth images from high-quality *in vivo* acquisitions in rats, and with a custom loss function. The network produces highly accurate images with restored sensitivity in the smaller blood vessels even when using heavily under-sampled data. We tested the network performance in a task-evoked functional neuroimaging application, demonstrating that time series of power Doppler images can be reconstructed with adequate accuracy to compute functional activity maps. Notably, the network reduces the occurrence of motion artifacts in awake functional ultrasound imaging experiments. The proposed platform can facilitate the development of this neuroimaging modality in any setting where dedicated hardware is not available or even in clinical scanners, opening the way to new potential applications based on this technology. To our knowledge, this is the first attempt to implement a convolutional neural network approach for power Doppler reconstruction from sparse ultrasonic datasets.

Introduction

Functional ultrasound (fUS) has recently emerged as an innovative neuroimaging modality capable of creating brain-wide neural activity maps at micrometer- and millisecond-scale resolution by tracking temporal changes in cerebral blood volume in the brain microvasculature[1]. Similar to blood oxygen level dependent functional magnetic resonance imaging (BOLD fMRI), the detected fUS hemodynamic response provides an indirect measurement of local spiking activity via neurovascular coupling[2]. However, fUS has higher sensitivity and spatiotemporal resolution than fMRI and uses more affordable and portable equipment, opening the possibility for functional neuroimaging performed directly at the bedside[3–5]. Preclinically, fUS enables imaging of neural activity in awake and freely behaving small animals, eliminating the confounds introduced by anesthesia/sedation or physical restraint in behavioral experiments[6,7]. In addition, fUS imaging has proven useful for imaging resting state and task-evoked functional connectivity in the rat and mouse brain[2,8,9], for visualizing vascular dynamics during spontaneous rapid-eye-movement sleep[10], and for mapping neural activation in primates during cognition tasks and visual stimulation[11,12]. In humans, fUS has been successfully used intraoperatively during brain tumor removal surgeries[4,5] and to visualize epileptic activity in neonates through the anterior fontanel window[3].

To detect the minute hemodynamic changes typical of the brain microvascular network, fUS relies on highly sensitive power Doppler sequences based on the use of plane-wave emissions. Specifically, unfocused ultrasound waves are used to 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 procedure makes it possible to continuously acquire long sequences of ultrasound data at high frame rates, to yield more efficient discrimination between the weak signals from slowly moving scatterers (i.e., red blood cells) in tiny blood vessels and the strong undesired static tissue clutter. The filtered Doppler signals are then squared and time-integrated to create power Doppler images with pixel signal amplitude proportional to the cerebral blood volume (Fig. 1a).

The effect of the sequence length on fUS sensitivity is depicted in Fig. 1a-f. With 250 compound frames, efficient clutter filtering can be obtained both in large and small vessels by using temporal and singular-value decomposition filters[7,13,14] (Fig. 1a-c). Conversely, clutter filtration becomes challenging when using shorter compound sequences (Fig. 1d-f). In particular, using sparse sequences hampers the blood flow detection sensitivity in smaller vessels where the blood-signal-to-clutter ratio is reduced and the lower blood velocities cause overlapping of the Doppler spectra relative to the tissue and blood signals (Fig. 1f). Hence, typical fUS sequences use hundreds of compound frames, posing high demands on the hardware platform in terms of data throughput, storage capacity, and computational power. For this reason, real-time fUS imaging is challenging even when processing is performed in a graphics processing unit (GPU) and often requires dedicated hardware implementations, effectively limiting the usability and clinical translation of this

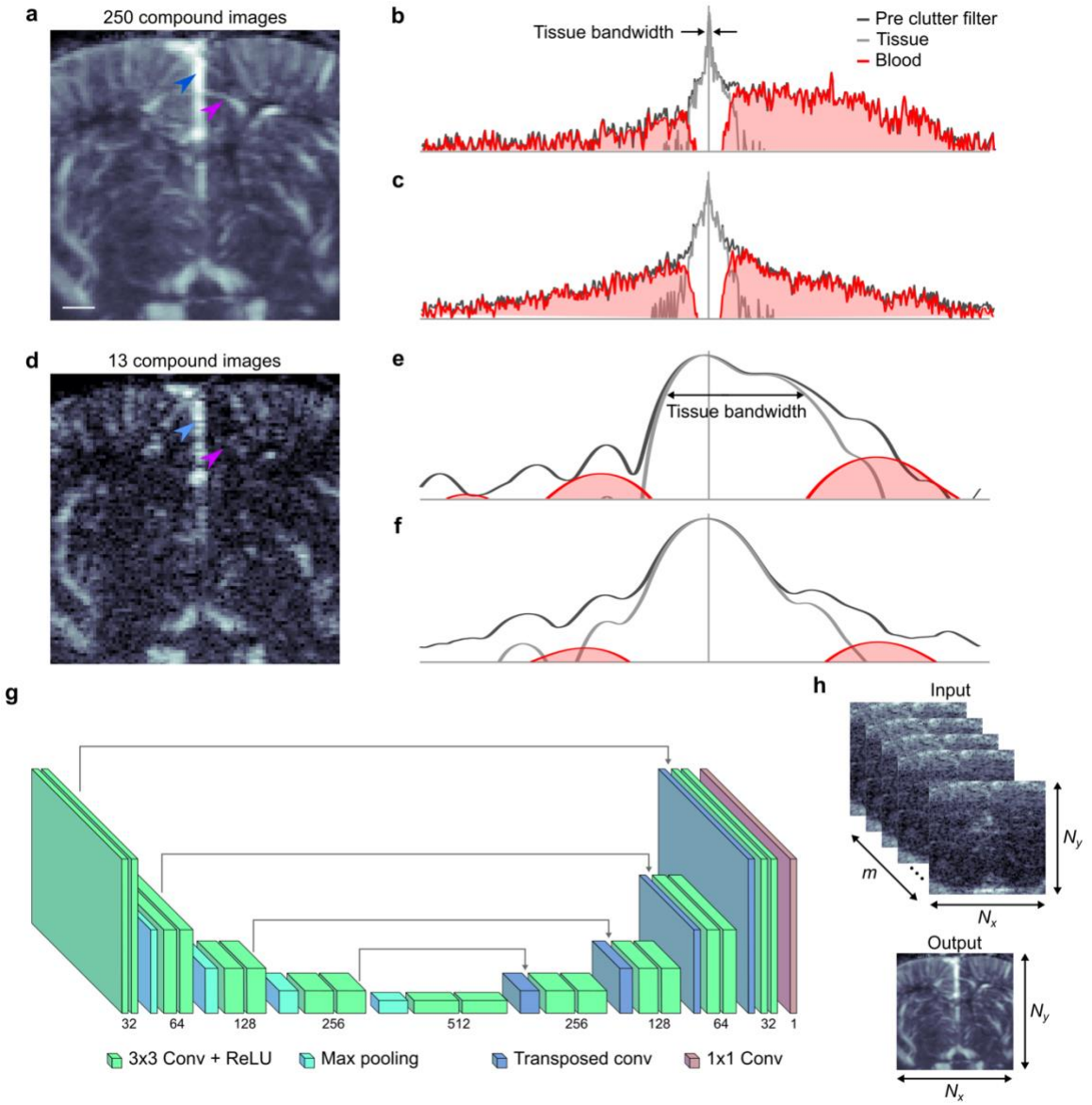

Fig. 1: Deep convolutional neural network for the reconstruction of power Doppler images from sparse compound ultrasound data. a, Coronal power Doppler image of the rat brain reconstructed with the conventional approach from 250 compound frames. Scale bar: 1 mm. **b**, **c**, Spectra of the Doppler signal in correspondence of a larger blood vessel (**b**; location pointed by the blue arrow in **a**) and a smaller blood vessel (**c**; location pointed by the purple arrow in **a**). The frequency axis is proportional to the Doppler velocity. The red shaded area is proportional to the power Doppler signal. **d**, Coronal power Doppler image of the rat brain reconstructed from 13 compound frames. **e**, **f**, Spectra of the Doppler signal in the same locations as in **b** and **c**. With a shorter sequence, tissue clutter filtration becomes difficult due to the broadening of the tissue bandwidth and the overlapping of the tissue and blood spectra. **g**, The Deep-fUS architecture uses a U-Net network consisting of an encoding and decoding branches. The network takes in input a sequence of compound ultrasound frames and outputs the power Doppler image (**h**).

imaging technology. This problem is even more relevant for recently developed four-dimensional fUS imaging sequences that aim to image volumetric cerebral blood volume changes[15,16]. In addition, long ultrasound exposure times raise concerns for the induced bioeffects, even at diagnostic levels of intensity[3,17].

Inspired by recent advances in the use of convolutional neural networks (CNNs) for the reconstruction and processing of biomedical images from sparse data[18–20], in this paper we propose Deep-fUS, an end-to-end deep learning approach to reconstruct power Doppler images from sub-optimally sampled compound datasets. Recently, CNN solutions have been proposed in medical ultrasound imaging for applications including contrast improvement[21] and image de-speckling[22], ultrasound contrast agent localization and tracking[23,24], and for undersampled and adaptive beamforming[25–28]. Here we implement a deep CNN based on the U-Net architecture[29] that learns a reconstruction mapping between the sequence of compound ultrasound data and the power Doppler output image (Figure 1g-h). Originally proposed for biomedical image segmentation, the U-Net network has been used for biomedical image reconstruction in applications spanning compressed sensing MRI[30], sparse-projection photoacoustic imaging[31,32], and sparse X-ray computed tomography[33,34]. We trained a modified U-Net on state-of-the-art power Doppler images from *in vivo* acquisitions in rats and using a custom loss function.

Results

The received sensor RF data were beamformed in a regular grid of pixels, and compound frames were created from 10 plane wave emissions. The sequence of compound frames was then processed to compute the power Doppler images (see Methods for details on the image acquisition and reconstruction). The conventional processing approach achieves a satisfactory level of detail in coronal images of the rat brain vasculature reconstructed from 250 compound frames (Fig. 2b). The resulting images were considered as ground truth in the Deep-fUS training and for evaluating the reconstruction performance. To test the processing with sub-optimal conditions, we created sparse sequences by selecting a subsample of the compound frames. We define the density index as the retained fraction of the full compound sequence in percent; for example, in the case of a density index of 10%, we retained only the first 25 compound frames in each sequence. The conventional processing produces power Doppler images that appear increasingly noisy as the sparsity increases, due to the reduced sensitivity to slower blood velocities. While larger blood vessels are still visible with a density of 25% or lower, the smaller vessels in cortical and subcortical regions vanish and become indistinguishable from the noise (Fig. 2c).

The Deep-fUS network solves a blind reconstruction function that directly extracts the power Doppler values from a collection of compound frames, without integrating any prior model-based knowledge. The modified U-Net takes in input a sparse sequence of m compound frames of size $N_x \times N_y$ and outputs the corresponding power Doppler image (see Fig. 1h-i and Methods for details). We chose to use beamformed frames instead of the sensor RF data to minimize data throughput and storage requirements. For an image of 96 × 96 beamformed pixels, for instance, the data burden is reduced by approximately 79-fold after beamforming (see Supplementary Notes). The network was trained for 20k epochs on ground truth images, and the training objective was to minimize a custom loss function defined as the weighted sum of the mean

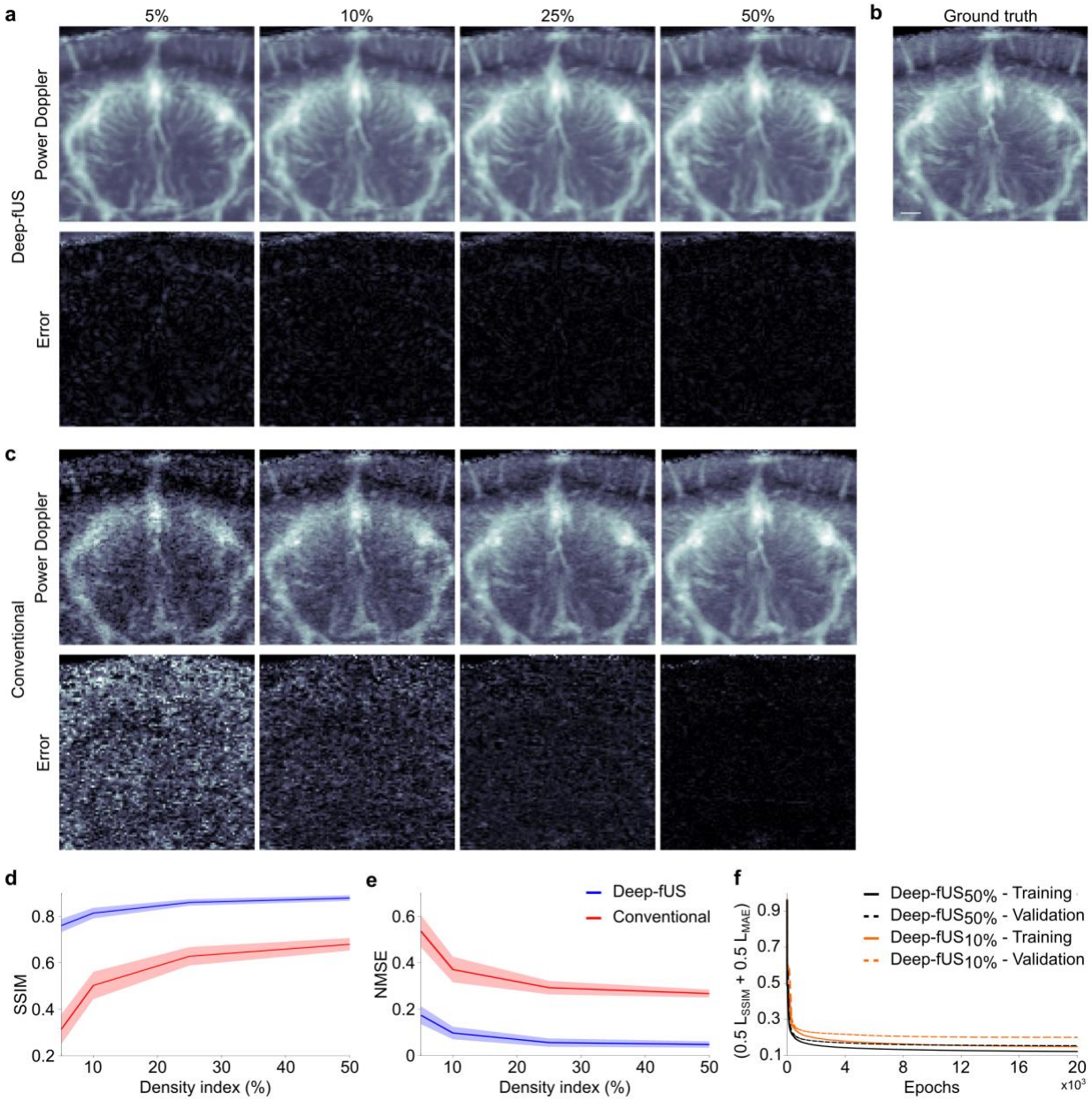

Fig. 2: Deep-fUS produces high-quality power Doppler images from sparse sequences of compound data. a, Representative power Doppler images of a coronal slice of the rat brain reconstructed by Deep-fUS with density indices of 5%, 10%, 25%, and 50% (Top) and error images calculated as a difference with the ground truth (Bottom). **b**, Ground truth image reconstructed with the conventional approach using 250 compound frames. **c**, Power Doppler images reconstructed with the conventional approach (Top) and respective error images (Bottom). **d**, **e**, Structural similarity index metric (SSIM) and normalized mean squared error (NMSE) of the power Doppler images reconstructed by Deep-fUS (blu) and by the traditional approach (red). The quantitative metrics were calculated versus the respective ground truth images using the testing examples. Results are reported as mean (solid line) and standard deviation (shaded area). **f**, Learning curves representing the custom training and validation losses for the Deep-fUS network trained using a density index of 50% (black) and 10% (orange). L_{SSIM}: Structural dissimilarity index metric loss. L_{MAE}: Mean absolute error loss. Coronal slices acquired at bregma -7.04 mm. Power Doppler images are plotted with a dynamic range of 40 dB. Scale bar: 1 mm. Native image resolution: 0.1 x 0.1 mm^2 in plane, 0.5 mm through plane, 250 ms compound frames.

absolute error between the predicted Deep-fUS image and the respective ground truth (L_{MAE}) and a structural dissimilarity index metric loss (L_{SSIM}). This loss function provides the best reconstruction accuracy

as determined by quantitative metrics (see Supplementary Table 1). Hyperparameters were heuristically optimized to provide smooth training and validation losses (see Fig. 2f and Supplementary Fig. 1) and to maximize quantitative evaluation metrics. Specifically, we based our optimization on a structural similarity index metric (SSIM), normalized mean squared error (NMSE), and peak signal-to-noise ratio (PSNR) between the predicted images and the respective ground truth. We used a total of 58 high-quality power Doppler datasets from *in vivo* acquisitions of coronal slices of the rat brain recorded between bregma +2.7 mm and bregma -7.04 mm[35]. Each acquisition lasted a maximum of 50 min to capture different hemodynamic states, levels of anesthesia, and heating/noise level of the scanner. We used 50 datasets for training and 4 for validating the network, randomly selecting up to 5 frames per dataset and using data augmentation, for a total of 740 training examples and 40 validation examples. To test the reconstruction performance of our network, we used the entire set of images from 4 acquisitions (1400 examples in total; see Methods for details on the network implementation and training).

The results of the Deep-fUS reconstruction with sparse sequences of density index 5% (m = 13), 10% (m = 25), 25% (m = 63), and 50% (m = 125) are shown in Fig. 2a. The Deep-fUS network appears to restore the reference blood flow sensitivity and is able to visualize the rat brain microvasculature when the reconstruction is performed from as few as 5% of the original compound frames. These results can be appreciated also by comparing the error images for the Deep-fUS (Fig. 2a, bottom) versus the conventional processing approach (Fig. 2c, bottom). Fig. 2d and 2e report the SSIM and NMSE calculated from the testing frames reconstructed with different density indices. The quantitative metrics confirm the considerably increased sensitivity of the Deep-fUS reconstruction when sparse compound sequences are used, reaching SSIM values in excess of 0.85 and NMSE equal to 0.05 with 25% of the original compound data. For a side-by-side comparison of the conventional and Deep-fUS approaches, four videos showing power Doppler images reconstructed by the two methods with different density indices are available in Supplementary Video 1-4. The mean prediction time for our network is 15.2 ms/frame with 50% density, 10.2 ms/frame with 25% density, 6.6 ms/frame with 10% density, and 5.8 ms/frame with 5% density.

We then sought to evaluate whether our network is able to provide accurate reconstruction of time series of power Doppler images in a functional neuroimaging application. Specifically, we used Deep-fUS to image visual-task-evoked brain activation in a rat exposed to binocular green light stimulation (see Fig. 3a and Methods for details on the stimulation protocol). We created activation maps by performing pixel-wise correlations between the power Doppler signals and the temporal stimulus pattern (Fig. 3b). A statistical threshold was applied to create binary masks showing only the significant pixels (see Methods). The ground truth activation map is shown for reference in Fig. 3c; this was created using power Doppler images reconstructed by the conventional approach from 250 compound frames and shows a significant activation in the primary visual cortex, with minor amounts of resolvable activation near the superior colliculus. Fig. 3d reports the activation maps created by using power Doppler time series reconstructed by Deep-fUS with

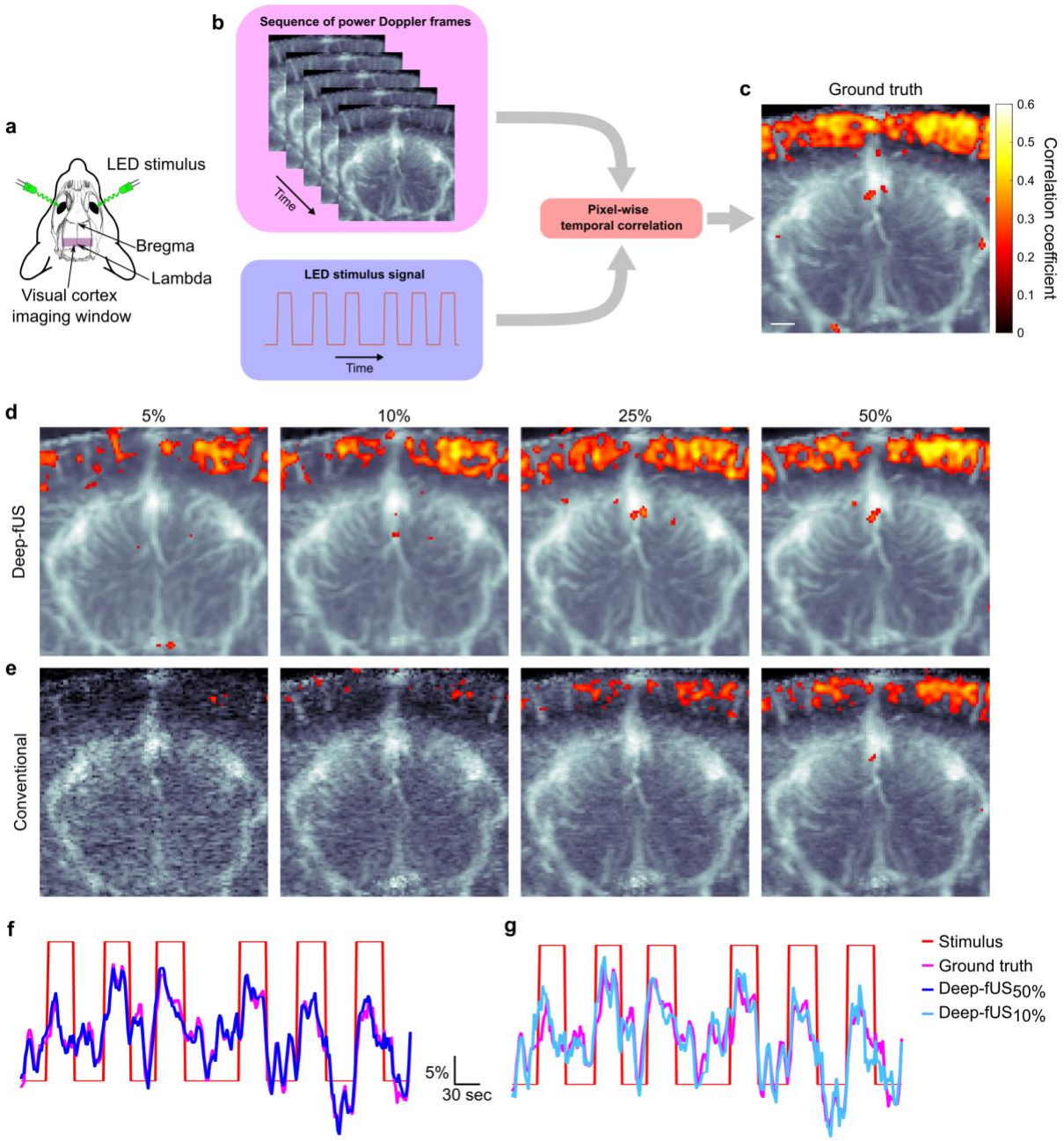

Fig. 3: Deep-fUS creates activation maps of visual evoked brain activation with sparse data. a, A visual stimulation protocol was implemented to test the Deep-fUS reconstruction performance in a functional neuroimaging application. Green LED light was used bilaterally. Ultrasound data were recorded continuously during the stimulation at bregma -7.04 mm. **b**, Time series of power Doppler images were correlated with the temporal stimulus pattern. The pattern consisted of 6 light stimuli, each with an on time of 30 s. The stimuli were distributed in a pseudo-random fashion. **c**, Ground truth activation map computed using power Doppler images reconstructed with the conventional approach from 250 compound frames. The map shows significant activation of the rat primary visual cortex. **d**, Activation maps computed using power Doppler images reconstructed by the Deep-fUS network with sparse compound sequences with 5%, 10%, 25%, and 50% density. **e**, Activation maps computed using power Doppler images reconstructed by the conventional approach with sparse compound sequences. **f**, **g**, Power Doppler signals in the activated regions from the ground truth and Deep-fUS activation maps with density index of 10% and 50%. The signals (proportional to cerebral blood volume) are shown as a variation relative to baseline, and present increases between 10% and 20% in response to the 30-s visual stimulus. Scale bar: 1 mm.

sparse compound sequences of density 5%, 10%, 25%, and 50%. Although the quality of the activation

maps appears to decrease with increasing data sparsity, significant bilateral visual cortex activation can be detected with a density index as low as 25%. Activation in the superior colliculus is also reproduced with 25% density index for the Deep-fUS processed images. On the contrary, when the reconstruction is performed with the conventional approach, it is difficult to detect any significant brain activation even with 50% density (Fig. 3e). The non-thresholded correlation maps computed using the Deep-fUS and conventional power Doppler reconstructions are available in Supplementary Fig. 2. In Fig. 3e-f we report the power Doppler signals in the activated brain regions of the ground truth activation map and Deep-fUS maps with density index of 10% and 50%. The signals show increases between 10% and 20% in response to the visual stimulus, which reflect increases in cerebral blood volume in the primary visual cortex due to neurovascular coupling.

To determine whether short acquisition sequences can help prevent the occurrence of motion artifacts, we used Deep-fUS to reconstruct a power Doppler time series from an awake rat fUS imaging experiment. We computed the SSIM of each image in the sequence versus the first image, then applied a threshold to

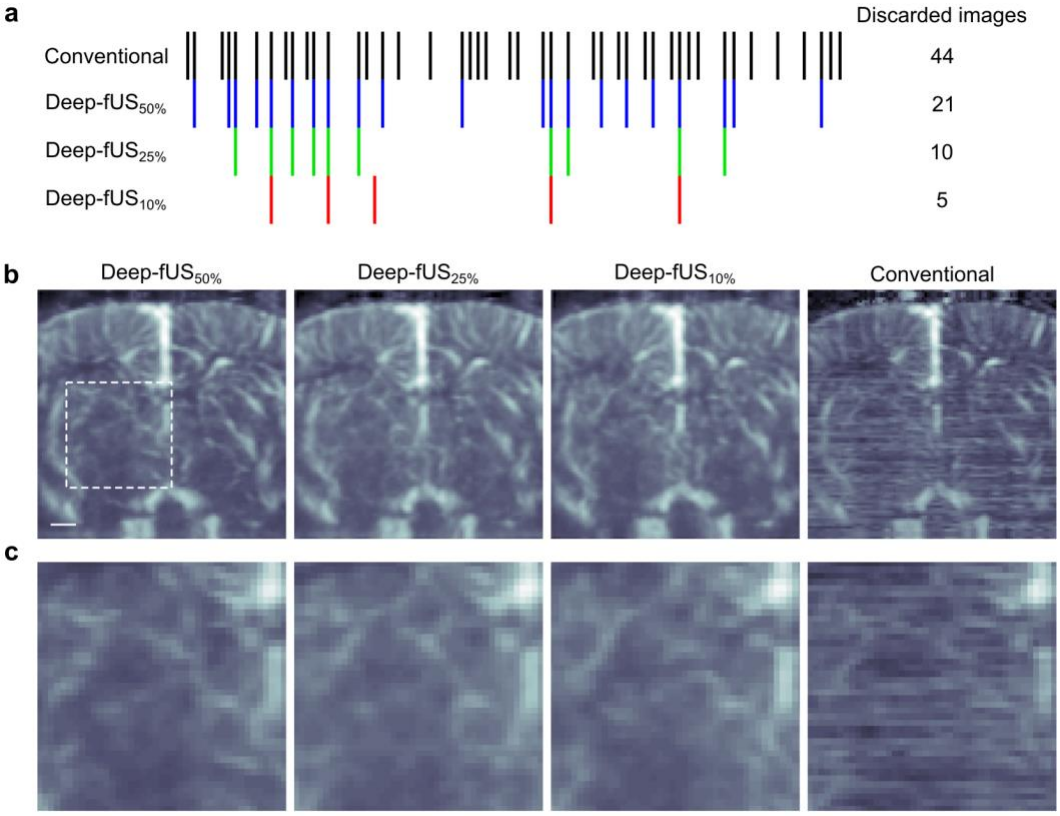

Figure 4: Sparse sequences reduce the occurrence of motion artifacts in awake functional ultrasound imaging. a, Motion-affected images in a time series of 500 power Doppler frames were filtered based on a structural similarity index metric threshold. Each vertical bar in the plot shows a discarded image in the sequence. The total number of discarded images for each reconstruction scenario is reported on the right. **b**, Power Doppler coronal image showing a case of significant motion artifacts in the conventional reconstruction that were completely resolved by using sparse sequences and Deep-fUS processing. Coronal slices were acquired at bregma +2.7 mm. Scale bar: 1 mm. **c**, Magnification of the region within the white dashed square in **b**.

discard images that showed significant motion artifacts (see Supplementary Fig. 3). Fig. 4a reports the discarded images for the sequence reconstructed with the conventional approach and by our neural network with a density index of 10%, 25%, and 50%. As expected, the number of discarded images decreases with increasing sparsity with an approximately linear relationship. In the power Doppler images shown in Fig. 4b, motion artifacts appear to be completely resolved in all the cases using sparse compound data. This result can be also appreciated by comparing the magnified power Doppler images in Fig. 4c. We provide a side-by-side comparison of the conventional and Deep-fUS reconstruction in the awake fUS experiment in Supplementary Video 5.

Discussion

Deep learning is steadily revolutionizing modern medicine and the medical imaging field as a whole[36], and medical ultrasound represents no exception. In the last few years, deep CNNs have been the object of increasing attention in anatomical ultrasound image reconstruction, with a particular emphasis on models that aim to restore high-end image quality while reducing data sampling, transmission, and processing[25,26,28,37]. With the exception of a single study using deep learning of color Doppler images[38], however, CNNs have not been applied as extensively to blood flow imaging, and to the best of our knowledge this is the first attempt to implement an end-to-end network to perform power Doppler reconstruction from sparse ultrasonic datasets (Fig. 1). The proposed Deep-fUS platform provides outstanding sensitivity and, in all cases, outperforms the conventional approach with sparse compound sequences, as clearly demonstrated by the presented quantitative metrics (Fig. 2). We expect the high quality of the *in vivo* rat brain images used during training to be the main contributing factor to the Deep-fUS reconstruction performance[31]. We demonstrate that our network can accurately predict time series of power Doppler images with heavily under-sampled data and can be used to compute functional activation maps in a task-evoked neuroimaging application (Fig. 3). Additionally, we show that the interpolation fidelity our network allows greater robustness to motion artifacts and the data scrubbing needed to account for them in an awake rat fUS imaging experiment (Fig. 4).

The main advantage of using sparse sequences is the net reduction in processing time and computational capacity demands. The suggested approach can facilitate the development of fUS neuroimaging in any setting where dedicated hardware is not available or even in clinical scanners, making this technology more affordable and opening the way to new potential applications based on this imaging modality[3,39]. Additionally, sparse sequences can prove beneficial in experimental situations where fUS acquisitions need to be interleaved with therapeutic ultrasound pulses, such as in the monitoring of focused ultrasound neurointerventions[40]. Importantly, this method significantly reduces the exposure time and lowers the risk of harmful bioeffects, making brain ultrasound imaging safer[3,17]. The network introduces a small processing time overhead while considerably reducing the beamforming time and eliminating the need for computationally demanding filters[13,14]; therefore, Deep-fUS simplifies real-time fUS implementations[2] and

has an additional potential to increase the imaging frame rate. Although high frame rates are not a primary concern in neuroimaging applications, due to the slow transfer function between neuronal and vascular responses[41], the proposed platform and conceptual framework can be readily adapted to other high-frame-rate Doppler ultrasound imaging modalities, including vector flow imaging, to expedite their deployment in mobile-device-based ultrasound systems[42,43]. In creating sparse sequences, we chose to select only the initial portion of the original sequence over zeroing interleaved frames (e.g., discarding 3 frames every 4 in the case of 25% density) to take advantage of the shorter observation window. This approach has the benefit of reducing the occurrence of motion artifacts and signal degradation due to data scrubbing, which are inevitable factors in mobile rodent fUS imaging experiments[7,44] and handheld applications[5] and are more likely to appear with longer observation times.

A main limitation of using ultrasound for brain imaging is the presence of the skull, which is highly absorbing, in particular at the higher imaging frequencies used for fUS. This has limited clinical fUS imaging to intraoperative applications or to scenarios with natural skull openings, such as the neonatal anterior fontanel window. This limitation can be partly overcome by using contrast agents that enhance the ultrasound signal[45]. While this paper is a first demonstration of the ability to learn power Doppler reconstruction functions from sparse data, we expect that using more advanced networks and training strategies can bring an additional boost in image quality with further reduced data density, to potentially overcome such current challenges.

Methods

Deep-fUS network and training

We modified a U-Net network[29] and trained it to perform the power Doppler reconstruction task. This fully convolutional neural network is based on an encoder (contracting) path followed by a symmetrically sized decoder (expanding) path. The encoder path progressively down-samples the input data and learns high-level features that are propagated to the following stages. The decoder path 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 to retain context information, which is propagated to the symmetric up-sampling layers. This network was specifically proposed for applications where little training data are available, such as biomedical imaging applications. As schematically illustrated in Fig. 1g, in our model we used a 5-layer encoder followed by a 4-layer decoder. Encoder layers are composed of 2 consecutive convolutional blocks implementing 3×3 convolutions, each followed by rectified linear unit (ReLU) activations. The second block is cascaded with a 2×2 max pooling operator that halves the resolution in both the image dimensions. Layers in the decoder part are composed of 2 convolutional blocks (3×3) followed by ReLU activations. The subsequent up-sampling function is performed by a 2×2 transposed convolution block with ReLU activations. The number of channels is progressively increased in the encoder path (32, 64, 128, 256, and 256 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. All the convolutional kernels are initialized using the He initialization[46]. We used a dropout probability of 50% on the convolutional blocks of encoder layer 5 and on all the decoder layers. This configuration provided the best reconstruction accuracy as determined by quantitative metrics (see Supplementary Table 2). In total, the network has 23 convolutional layers and 8,665,633 trainable parameters.

We trained the network to learn a function $y = f(x)$ that maps the input sequence x_i of m compound frames of $N_x \times N_y$ pixels to the output power Doppler image y_i of dimensions $N_x \times N_y$ (Fig. 1h). For all the experiments, we considered images of $N_x = N_y = 96$ pixels. Only the real part of the beamformed data was used to avoid GPU memory overflow. The input data were normalized, and the output data standardized. Ground truth images y_i were obtained from *in vivo* acquisitions reconstructed using state of the art power Doppler processing (see “Power Doppler processing” section below). To improve the network regularization, we selected up to 5 crops from each training image, and a random horizontal flipping was applied with a probability of 50%. We did not consider including vertical flipping due to the nature of the ultrasound data, which see an exponentially decaying signal-to-noise ratio caused by the depth-dependent acoustic attenuation. Therefore, inverting this curve would give artifactual results that do not represent a realistic situation. Moreover, it is worth noting that the exact vertical placement of the transducer with respect to the brain varied slightly at each acquisition, and we used two ultrasound imaging probes with

different elevation foci (see “Ultrasound system and data acquisition” section below). The elevation foci change the beam shape in the elevation plane, therefore adding an additional source of randomness that is expected to help regularize the network.

At each iteration, the network predicts a new estimate $\hat{y}_i$, and the parameters are updated via stochastic gradient descent to minimize the loss function

$$L(y, \hat{y}) = \frac{1}{2}\left(L_{MAE}(y, \hat{y}) + L_{SSIM}(y, \hat{y})\right)$$

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

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

with $||\cdot||_1$ denoting the l_1 norm, N the number of image pixels, and n the number of training 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[47,48]. A kernel of 3×3 pixels was used for the SSIM computation. We trained the network using the Adam optimizer[49] with a learning rate of 0.000005 and $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-7}$. The mini-batch size was set to 1. We heuristically optimized the training hyperparameters to provide smooth losses (see Fig. 2f and Supplementary Fig. 1) and based our optimization on the SSIM of the predicted validation images versus the respective ground truth, on 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 our network in Python using TensorFlow 2.1 with Keras API. The network was trained for 20k epochs with a single NVIDIA Titan RTX GPU with 24 GB of RAM. Training each epoch took approximately 12.5 ms.

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. Each compound frame was created by beamforming the received sensor RF data in a regular grid of pixels of 100 × 100 μm^2 and by successively compounding the beamformed data from tilted plane waves at angle (-6°, -3°, 0°, 3°, 6°) emitted with a pulse repetition frequency of 19 kHz. Two 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 250 frames and at a frame rate of 1 kHz in an observation time window of 250 ms. The acquired data from a full sequence were transferred to the host computer and beamformed in an NVIDIA Titan RTX GPU using

a GPU beamformer[50]. The compound frames were then saved on the host machine for offline processing. The 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 final power Doppler frame rate was 0.6 frames/sec.

Power Doppler processing

Sequences of *m* compound ultrasound frames were processed in Matlab (MathWorks) 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 a singular value decomposition filter that eliminates the first singular value[7]. 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 a temporal high-pass filter. Singular value decomposition filters are instead based on the assumption that, while blood signals are highly incoherent due to the time-varying stochastic distribution of the moving scatterers (red blood cells), tissue signals maintain a high degree of correlation over time, and therefore aim to eliminate the highly coherent components. At each pixel location (x, y), the intensity of the filtered signal $s(x, y, t)$ was then calculated to find the power Doppler value $I(x, y) = \int s^2(x, y, t)\, dt$. For the ground truth data (m = 250), the entire time window of 250 ms was integrated. For the sparse compound sequences, integration windows were 125 ms (50%), 63 ms (25%), 25 ms (10%), and 13 ms (5%).

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. A total of N = 20 male Long Evans and Sprague Dawley rats (Charles River) were used in this study. We prepared the animals by performing a bilateral surgical craniotomy and chronic prosthesis implant using previously published protocols[6]. Briefly, animals were anesthetized with 3.5% isoflurane and maintained with 1.5% isoflurane. Animals 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 anti-inflammatory 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, and sealed in place using dental cement (Tetric EvoFlow) a plastic polymethylpentene prosthesis covering the entire craniotomy. The space between the dura mater and the polymer film was filled with 0.9% sterile saline. Animals were then allowed to recover for 1 week before the first imaging session.

During the data acquisition experiments, animals were either anesthetized and kept under anesthesia with 1.5% isoflurane while placed in a stereotaxic frame (41 acquisitions) or were lightly sedated with 0.5% isoflurane and kept in a restraining apparatus that prevented head movement[51]. The latter protocol was also used in the awake fUS experiment of Fig. 4. In the visual-task-evoked experiment, 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. Bilateral visual stimuli were delivered using two green light LEDs (Fig. 3a) driven by a custom power supply circuit. We controlled the stimulus pattern through a microcontroller board (Arduino Uno) connected to Matlab via the serial port and interfaced with the Verasonics platform 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. This stimulation protocol was shown to maximize visual cortex response in prior fUS imaging studies[52]. Note that the dataset acquired during the visual stimulation was also one of the 4 sets used for testing the reconstruction performance of the network.

Functional activation maps

Functional activity maps were created by calculating the Pearson's correlation coefficient r between the power Doppler temporal signal and the stimulus pattern[1,7]. We then used a Fisher's transformation to calculate the z score as

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

with N the number of temporal samples. Each pixel was considered significant for $z > 3.1$, which corresponds to $P < 0.001$ in a one-tailed t-test. We used this threshold to create binary activation maps that show only the significant pixels.

Data availability

The code to reproduce the results of the paper is available at https://github.com/todiian/deep-fus.

Acknowledgements

This work was supported by a Seed Grant from the Stanford Wu Tsai Neurosciences Institute and the NIH BRAIN Initiative (NIH/NIMH RF1MH114252) and HEAL Initiative (NIH/NINDS UG3NS115637). TDI was supported by a Stanford University School of Medicine Dean's Postdoctoral Fellowship. We would like to thank Jeremy Dahl, PhD for support with the ultrasound equipment and Dongwoon Hyun, PhD for technical assistance with the GPU beamformer. We would also like to thank Sam Baker, DVM, and Madeline Hughes Cooper for assistance with the surgical protocols. We would also like to thank the Airan Lab for helpful discussions.

Author Contribution

TDI designed and implemented the functional ultrasound imaging sequences and the computational approach for data analysis. RDA provided contextual advising, the hardware for data acquisition, and secured funding support. All authors together drafted and reviewed the figures and manuscript.

Competing Interests

Patent applications have been filed on the technologies described in this manuscript (63/084816 – Provisional application with Stanford University; T.D.I. and R.D.A.).

References

1. Macé, E. *et al.* Functional ultrasound imaging of the brain. *Nat. Methods* **8**, 662–664 (2011).
2. Macé, É. *et al.* Whole-Brain Functional Ultrasound Imaging Reveals Brain Modules for Visuomotor Integration. *Neuron* **100**, 1241–1251 (2018).
3. Demené, C. *et al.* Functional ultrasound imaging of brain activity in human newborns. *Sci. Transl. Med.* **9**, (2017).
4. Imbault, M., Chauvet, D., Gennisson, J. L., Capelle, L. & Tanter, M. Intraoperative Functional Ultrasound Imaging of Human Brain Activity. *Sci. Rep.* **7**, 1–7 (2017).
5. Soloukey, S. *et al.* Functional Ultrasound (fUS) During Awake Brain Surgery: The Clinical Potential of Intra-Operative Functional and Vascular Brain Mapping. *Front. Neurosci.* **13**, 1–14 (2020).
6. Sieu, L.-A. *et al.* EEG and functional ultrasound imaging in mobile rats. *Nat. Methods* **12**, 831–834 (2015).
7. Urban, A. *et al.* Real-time imaging of brain activity in freely moving rats using functional ultrasound. *Nat. Methods* **12**, 873–878 (2015).
8. Osmanski, B.-F., Pezet, S., Ricobaraza, A., Lenkei, Z. & Tanter, M. Functional ultrasound imaging of intrinsic connectivity in the living rat brain with high spatiotemporal resolution. *Nat. Commun.* **5**, 5023 (2014).
9. Ferrier, J., Tiran, E., Deffieux, T., Tanter, M. & Lenkei, Z. Functional imaging evidence for task-induced deactivation and disconnection of a major default mode network hub in the mouse brain. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 15270–15280 (2020).
10. Bergel, A., Deffieux, T., Demené, C., Tanter, M. & Cohen, I. Local hippocampal fast gamma rhythms precede brain-wide hyperemic patterns during spontaneous rodent REM sleep. *Nat. Commun.* **9**, 5364 (2018).
11. Blaize, K. *et al.* Functional ultrasound imaging of deep visual cortex in awake nonhuman primates. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 14453–14463 (2020).
12. Dizeux, A. *et al.* Functional ultrasound imaging of the brain reveals propagation of task-related brain activity in behaving primates. *Nat. Commun.* **10**, 1–9 (2019).
13. Demené, C. *et al.* Spatiotemporal Clutter Filtering of Ultrafast Ultrasound Data Highly Increases Doppler and fUltrasound Sensitivity. *IEEE Trans. Med. Imaging* **34**, 2271–2285 (2015).
14. Baranger, J. *et al.* Adaptive Spatiotemporal SVD Clutter Filtering for Ultrafast Doppler Imaging Using Similarity of Spatial Singular Vectors. *IEEE Trans. Med. Imaging* **37**, 1574–1586 (2018).
15. Rabut, C. *et al.* 4D functional ultrasound imaging of whole-brain activity in rodents. *Nat. Methods* **16**, 994–997 (2019).
16. Sauvage, J. *et al.* 4D Functional Imaging of the Rat Brain Using a Large Aperture Row-Column Array. *IEEE Trans. Med. Imaging* **39**, 1884–1893 (2020).
17. Ang, E. S. B. C., Gluncic, V., Duque, A., Schafer, M. E. & Rakic, P. Prenatal exposure to

ultrasound waves impacts neuronal migration in mice. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 12903–12910 (2006).

18. Zhu, B., Liu, J. Z., Cauley, S. F., Rosen, B. R. & Rosen, M. S. Image reconstruction by domain-transform manifold learning. *Nature* **555**, 487–492 (2018).
19. Vishnevskiy, V., Walheim, J. & Kozerke, S. Deep variational network for rapid 4D flow MRI reconstruction. *Nat. Mach. Intell.* **2**, 228–235 (2020).
20. Shen, L., Zhao, W. & Xing, L. Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning. *Nat. Biomed. Eng.* **3**, 880–888 (2019).
21. Luchies, A. C. & Byram, B. C. Deep Neural Networks for Ultrasound Beamforming. *IEEE Trans. Med. Imaging* **37**, 2010–2021 (2018).
22. Hyun, D., Brickson, L. L., Looby, K. T. & Dahl, J. J. Beamforming and speckle reduction using neural networks. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **66**, 898–910 (2019).
23. Hyun, D. *et al.* Nondestructive Detection of Targeted Microbubbles Using Dual-Mode Data and Deep Learning for Real-Time Ultrasound Molecular Imaging. *IEEE Trans. Med. Imaging* 1–11 (2020). doi:10.1109/TMI.2020.2986762
24. Youn, J. *et al.* Detection and Localization of Ultrasound Scatterers Using Convolutional Neural Networks. *IEEE Trans. Med. Imag.* (2020). doi:10.1109/TMI.2020.3006445
25. Gasse, M. *et al.* High-quality plane wave compounding using convolutional neural networks. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **64**, 1637–1639 (2017).
26. Yoon, Y. H., Khan, S., Huh, J. & Ye, J. C. Efficient B-Mode Ultrasound Image Reconstruction From Sub-Sampled RF Data Using Deep Learning. *IEEE Trans. Med. Imaging* **38**, 325–336 (2019).
27. Luijten, B. *et al.* Adaptive Ultrasound Beamforming using Deep Learning. *IEEE Trans. Med. Imag.* 1–12 (2019). doi:10.1109/TMI.2020.3008537
28. Nair, A. A., Washington, K. N., Tran, T. D., Reiter, A. & Bell, M. A. L. Deep learning to obtain simultaneous image and segmentation outputs from a single input of raw ultrasound channel data. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* (2020). doi:10.1109/tuffc.2020.2993779
29. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **9351**, 234–241 (2015).
30. Yang, G. *et al.* DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction. *IEEE Trans. Med. Imaging* **37**, 1310–1321 (2018).
31. Davoudi, N., Deán-Ben, X. L. & Razansky, D. Deep learning optoacoustic tomography with sparse data. *Nat. Mach. Intell.* **1**, 453–460 (2019).
32. Guan, S., Khan, A. A., Sikdar, S. & Chitnis, P. V. Fully Dense UNet for 2-D Sparse Photoacoustic Tomography Artifact Removal. *IEEE J. Biomed. Heal. Informatics* **24**, 568–576 (2020).

33. Jin, K. H., Mccann, M. T., Froustey, E. & Unser, M. Deep Convolutional Neural Network for Inverse Problems in Imaging. *IEEE Trans. Image Process.* **26**, 4509–4522 (2017).

34. Han, Y. & Ye, J. C. Framing U-Net via Deep Convolutional Framelets: Application to Sparse-View CT. *IEEE Trans. Med. Imaging* **37**, 1418–1429 (2018).

35. Paxinos, G. & Watson, C. *The Rat Brain in Stereotaxic Coordinates: Hard Cover Edition*. (Academic Press, 1998).

36. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44–56 (2019).

37. Wiacek, A., Gonzalez, E. & Bell, M. A. L. CohereNet: A Deep Learning Architecture for Ultrasound Spatial Correlation Estimation and Coherence-Based Beamforming. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* (2020). doi:10.1109/tuffc.2020.2982848

38. Huijben, I. A. M., Veeling, B. S., Janse, K., Mischi, M. & van Sloun, R. J. G. Learning Sub-Sampling and Signal Recovery with Applications in Ultrasound Imaging. *IEEE Trans. Med. Imag.* **0062**, 1–13 (2019).

39. Fishell, A. K. *et al.* Portable, field-based neuroimaging using high-density diffuse optical tomography. *Neuroimage* **215**, (2020).

40. Wang, J. B. *et al.* Focused Ultrasound for Noninvasive , Focal Pharmacologic Neurointervention. *Front. Neurosci.* **14**, (2020).

41. Aydin, A. K. *et al.* Transfer functions linking neural calcium to single voxel functional ultrasound signal. *Nat. Commun.* **11**, 1–10 (2020).

42. Di Ianni, T. *et al.* A Vector Flow Imaging Method for Portable Ultrasound Using Synthetic Aperture Sequential Beamforming. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **64**, (2017).

43. Di Ianni, T., Hemmsen, M. C., Muntal, P. L., Jorgensen, I. H. H. & Jensen, J. A. System-Level Design of an Integrated Receiver Front End for a Wireless Ultrasound Probe. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **63**, (2016).

44. 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* **222**, 117231 (2020).

45. Errico, C. *et al.* Transcranial functional ultrasound imaging of the brain using microbubble-enhanced ultrasensitive Doppler. *Neuroimage* **124**, 752–761 (2016).

46. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 770–778 (2016). doi:10.1109/CVPR.2016.90

47. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004).

48. Zhao, H., Gallo, O., Frosio, I. & Kautz, J. Loss Functions for Image Restoration With Neural Networks. *IEEE Trans. Comput. Imaging* **3**, 47–57 (2016).

49. Kingma, D. P. & Ba, J. L. Adam: A method for stochastic optimization. *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.* 1–15 (2015).

50. Hyun, D., Trahey, G. E. & Dahl, J. J. Real-time high-framerate in vivo cardiac SLSC imaging with a GPU-based beamformer. *2015 IEEE Int. Ultrason. Symp. IUS 2015* 1–4 (2015). doi:10.1109/ULTSYM.2015.0077

51. Stenroos, P. *et al.* Awake rat brain functional magnetic resonance imaging using standard radio frequency coils and a 3D printed restraint kit. *Front. Neurosci.* **12**, (2018).

52. Gesnik, M. *et al.* 3D functional ultrasound imaging of the cerebral visual system in rodents. *Neuroimage* **149**, 267–274 (2017).