Neural Operators for Accelerated Functional Ultrasound Imaging

Bahareh Tolooshams[1]   Lydia Lin[2]
Thierri Callier[2]   Jiayun Wang[1]   Sanvi Pal[1]   Aditi Chandrashekar[1]
Claire M. Rabut[2]   Sumner Norman[2]   Mikhail G. Shapiro[2]   Charles Liu[3]
Kamyar Azizzadenesheli[4]   Richard A. Andersen[2,5]   Anima Anandkumar[1]

[1] Division of Engineering and Applied Science, California Institute of Technology
[2] Division of Biology and Biological Engineering, California Institute of Technology
[3] Departments of Neurosurgery and Neurology, University of Southern California
[4] NVIDIA   [5] T & C Chen Brain-machine Interface Center, California Institute of Technology

Abstract

Functional ultrasound (fUS) is an emerging technique for non-invasive neuroimaging that infers neural activity by detecting changes in blood volume. fUS has found its applications in neuroscience studies with freely moving animals and brain-computer interfaces (BCIs) as it offers minimally invasive high spatiotemporal resolution and is a low-cost and portable technology compared to prior neurorecording techniques such as electrophysiology and functional magnetic resonance imaging (fMRI). However, the current classical fUS methods require a relatively large number of compounded images to successfully remove tissue clutter. This property has not only caused computational, memory, and communication complexity for fUS hardware technologies but also has resulted in an undesirable wait period to construct one brain image. The latter, particularly, has negatively impacted the use of fUS for real-time BCIs. Therefore we propose accelerated fUS through a deep learning technique called the neural operator for functional ultrasound (NO-fUS). NO-fUS tackles the technical challenges: it reduces the wait period of frame collections by 90% and the sampling rate at inference time by 50%. This extensive reduction on the number of input frames is a step toward more efficient fUS technology such as for 3D volumetric imaging fUS technology, reducing number of ultrasound pulses needed to image and, in turn, reduce potential probe heating and computational cost. Unlike conventional, data-driven deep neural architecture, NO-fUS is generalizable across experiment sessions and animals; we highlight this generalization in mouse, monkey, and human data. Finally, we demonstrate the BCI applications of NO-fUS in behavioral decoding. Specifically, our results suggest that NO-fUS not only offers high-quality images, but also preserves behavior-related information required to decode the subject's thoughts and planning.

1 Introduction

The advent of novel neurotechnologies provides scientists the opportunity to explore the brain at resolutions not seen before, which can guide future clinical discoveries and interventions for neurological diseases. Electrophysiological recording technology such as Utah arrays provide direct electrical recordings of single neuron level activity and have been used for brain computer interfaces (BCIs) [1, 2] and other applications. However, these techniques require highly invasive surgery, have a small field of recording, and are prone to probe degradation [3, 4, 5]. Other techniques such as fMRI and EEG are minimally invasive and have long been used both clinically and in research to provide a large field of view of brain activity. Even so, these techniques are constrained by their limited spatial and temporal resolution.

Functional ultrasound (fUS) [6, 7] is an up-and-coming neuroimaging technique that is able to image from outside of the dura with a high sensitivity, high spatial resolution, and large field of view, bridging the gaps between prior neurorecording techniques. Similar to fMRI, fUS leverages neurovascular coupling to infer population level neural activity from changes in cerebral blood volume (CBV) for neuroimaging. However, in contrast, fUS is relatively low cost and portable given the small size of current transducers and

probes, allowing for frequent bedside use. Additionally, fUS is not constrained by machinery and thus can be simultaneously used on freely moving subjects.

These advantages make fUS an optimal tool for minimally invasively examining brain activity both in the clinic and during neuroscience studies. Clinically, fUS has already been used to map brain activity and vasculature intraoperatively in humans for tumor removal [8] and monitor functional connectivity and epileptic activity in human neonates at bedside [9, 10]. Several prior studies have used fUS to record neural activity in behavioral tasks in freely moving rodents [11], non-human primates [12, 13], and humans [14]. Moreover, fUS has been increasingly used for brain computer interfaces (BCIs), a technology that decodes and converts brain signals into computer commands [12, 15, 13], and has the potential to power next-generation minimally invasive alternatives for BCIs.

However, fUS' potential is restricted by hardware and computational demands. To produce a single, final fUS image or "power Doppler" (DOP) image, hundreds of ultra-fast ultrasound pulses sent out as plane waves are transmitted and received, capturing backscattered signal from red blood cells and providing fast and precise measurements of changes in CBV. Data from these plane wave emissions are reconstructed or "beamformed" and coherently compounded to produce an intermediate "in phase and quadrature" image frame (IQ frame). One DOP image is finally then composed of the squared average of hundreds of IQ frames. Singular value decomposition (SVD) can then be used to filter out tissue motion signals from blood volume information [16, 17, 18, 19].

These processes require a large amount of data storage, can lead to dangerous heating of the ultrasound probe, and can cause slow wait times during Doppler image generation, limiting the capabilities and applications of fUS. Furthermore, there are currently two primary computational obstacles that constrain fUS neuroimaging and processing: a) fUS faces computational complexity and memory storage challenges due to the number of compounded plane wave images that must first be acquired and beamformed per DOP and b) SVD, used for tissue motion filtering, involves slow computations, struggling to scale for large datasets and requires domain experts to tune the filtering parameters.

To address the first challenge of computational and memory complexity, deep-learning-based approaches such as DeepfUS [20] have been proposed to achieve reasonable Doppler imaging. However, this approach reduces the spatial resolution, discards the imaginary component of the complex in-phase and quadrature (IQ) signal input, and ignores the inherent temporal continuity of fUS. Thus, DeepfUS is unable to operate on data of different dimensions across the time domain. The second challenge is partially addressed by general deep learning-based methods with their scalability using graphical processing unit (GPU) infrastructures and by pushing the domain knowledge to the training stage for the preparation of example pairs.

We propose a solution to these gaps. Neural Operator for Functional Ultrasound (NO-fUS) is a deep-learning-based framework that accelerates fUS acquisition using a limited number of time frames, thereby alleviating the computational, memory, and communication complexity of fUS technology and reducing the latency for fUS-based BCIs. NO-fUS uses a concept from neural operator learning, which learns mappings between two infinite-dimensional functions [21, 22, 23]. Neural Operators (NOs) are widely used for finding solutions to partial differential equations (PDEs) and have applications in many scientific and engineering disciplines [24], such as computational fluid dynamics [25], weather forecasting [26], seismology [27], and lithography modeling [28]. A notable property of neural operators is their parameterization, which allows for different discretizations. NO-fUS leverages this fundamental property to enable changes in hardware sampling rates in time or to accelerate temporal acquisition during inference without impacting imaging quality beyond a lower bound. NO-fUS is the first neural operator-empowered method for fUS. Unlike existing deep-learning-based fUS methods, NO-fUS utilizes the complex nature of input IQ signals and respects the inherent temporal attributes of image frames; the latter is particularly crucial for learning generalizable temporal filtering.

We apply NO-fUS to brain images collected from of mouse, monkey, and human and demonstrate NO-fUS state-of-the-art performance in achieving high-quality reconstructions of fUS data using less temporal data. While NO-fUS is trained on image frames from mice, we showcase its superior practical generalization to unseen experiment sessions and not only in new animals but also across species. Additionally, we show that NO-fUS preserves the features necessary for behavioral decoding in monkeys and humans while reducing the collection wait time-i.e., the Doppler sampling rate is reduced by 90%, improving NO-fUS-powered BCIs. Finally, we show that NO-fUS can run on both a CPU and GPU. It can be trained on a single GPU with less than 24 GB of memory in a short amount of time, and its inference is highly parallelizable on a GPU.

Overall, NO-fUS offers the following in a single package: a) a pipeline for training NO-fUS on experimental datasets, b) a fast procedure for fine-tuning NO-fUS on a single session for better precision, c) an inference framework for accelerated and efficient imaging in an online setting, d) a low-latency imaging tool that preserves decoding information needed for BCI-based decoding in monkeys and human and e) low-complexity pre-processing that improves the robustness of behavioral decoding within a BCI framework.

2 Results

Deep neural operator framework for accelerated fUS. NO-fUS (fig. 1) is a deep learning framework that offers accelerated fUS imaging from limited time frames, increasing the DOP sampling rate and hardware efficiency. Current classical methods require a high number of frames (e.g., 250 or 300) to achieve high-quality fUS imaging. NO-fUS maintains this performance using only $10 - 15\%$ of the number of frames (e.g., 32 frames). NO-fUS is GPU enabled, and its imaging process can be parallelized over multiple imaging examples.

NO-fUS consists of three phases: training, fine-tuning, and inference. During the training stage (fig. 1a), NO-fUS is trained on data from multiple mice recording sessions; this step leverages techniques from deep learning, such as data augmentation, for improved performance. We offer our trained model as a foundation model that can be used for imaging new sessions and new animals, with the option of a simple fine-tuning stage. NO-fUS can be fine-tuned (fig. 1b) using only a small amount of input data captured over a few minutes at the beginning of an hours-long experiment session. This fine-tuning stage is optional and is meant to further improve performance; we, indeed, skip this step in some of our experiments and show that NO-fUS generalizes to a new session of the same animal without a need for fine-tuning. Finally, during the inference stage, NO-fUS constructs a fUS image from limited IQ time frames in a fraction of a second.

NO-fUS combines techniques from deep learning and operator learning (fig. 1c). It consists of a U-shaped architecture [29, 30] that performs decomposed spatial and temporal filtering on the beamformed complex in-phase/quadrature (IQ) input signal. NO-fUS learns spatial and temporal correlations of IQ frames, which are central to principled optimization-based fUS frameworks; such tissue motion and blood flow statistics are often overlooked in data-driven deep learning approaches. NO-fUS achieves spatial filtering through convolutional blocks in the model, while temporal filtering is performed globally via the Fourier neural operator (FNO) [22], making the model discretization agnostic. Three distinct characteristics make NO-fUS highly generalizable in practice. First, unlike prior neural networks [20], NO-fUS respects the complex nature of the IQ input signal. Second, NO-fUS learns global frequency-based features crucial for filtering the temporally slow-moving tissue components from the blood volume flow. We later show that this operator-based temporal learning enables NO-fUS to be used at a sampling rate different from the trained frame rate. This allows NO-fUS to acquire IQ frames at a lower sampling rate, increasing the memory and communication efficiency of the fUS hardware technology. Third, NO-fUS guides the data-driven computations using an approximate DOP image, constructed from the limited acquired frames using SVD-based filtering [19]. This guidance is inspired by the observation that SVD on limited frames offers high structural similarity relative to the ground-truth image; hence, the approximated image can guide NO-fUS with an inferred spatial location of blood flow. We refer to our method without this guidance as `NO-fUS`, while in the presence of SVD guidance, we call our method `NO-fUS-G`.

We use the DOP images that have been processed by SVD, the most common tissue decluttering method in practice, as the ground truth. Given the paper's focus on reducing the number of IQ frames used during data acquisition, we define the `ground-truth` image as the one captured by SVD using the full set of T frames, where SVD filters out tissue clutter across all time frames, and the power is averaged over the entire set of T frames (i.e., $\texttt{DOP}_{\texttt{full}} + \texttt{SVD}_{\texttt{full}}$). This SVD filtering approach is the gold standard for fUS in neuroscience studies, so the ultimate goal is to achieve an image as close to this `ground-truth` while utilizing a lower number of IQ frames. The central frequency of fUS used during imaging varied depending on the species imaged and, in turn, the total set of T frames. To provide a comparison of what fUS image quality would look like using the K limited number of frames used by NO-fUS, we use an upper bound, $\texttt{DOP}_{\texttt{lim}} + \texttt{SVD}_{\texttt{full}}$, and a lower bound, $\texttt{DOP}_{\texttt{lim}} + \texttt{SVD}_{\texttt{lim}}$. We define $\texttt{DOP}_{\texttt{lim}} + \texttt{SVD}_{\texttt{full}}$, as the DOP image generated by averaging over only the limited observed K frames. Instead of using the K frames for SVD, we look into the "future" by using all T frames for SVD to remove the tissue motion clutter, providing improved tissue decluttering. This $\texttt{DOP}_{\texttt{lim}} + \texttt{SVD}_{\texttt{full}}$ serves as an upper bound on performance. We use both the `ground truth`
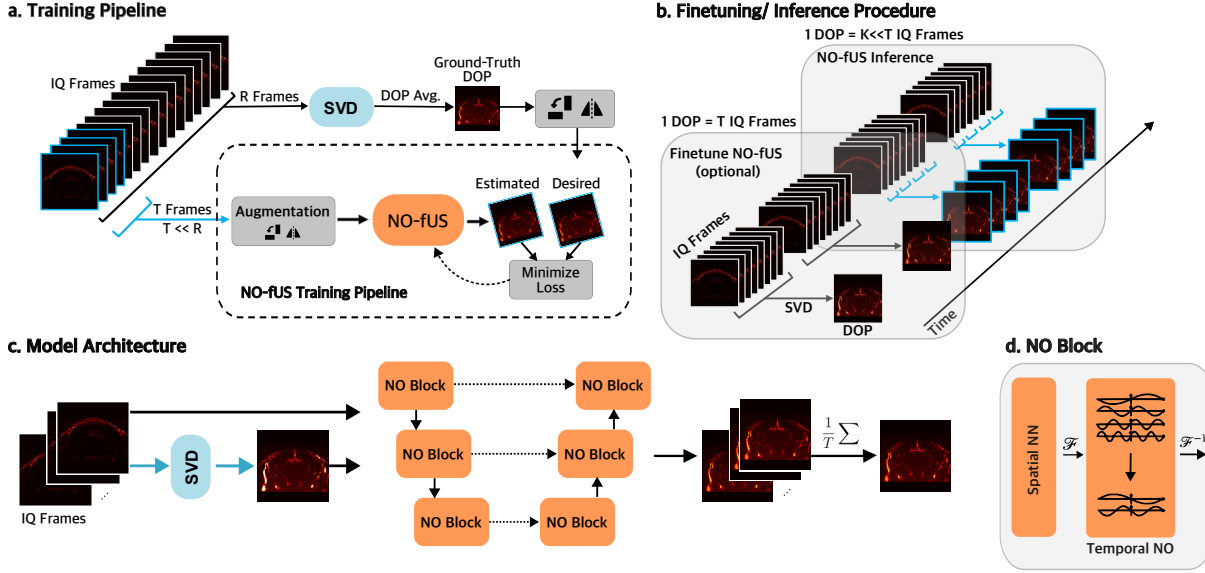
Figure 1: **Neural Operator for Functional Ultrasound (NO-fUS) Imaging Pipeline. a**, Training pipeline. The pipeline takes several frames of beamformed IQ complex signals as input. The input passes through an augmentation step (e.g., rotation, flipping) and a standardization step. The model outputs a DOP image, which is the average power of the filtered and decluttered IQ frames. Finally, a loss is computed based on the similarity between the model output and the desired DOP image; this loss is used via autograd and backpropagation to update the trainable parameters. **b**, Fine-tuning procedure and inference procedure. This module fine-tunes NO-fUS on a new session or animal. It takes a handful of IQ frames from the beginning of the session and fine-tunes the model by following the training pipeline within a short period of time. The inference stage offers efficient low-latency real-time fUS imaging using the trained or fine-tuned NO-fUS. For example, at a sampling rate of 500 Hz using 32 IQ frames at a time, NO-fUS can generate one DOP image every 64 ms, excluding processing latency. Notably, this is approximately $6 - 9$ times faster than the traditional SVD-based method, which uses $250 - 300$ input IQ frames. **c**, GPU-accelerated NO-fUS architecture. NO-fUS, a deep learning framework, consists of neural blocks that perform decomposed convolutional spatial filtering combined with frequency-based temporal operator mapping. NO-fUS takes a 3D input (time frames $\times$ image height $\times$ image width) of beamformed IQ signals. When guided by SVD, i.e., `NO-fUS-G`, it also takes an approximate DOP (image height $\times$ image width), which is computed using SVD on a limited number of time frames. The data dimensions remain the same across the time dimension as they are processed via the temporal neural operator. In contrast, the spatial dimensions decrease and increase via a U-Net architecture. The final layer takes the frame's average to output a single image of dimension (image height $\times$ image width) representing the DOP image (see Supplementary material for detailed architecture design). **d**, NO block. It performs decomposed spatiotemporal filtering; a neural network parameterizes the spatial filtering, and the temporal filtering is achieved through functional mapping with Fourier neural operators. See Supplementary material for a more detailed architecture.

and $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{full}}$ to evaluate the performance of NO-fUS which constructs the DOP image from only $K \ll T$ frames. We note that $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{full}}$ is not a practical method which uses the "future frames" for decluttering and it is only being used in our analysis to provide an upper bound in regard to the ground-truth. Finally, we compare NO-fUS to $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{lim}}$ that removes tissue clutter using only the K frames, which are then power-averaged to construct the DOP image. $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{lim}}$ is a more realistic representation of how DOP images would appear if using reduced K frames.

NO-fUS is the state-of-the-art accelerated fUS that runs at a faster frame rate during inference. Using the training pipeline, we trained NO-fUS on a fUS imaging dataset from five mouse brains, where the target image to construct was the DOP image provided by SVD-based fUS at the original frame

rate (see Supplementary Material for more details). In the remainder of the manuscript, we refer to this dataset as MultiMice.
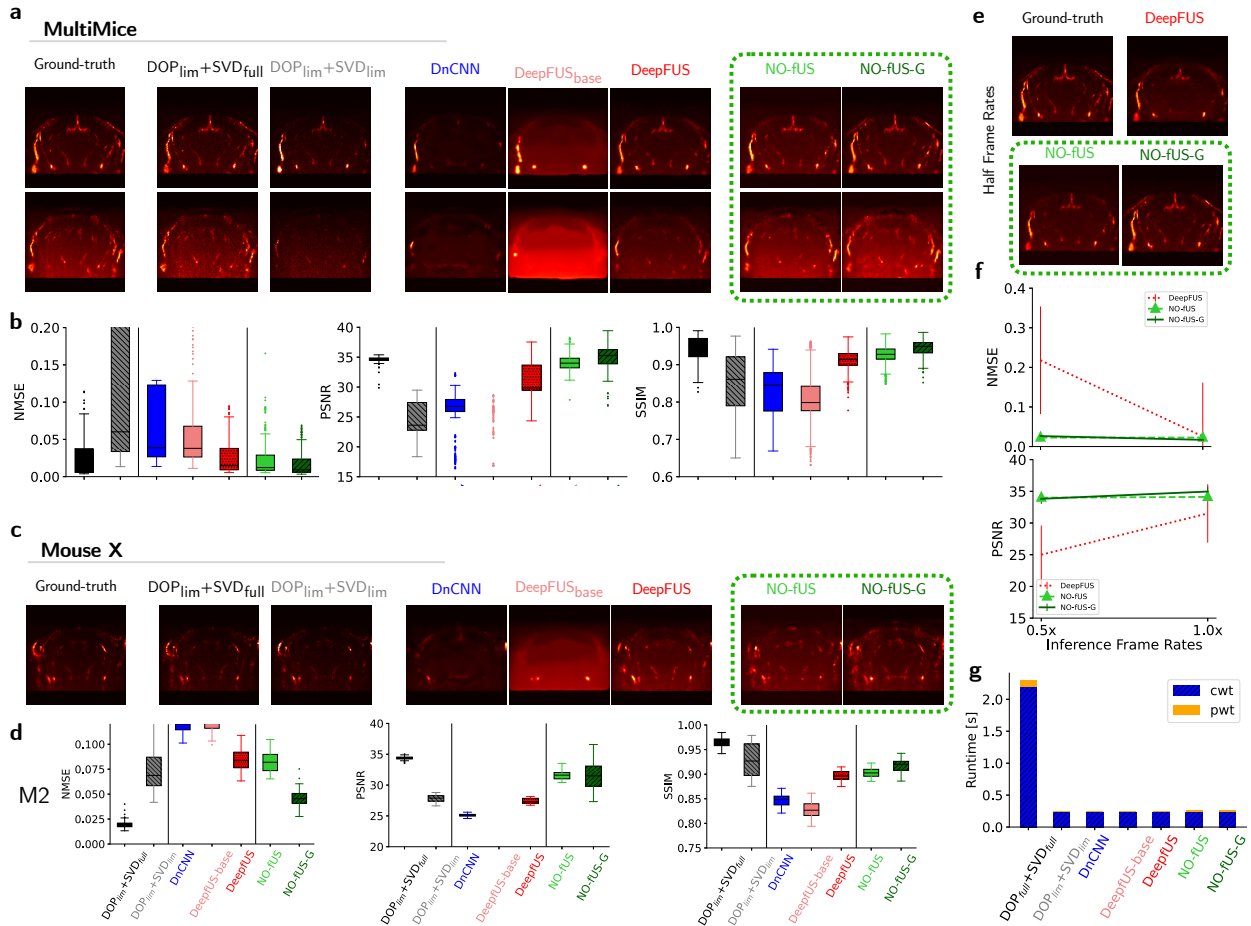

Figure 2: **Outperformance generalization of NO-fUS on unseen frames from trained (MultiMice) and new mouse (Mouse X). a**, Qualitative analysis of NO-fUS and the baselines on the test set (unseen frames) from the MultiMice dataset. The best-case and worst-case performance within MultiMice are shown in each row. From left to right, columns represent: Ground-truth DOP (computed by performing SVD-based de-cluttering and power averaging using the full 300 frames), $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{full}}$, $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{lim}}$ [16, 17, 19], DnCNN [31], DeepFUS [20], $\texttt{DeepFUS}_{\texttt{base}}$, NO-fUS, and NO-fUS-G. **b**, Quantitative comparison of the proposed method (NO-fUS, and NO-fUS-G) with baselines in terms of NMSE, PSNR, and SSIM on the test set, MultiMice. **c**, Qualitative analysis of the performance on a Mouse X for activity during visual stimulus presentation and the worst case performance. **d**, Quantitative comparison of the proposed method (NO-fUS, and NO-fUS-G) with baselines on Mouse X to highlight generalization to unseen brains. **e**, Qualitative analysis of DeepFUS, DeepFUS, NO-fUS, and NO-fUS-G on MultiMice when the inference frame rate is $\times 0.5$ of the trained IQ frame rate. The visualization shows the generalization capability of NO-fUS against DeepFUS. **f**, Performance of the methods on MultiMice at full ($1.0\times$) and half ($0.5\times$) inference frame rates (NMSE and PSNR are shown; see Supplementary Materials for SSIM). **g**, The runtime of methods, divided into collection-wait time (cwt) and process-wait-time (pwt); cwt includes the transmission and receiving of the ultrasound pulses, the beamforming process, and the storage time along with a buffer for safety.

<u>Baselines.</u> The ground-truth images are constructed using $T = 300$ IQ frames, while NO-fUS uses only $K = 32$ IQ frames (See Supplementary Material for $K = 16$ or 64, and the Monkey experiments where $T = 250$). In addition to $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{lim}}$, we compared NO-fUS with data-driven deep learning-based methods (fig. 2). They all were trained similarly to NO-fUS for a fair comparison, thus any differences in performance

highlight the architectural advantages of NO-fUS. We report performance based on normalized mean square error (NMSE, $\downarrow$), peak signal to noise ratio (PSNR, $\uparrow$), and structural similarity index measure (SSIM, $\uparrow$). NMSE and SSIM range between 0 to 1. Using the MultiMice test set, we compare NO-fUS to a variety of models and demonstrate the superiority of NO-fUS (fig. 2a,b).

Importance of temporal filtering. First, we show that a decrease in the number of IQ frames increases the difficulty of differentiating between tissue clutter and blood flow, resulting in poor $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{lim}}$ performance. NO-fUS is trained to learn the temporal correlations from limited IQ frames, hence, it improves upon $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{lim}}$ and shows a close performance to $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{full}}$. Second, a data-driven filtering approach should be based on the inherent difference between the desired blood flow signal and the undesired signal of tissue clutter. Otherwise, they fail to generalize. To demonstrate this, we compare NO-fUS to DnCNN [31] and DeepFUS-base [20]. DnCNN is a convolutional neural network that operates on the real component of the IQ time frames and has been previously used in the literature for general image denoising [31]. DeepFUS-base is the basic version of DeepFUS, specifically designed for fUS, but it does not include 3D convolution over time and space to mix time frames at the beginning of the network architecture [20]. NO-fUS outperforms DnCNN and DeepFUS-base (fig. 2b). We attribute the inability of DnCNN and DeepFUS-base to perform on multiple brains to the lack of temporal processing of the IQ time frames in their neural architectures. For instance, we observe that while DnCNN and DeepFUS-base perform well on data from one mouse within the MultiMice training set, they fail on another (fig. 2a). Finally, we note that we also attempted to apply a Robust PCA optimization-based approach for fUS [32], as suggested in [33]. However, despite varying the hyperparameters, the method—taking several minutes per image—failed to perform well. We attribute this to using the limited number of frames, $K = 32$, as opposed to the full number of $T = 300$ frames.

Comparison to the best available method. Next, we compare NO-fUS to DeepFUS [20], the current best published deep learning method available for reconstructing fUS data from sparse data (fig. 2a,b). DeepFUS is a deep learning framework that operates only on the real component of the IQ time frames, treats the time frames as channels, and ignores the temporal aspect of the data. DeepFUS, an extension of DeepFUS-base, includes only one 3D convolutional block at the beginning of its architecture to combine the time frames temporally before the deep processing layers. Our results show that NO-fUS has better performance than DeepFUS. The improved performance metrics of NO-fUS compared to DeepFUS highlight the importance of two features present in NO-fUS but lacking in DeepFUS: processing the *complex* IQ frames rather than only the real portion and incorporating a series of *temporal* processing layers in addition to spatial filtering in the deep layers.

Lastly, we show that the generalization of NO-fUS can further improve by adding SVD Guidance (see fig. 1d and Methods section); in this setting, NO-fUS-G outperforms NO-fUS on the MultiMice test set, achieving state-of-the-art (SOTA) results. Overall, fig. 2a visualizes image examples from each of the mice. Another notable advantage of NO-fUS-G is that it captures detailed spatial segments which may be entirely neglected by fully data-driven methods, such as DeepFUS and NO-fUS.

Generalization on 0.5× IQ frame rate. An important property of NO-fUS, compared to DeepFUS, is that NO-fUS uses neural operators to learn global temporal processing instead of localized temporal filtering. This allows NO-fUS and NO-fUS-G to run at 0.5× the IQ frame rate, skipping every alternate frame without performance decline. In contrast, testing on MultiMice testset, DeepFUS's performance deteriorates when IQ time frames are acquired at half the frequency compared to the trained frame rate (fig. 2e,f). The slight performance decline in NO-fUS-G is due to the fact that the framework relies on a guided SVD, which is computed from a limited half frame rate IQ signal and thus has lower structural similarity to the ground truth. Overall, the ability of NO-fUS to run at 0.5× the original fUS frame rate during inference is a crucial capability, as it reduces the computational, memory, and communication cost of data acquisition by an additional 50%.

Runtime. Finally, we compare the runtime of NO-fUS with other frameworks (fig. 2g). We divide the runtime into two components: one is the time taken to collect all frames, i.e., collection-wait-time (cwt), which includes frame acquisition, beamforming processing time, and data storage, and another is the time taken to process the IQ input signal into a Doppler image , i.e., processing-wait-time (pwt), which includes cohesive compounding of frames and SVD. To acquire the ground truth image, cwt is 2.2 seconds (equivalent to collecting and storing $T = 300$ frames at 500 Hz IQ sampling rate) and pwt is 0.0949 seconds. The proposed pipeline enables the approximately linear reduction of the cwt by $\approx 90\%$ (equivalent to collecting $K = 32$ frames at 500 Hz IQ sampling rate); hence, the cwt of all the discussed deep learning-based methods

are $^{32}/_{300} \times 2.2 = 0.23$ s.

We observe that $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{lim}}$ using K limited IQ frames has lower pwt (0.0055 s) than $\texttt{NO-fUS}$ (0.0225 s) and $\texttt{NO-fUS-G}$ (0.0230 s); however, we note that $\texttt{NO-fUS}$ and $\texttt{NO-fUS-G}$ have better image quality metrics than when using SVD. The pwt of other deep learning frameworks are $\texttt{DnCNN}$ (0.0026 s), $\texttt{DeepFUS-base}$ (0.0031 s), and $\texttt{DeepFUS}$ (0.0031 s). Although $\texttt{NO-fUS}$ pwt is a bit longer than the other deep learning frameworks due to its delicate design of temporal filtering using Fourier neural operators (FNO)s [22, 23], $\texttt{NO-fUS}$ has the best imaging quality among all, and it is still much faster than full-frame T SVD with a pwt of 0.0949 s.

We emphasize that the cwt is the runtime bottleneck compared to the pwt. This brings the total effective runtime of $\texttt{NO-fUS}$ to 0.2525 s and $\texttt{NO-fUS-G}$ to 0.2530 s. These runtimes are on the same order as other deep learning frameworks and are faster compared to the $\texttt{ground-truth}$ method, $\texttt{DOP}_{\texttt{full}}+\texttt{SVD}_{\texttt{full}}$, (2.2949 s) run at full collection wait time with SVD processing. Overall, the significant reduction of cwt ($\approx 90\%$) and pwt ($\approx 0.75\%$), results in remarkable speed improvement and achieves DOP imaging every 0.2525 s as opposed to 2.2949 s for the ground-truth. In the Monkey experiments discussed later, we reduce the number of IQ frames from 250 to 32 for the construction of one DOP image, resulting in the reduction of cwt by 87.2%. This is of great advantage to overcome the latency limitations and also hardware efficiency of fUS for real-time brain-computer interfaces (BCIs) [13].

Generalization to a new mouse without fine-tuning. Next, we discuss the generalization of $\texttt{NO-fUS}$ and $\texttt{NO-fUS-G}$, for fUS imaging from a new mouse whose brain's images are not seen during the deep learning training stage. We call this mouse, Mouse X, and consider one experimental session containing DOP images from the resting state (Mouse X $\texttt{M1}$) and during visual stimulation (Mouse X $\texttt{M2}$). For the visual stimulation experiment, the animal's brain was activated with a visual stimulus every 10 DOP images, using 300 IQ frames per DOP(fig. 2c,d). For the remainder of the manuscript, we compare our proposed framework to only $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{lim}}$ and the best performing published deep learning framework, $\texttt{DeepFUS}$. Given $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{full}}$ performance on $\texttt{M1}$ and $\texttt{M2}$, $\texttt{NO-fUS-G}$ provides high-quality DOP imaging in the absence of any fine-tuning. This improves upon NMSE and PSNR evaluation metrics compared to $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{lim}}$ performed on limited IQ frames. In this case, $\texttt{NO-fUS}$ does not perform as well as $\texttt{NO-fUS-G}$. Lastly, $\texttt{DeepFUS}$ loses the ability to capture detailed information on the blood flow shown by its poor performance metric, performing worse than $\texttt{NO-fUS}$ and $\texttt{NO-fUS-G}$.

We attribute the superior generalization performance of $\texttt{NO-fUS-G}$ to two main architectural properties: One is that, unlike DeepFUS, NO-fUS processes complex IQ signals. Second is that $\texttt{NO-fUS-G}$ leverages a spatial guidance provided by limited-time-frame SVD, a feature that is missing in DeepFUS. To support this, we performed additional experiments with all the deep learning blocks operating in the real domain, using only the real portion of IQ input signal for NO-fUS (fig. 6). At first glance, the quantification results on MultiMice test set may suggest that $\texttt{NO-fUS-real}$ performs better than $\texttt{NO-fUS}$. However, having a closer look, the complex framework $\texttt{NO-fUS}$ has a much better generalization to Mouse X than $\texttt{NO-fUS-real}$ (fig. 6b,c), which throws away the information contained in the imaginary part of the IQ signal.

Despite the aforementioned state-of-the-art and promising performance, one of the obstacles to the widespread usage of deep learning in practice is its lack of generalization and reliability beyond the trained dataset, a problem also known as domain shift [34]. In medical imaging, this challenge, known as out of distribution (OOD), is particularly evident when generalizing deep learning models from one experimental session to another or from one brain to another [35]. As we demonstrated above, we have addressed this challenge by guiding NO-fUS with an approximate DOP image constructed by the SVD method, improving generalization to new Mouse X in the absence of fine-tuning. Next, we propose an additional strategy to further enhance generalization and expand the practical usage of NO-fUS to new species and new sessions.

Practical generalization of NO-fUS on new sessions and new animals. We discuss a highly efficient and fast fine-tuning process for NO-fUS to construct DOP images across different species. In this setting, we take NO-fUS, trained on MultiMice, and fine-tune it on the session of interest for different target species. We demonstrate that NO-fUS can be fine-tuned efficiently and quickly on a single GPU (e.g., NVIDIA GeForce RTX 4090), allowing users to apply NO-fUS for imaging from new mice, monkeys, and humans (fig. 3). For details on the fine-tuning process, see Supplementary Methods.

Mouse experiment. Focusing on Mouse X, which was excluded from the MultiMice train set (fig. 3a,b), we fine-tune (ft) NO-fUS, trained on MultiMice, on only two DOP images from the beginning of $\texttt{M1}$ for 1600 steps. This fine-tuning process took only 6.13 and 6.15 minutes for $\texttt{NO-fUS}_{\texttt{ft}}$ and $\texttt{NO-fUS-G}_{\texttt{ft}}$ on an NVIDIA GeForce RTX 4090 GPU-accelerated machine; we note that using a more powerful GPU can further
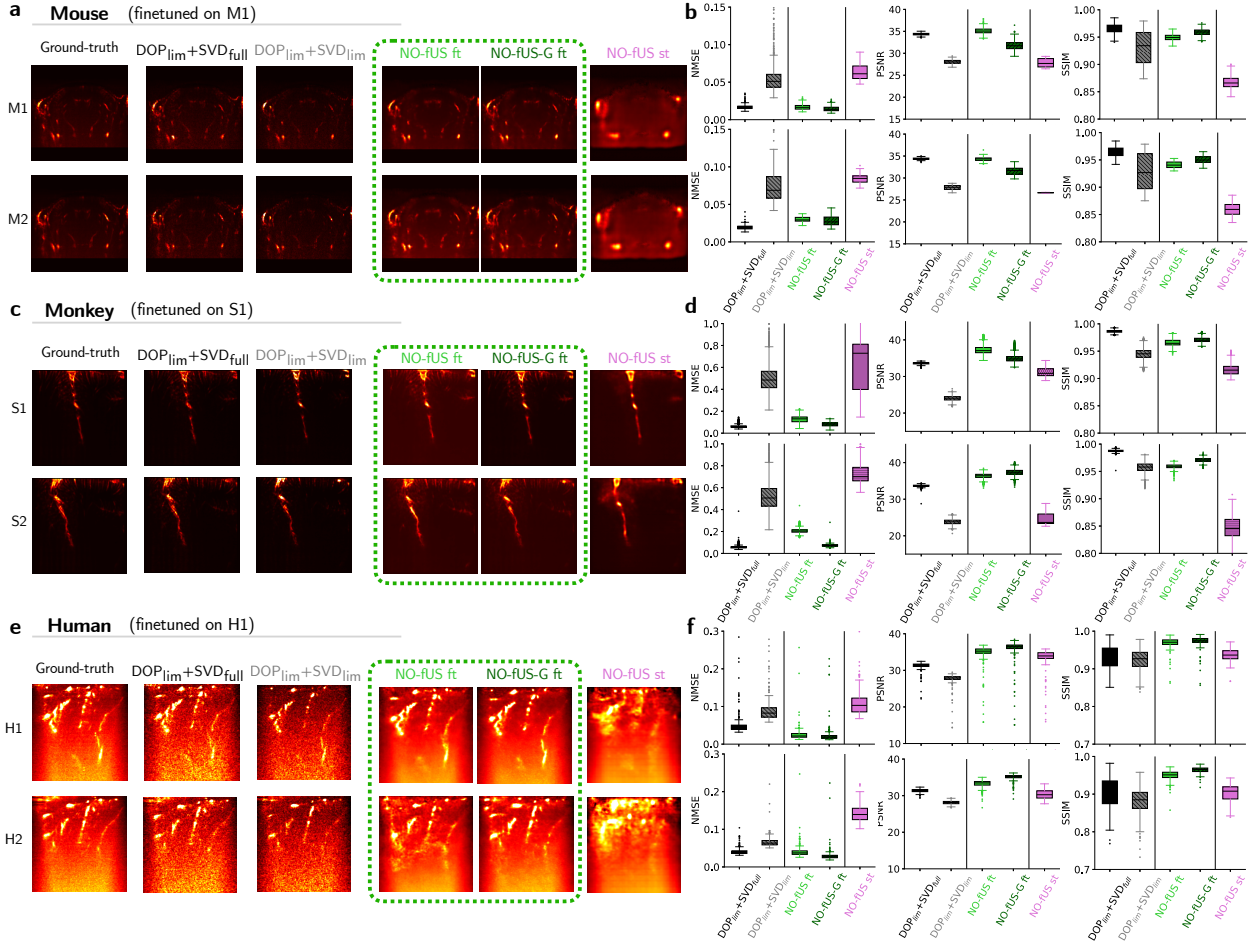
Figure 3: **Generalization demonstration of NO-fUS on new sessions and new animals.** Ground-truth DOP is computed using SVD de-cluttering based on 300 IQ frames for Mouse and Human, and 250 IQ frames for Monkey data. The remainder of the methods, $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{full}}$, $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{lim}}$, $\texttt{NO-fUS}_{\texttt{ft}}$, $\texttt{NO-fUS-G}_{\texttt{ft}}$, and $\texttt{NO-fUS}_{\texttt{st}}$, are all based on 32 IQ frames. **a-b**, Performance visualization and quantitative analysis from mouse brain for M1 and M2 runs. **c-d**, Monkey visualization and quantitative results from two sessions of S1 and S2. **e-d**, Visualization and quantitative results for two sessions of H1 and H1 from Human experiments.

shorten the fine-tuning duration. While $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{lim}}$ using limited IQ frames ($K = 32$) captures the spatial properties of blood flow, it is noisy and sensitive to the SVD threshold, and can end up removing blood flow information in some cases (fig. 3a); this results in relatively low structural similarity in some examples, and overall poor average reconstruction even though it offers high SSIM for a portion of the dataset (fig. 3b). NO-fUS' data-driven methods can resolve this problem. We observe that $\texttt{NO-fUS}_{\texttt{ft}}$ reconstructs DOP imaging better than $\texttt{DeepFUS}_{\texttt{ft}}$ (See Supplementary materials). In particular, $\texttt{NO-fUS}_{\texttt{ft}}$ has better generalization on M2, which is not seen during the fine-tuning stage.

We attribute this superior performance to the temporal IQ frame processing by NO-fUS, which captures an intrinsic property of the Doppler image that is missing in DeepFUS. However, we observe that NO-fUS misses some structural details in the DOP image, which can be improved by adding SVD guidance, i.e., $\texttt{NO-fUS-G}_{\texttt{ft}}$ achieves results that is close to the upper limit $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{full}}$. This highlights the practicality of the NO-fUS fine-tuning pipeline. Indeed, training NO-fUS from scratch, $\texttt{NO-fUS}_{\texttt{st}}$, for the same duration as fine-tuning, results in a poor DOP imaging (see Supplementary Materials for fine-tuning results using 16 images). Having observed the superior generalization of NO-fUS over DeepFUS, we focus on the comparison between our proposed NO-fUS method and SVD-based methods for the remainder of the manuscript.

Monkey experiment.   We highlight the applicability of NO-fUS for functional ultrasound neuroimaging

in Rhesus macaques (fig. 3c,d). We consider three sessions - $\mathtt{S1}$, $\mathtt{S2}$, and $\mathtt{S3}$ - from Monkey Y. We fine-tuned NO-fUS, which was previously trained on MultiMice, on 120 DOP images from the beginning of $\mathtt{S1}$ using only 3600 gradient steps; this process took 7.07 and 7.15 minutes on our GPU-enabled machine for $\mathtt{NO\text{-}fUS_{ft}}$ and $\mathtt{NO\text{-}fUS\text{-}G_{ft}}$, respectively (see Supplementary Material for training on 5 DOP images and performance from lower number of training steps). We then evaluate the performance on the remainder of the frames in $\mathtt{S1}$, and the full sessions of $\mathtt{S2}$ and $\mathtt{S3}$, which were not seen during the fine-tuning stage. The desired DOP image in the monkey experiments was computed using $T = 250$ IQ frames. We construct the DOP using only $K = 32$ IQ frames. $\mathtt{DOP_{lim}+SVD_{lim}}$ captures the general spatial properties of the blood but misses some details due to the sensitivity of SVD thresholding performed on limited times frames. As discussed before, this is evident from the large variance in reported NMSE across the dataset and low PSNR. $\mathtt{NO\text{-}fUS_{ft}}$ shows much better performance than $\mathtt{DOP_{lim}+SVD_{lim}}$, but it still misses some of the fine blood flows (fig. 3c). This limitation arises from the fully data-driven nature of NO-fUS. We address this limitation using our best performing framework, $\mathtt{NO\text{-}fUS\text{-}G_{ft}}$, which shows performance very close to the upper bound $\mathtt{DOP_{lim}+SVD_{full}}$.

Given the fine-tuned NO-fUS on $\mathtt{S1}$, we next highlight its ability to generalize on new unseen sessions, $\mathtt{S2}$ (fig. 3c, second row). First, we emphasize that training NO-fUS from scratch $\mathtt{NO\text{-}fUS_{st}}$ on these sessions results in very poor DOP imaging. While $\mathtt{DOP_{lim}+SVD_{lim}}$ is better than $\mathtt{NO\text{-}fUS_{st}}$, it suffers from the same above-discussed limitations. $\mathtt{NO\text{-}fUS\text{-}G_{ft}}$ significantly improve upon $\mathtt{DOP_{lim}+SVD_{lim}}$. Notable performance of $\mathtt{NO\text{-}fUS\text{-}G_{ft}}$ is that it gets very close to $\mathtt{DOP_{lim}+SVD_{full}}$ and offers a less noisy, hence higher PSNR, DOP image on a session that is not observed during the fine-tuning stage.

Human experiment.   Finally, we show the applicability of NO-fUS for imaging from a human subject with an acoustic cranial implant (fig. 3e,f). In these experiments, we used the NOfUS that is already trained on MultiMice and only fine-tuned it. For evaluating imaging quality, the desired DOP image was computed using $T = 300$ IQ frames, the $\mathtt{ground\text{-}truth}$ ($\mathtt{DOP_{full}+SVD_{full}}$). We focus on performance in two separate human sessions ($\mathtt{H1}$, $\mathtt{H2}$). We fine-tuned the data-driven models on 32 DOP images captured from the beginning of the $\mathtt{H1}$ session at the original Doppler imaging rate using 300 IQ frames. Results highlight that while $\mathtt{DOP_{lim}+SVD_{lim}}$ is sensitive to the threshold used for the removal of the tissue clutter, NO-fUS successfully filters out a majority of the tissue clutter without removing useful blood flow information (fig. 3e). Moreover, NO-fUS provides a DOP image that is less noisy than $\mathtt{DOP_{lim}+SVD_{full}}$. This results in better quantitative performance of $\mathtt{NO\text{-}fUS\text{-}G_{ft}}$ than $\mathtt{DOP_{lim}+SVD_{full}}$ (fig. 3f). For reference, we show that the generalization of NO-fUS is unreachable if one trains the model from scratch, i.e., $\mathtt{NO\text{-}fUS_{st}}$. Our reported results also apply to the new human session $\mathtt{H2}$ which is not seen at all during the fine-tuned stage.

Behavioural decoding using NO-fUS. fUS-based BCI is an emerging technology that can decode movement and planning from fUS signal, advancing technology that can allow subjects to control robotic limbs and other assistive devices with their thoughts [12, 13, 14]. We propose to use NO-fUS to power BCIs to reduce the latency and efficiency of decoding. Unlike prior published deep-learning methods, we show in both monkey and human experiments that NO-fUS not only produces high quality DOP reconstructions but also preserves important behavior-related information within reconstructed images that can be useful for decoding purposes in BCIs.

Monkey experiment.   NO-fUS is capable of computing DOP images from fewer frames of the acquisition. The original DOP frame rate offers one image per 250 IQ frames resulting in total of 3 DOP images within a 1.5 s memory period; NO-fUS increases this rate by approximately $8\times$ and provides 24 DOP images within the same memory period, each from 32 IQ frames. Similar to the above-discussed results, NO-fUS was fine-tuned on the first 120 DOP frames from $\mathtt{S1}$ session and then used for inference across all other sessions. This increased imaging rate is well-suited for implementation in an online decoder, suggesting potential for real-time decoding applications. We evaluated the performance of $\mathtt{NO\text{-}fUS}$ and $\mathtt{NO\text{-}fUS\text{-}G}$ against $\mathtt{DOP_{lim}+SVD_{lim}}$ using 32 IQ frames and the oracle $\mathtt{DOP_{lim}+SVD_{full}}$, which de-clutters the fUS signal using all 250 IQ frames.

fUS data was obtained from a rhesus macaque monkey while it performed two different eye saccade tasks: a task that varies direction of saccades ($\mathtt{S1}$ and $\mathtt{S2}$, fig. 4a) and a task that varies the saccade distances ($\mathtt{S3}$, fig. 4c). The goal was to decode the monkey's movement planning prior to the action. We used an offline decoding pipeline based on principal component analysis (PCA) for dimensionality reduction and linear discriminant analysis (LDA) for decoding to predict the monkey's planned movement directions/distance after complete data collection (see Supplementary material for more details).

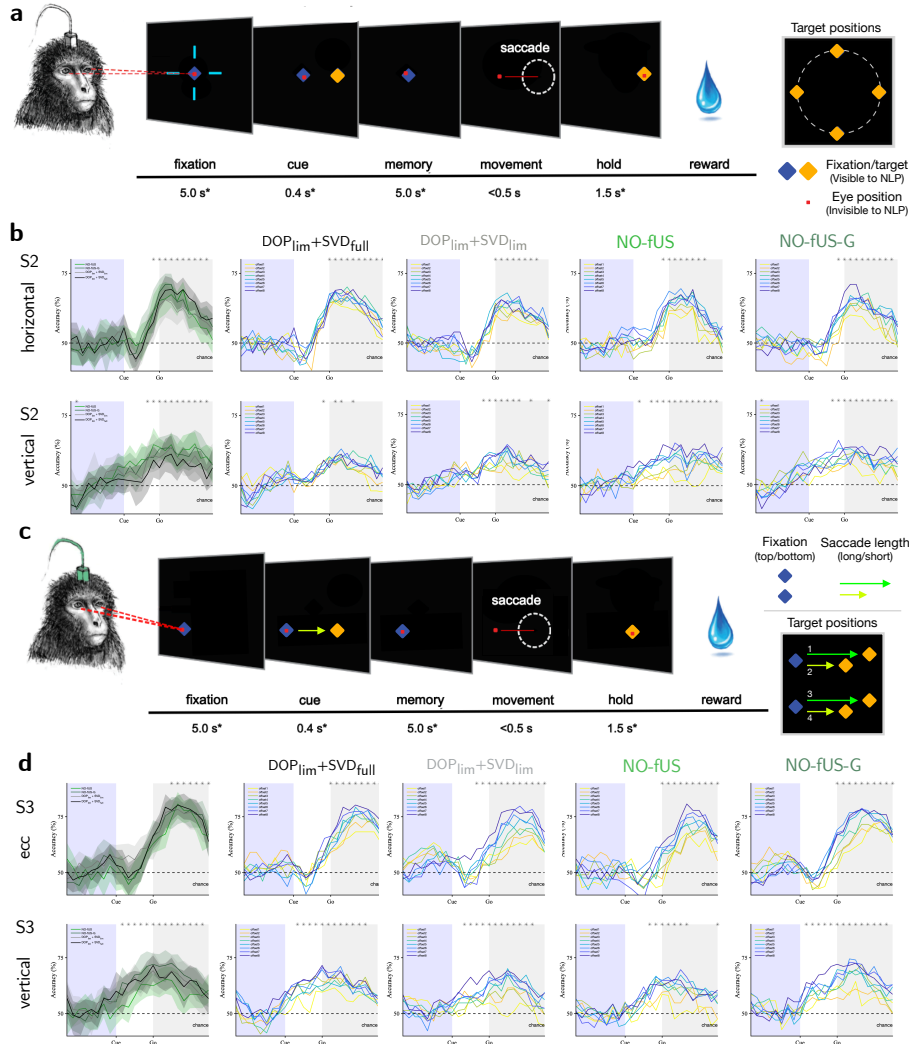Using this pipeline, we evaluated the decoding performance over a sliding window of 3 DOP images

Figure 4: **NO-fUS-powered BCI to Decode Thoughts from Monkey. a**, Memory guided saccade task (sessions `S1` and `S2`) for decoding varied directions for target/fixation points. ±1000 ms of jitter for fixation and memory periods; ±500 ms of jitter for hold period. The peripheral cue was chosen out of four possible target locations (up, down, left, and right). The red square signifies the monkey's line of sight. Horizontal and vertical decoding was done. **b**, Decoding performance for `S2`. Left panels show decoding performance comparing all methods. The remaining panels on the right show decoding performance individually for each method as DOP frame rate in the decoding sliding window decreases from full 8 DOP images to 1 DOP per second. The vertical lines denote the Cue and Go onsets. **c**, Memory guided saccade task (session `S3`) for decoding varied distances or "eccentricity (ecc)" between fixation and target. ±1000 ms of jitter for fixation and memory periods; ±500 ms of jitter for hold period. The green line is for long-range right-ward saccades, and the yellow line is for short-range rightward saccades. The peripheral cue was chosen out of four possible target locations (fixation point at top, short-range rightward saccade, fixation point at bottom, short-range rightward saccade, fixation point at top, long-range rightward saccade, and fixation point at bottom, long-range rightward saccade). The red square signifies the monkey's line of sight. Since the fixation point varied vertically, both eccentricity and vertical decoding were done. **d**, Decoding performance from `S3`. The first row shows eccentricity (ecc), and the second shows decoding vertical movement planning. Panel format is similar to b.
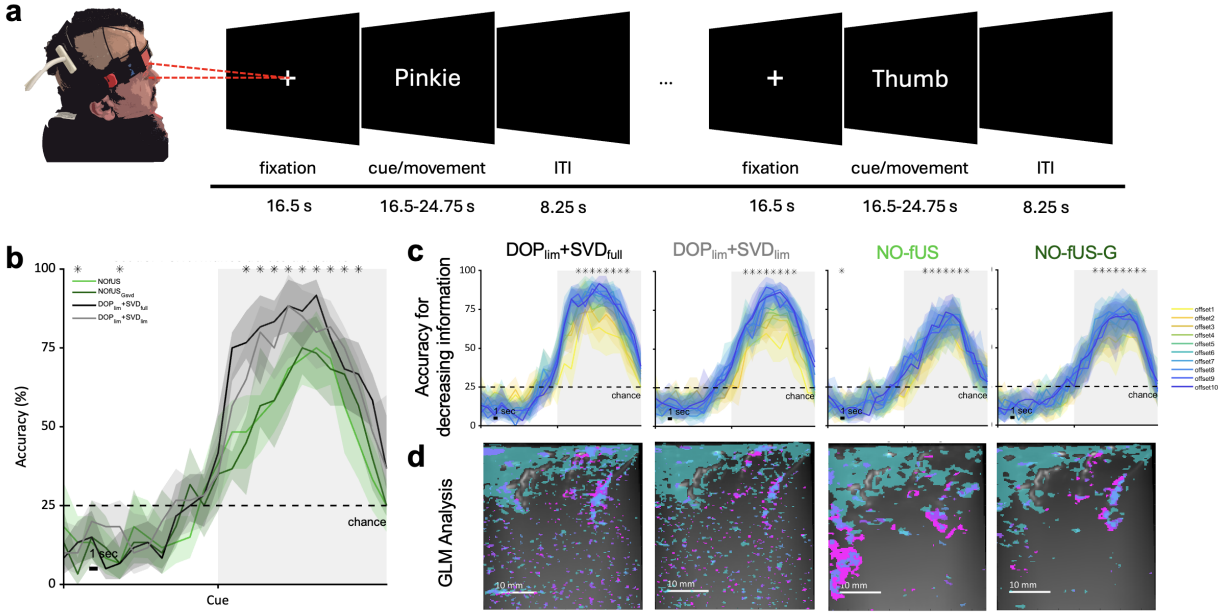
Figure 5: **NO-fUS-powered BCI to Decode Planning from Human. a**, Instructed delay motor task for human experiments. The participant was asked to fixate on a central fixation cross and then perform the cued movement upon cue. The participant was asked to repeat the movement until the cue disappeared. Trials were separated by an inter trial interval (ITI) to allow fUS signal to return to baseline. **b**, The decoding accuracies over time across a trial for all methods using full DOP information from all 9 DOP images. **c**, The decoding accuracy across time for each method when reducing the number of images averaged to generate one DOP image. Each sub-figure corresponds to one method, containing decoding curves when 1 DOP (yellow) to 9 DOP (full information, dark blue) images are used. **d**, Combined mapping of statistically significant regions of interest generated from general linear modeling (GLM) for each effector across each model.

in length. We first compared the average decoding accuracy over time across a trial for different models, $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{full}}$, $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{lim}}$, $\texttt{NO-fUS}$ and $\texttt{NO-fUS-G}$, using all 8 images. Then we examined how decoding accuracy changed as the amount of information per DOP was reduced. This was done by comparing decoding accuracy as the total number of images averaged to generate one DOP image, ranging from 8 images (dark blue) to 1 image (yellow), was reduced, effectively reducing frequency of DOP sampling.

Results confirm that when full information is used, $\texttt{NO-fUS}$ and $\texttt{NO-fUS-G}$ can preserve behavioral features and achieve similar accuracy as the SVD-based decluttering baseline (fig. 4b,d). Moreover, findings in fig. 4b,d suggest that $\texttt{NO-fUS}$ and $\texttt{NO-fUS-G}$ are more robust to the decreasing frequency of DOP sampling (dark blue to yellow color) compared to $\texttt{DOP}_{\texttt{lim}}+\texttt{SVD}_{\texttt{lim}}$.

<u>Human experiment.</u>   We demonstrate that NO-fUS is similarly able to maintain behavioral information in a human subject using the same analysis for monkey decoding. We compared the decoding performance of $\texttt{NO-fUS}$ and $\texttt{NO-fUS-G}$ with SVD-based approaches using data recorded from a human subject with an acoustic cranial implant. For DOP reconstruction, all methods provide one image per 32 IQ frames from the original 300 IQ frames. This results in $9\times$ more DOP images compared to the original ground truth using all 300 IQ frames to construct one DOP image.

fUS data was recorded from the primary motor cortex of a human participant with an acoustic cranial window implant while they performed randomized instructed movement tasks in block task format. The subject first fixated on a central fixation cross and then was cued to move the chosen movement effector for a set period of time. During this period, the subject was asked to repeat the movement until the cue no longer remained on the screen. Movement effectors included the right index finger, right wrist, lip, and tongue (fig. 5a). DOP data was then processed and analyzed offline using general linear modeling (GLM) to map significant task-correlated voxels ($p < 10^{-3}$). DOP data underwent the same offline decoding pipeline as used for the above monkey decoding experiments to decode cued movement effectors. Using traditional,

groundtruth DOP images, movement effectors could be significantly decoded above chance level at more than 75 percent accuracy.

We used a causal sliding window of 1 DOP image in length to calculate decoding performance over time across a trial. We also examined performance as the amount of information per DOP image was reduced. Like in monkey, this was done by comparing decoding accuracy as the total number of images averaged to generate one DOP image was reduced. However, in human, this ranged from 9 images (dark blue) to 1 image (yellow). When comparing decoding accuracies using the full number of images across models, `NO-fUS` and `NO-fUS-G` had slightly lower decoding accuracy than $DOP_{lim}+SVD_{full}$, and $DOP_{lim}+SVD_{lim}$ (fig. 5b). However, when examining their performance using fewer DOP images, NO-fUS shows decoding performance similar to that of the SVD-based methods. Moreover, the decoding accuracy of SVD-based approaches is not robust to the amount of information provided in each window. Their performance decays once they reach only one DOP image. On the other hand, NO-fUS robustly maintains decoding performance as the number of images is reduced (fig. 5c). Finally, the GLM analysis demonstrates that `NO-fUS` and `NO-fUS-G` preserve the functional behavior mappings similar to the SVD-based approaches. However, while `NO-fUS` and `NO-fUS-G` produce similar representations, they also include substantial noise.

3   Discussion

Functional ultrasound (fUS) imaging is an emerging neurotechnology with significant potential for widespread application in long-term, minimally invasive neurorecording during active movement and behavior. fUS offers advantageous sensitivity and spatial resolution compared to fMRI and EEG, and is a less invasive option than microelectrode recording with a greater field of view (although less temporal resolution) [6]. A major disadvantage of fUS is that it is limited by the large number of cohesively compounded intermediate frames required to construct a single Doppler power (DOP) image (approximately 250-300 frames). Thus fUS acquisition is computationally inefficient. It requires thousands of pulses per DOP, which can result in delays due to increased time needed to store and transfer the data needed to construct a single DOP image. Additionally, this process can result in in detrimental probe heating during prolonged imaging, particularly with an increased number of probe elements. These limitations are expected to become even more significant as advances in fUS imaging technology, such as 3D volumetric imaging, further increase the data requirements for providing a single DOP image.

We propose the neural operator for functional ultrasound (NO-fUS), a deep learning solution that can reduce the frames needed to reconstruct a high-quality fUS image. We highlight the capability of NO-fUS in the reconstruction of DOP images while the number of input frames is reduced by 90% (from 250-300 frames to 32). We provide two variants of NO-fUS, one that is fully data driven, and another that is guided by an approximated DOP image. NO-fUS is carefully designed for spatiotemporal processing via neural operators, allowing for a change in the in-phase and quadrature (IQ) sampling rate without loss of image quality. This work highlights this capability by reducing the sampling rate by 50%. Optimizing hardware can only partially address this issue. Moreover, existing deep-learning methods for reconstructing DOP images from fUS data a) do not take into account spatiotemporal filtering of fUS data, b) do not support change of IQ sampling rate or number of frames, and c) fail to test the extent reconstructed images can be used to decode behavioral variables [20]. NO-fUS is a highly generalizable deep learning framework and achieves the state-of-the-art. NO-fUS not only offers structurally similar high-quality DOP images from significantly reduced frames across multiple species and sessions but also retains decodable information within DOP images shown by the comparable behavioral decoding accuracies computed from NO-fUS reconstructed images.

NO-fUS provides high-quality DOP reconstruction from limited complex IQ signals, achieving visual quality and image similarity metrics comparable to the current fUS standard SVD, while using a fraction of the time and data. This reduces the need for computationally demanding processing and filtering [19]. NO-fUS consistently outperforms other deep learning methods, such as DeepfUS [20], in these metrics. We attribute this generalization to the novel architecture of NO-fUS, which respects the complex and temporal nature of the data via neural operators, unlike DeepfUS. Furthermore, NO-fUS is capable of constructing high-quality DOP images across different sessions, planes, and even organisms, both with and without fine-tuning a small subset of data. Training and fine-tuning can be efficiently performed on a single GPU. Thus, NO-fUS presents a user-friendly and highly practical technique for providing DOP images from reduced

IQ frames, applicable to a broad range of fUS experiments without requiring substantial modifications to the fUS acquisition pipeline.

Beyond structural image similarity metrics, NO-fUS is the first deep-learning method capable of reconstructing DOP images from reduced IQ frames with flexible IQ frame rates while preserving accurate decoding of complex behavioral information within the reconstructed images. While, DeepfUS is able to produce task-evoked functional activation maps similar to state of the art methods, these maps degrade as data sparsity increases, and they do not examine DeepfUS' ability to decode complex behavioral information, focusing only on task-correlated activity for a visual-stimulation task.

NO-fUS achieves high decoding accuracy comparable to the ground-truth SVD-based approaches across various behavioral tasks in both NHPs and human subjects for multiple task conditions. Importantly, this decoding accuracy remains robust as the amount of information used per DOP is reduced unlike in DeepfUS, whereas the accuracy of the SVD-based fUS declines significantly when the information is reduced to 32 frames. This capability supports applications of fUS for decoding behavioral information as part of a brain-machine interface (BMI) by enabling up to a 7 to 9-fold increase in the number of DOP images generated per unit of time, thereby providing more DOP images for decoder training, and allowing for a faster frame rate of imaging. Additionally, we demonstrate that NOfUS produces similar task-correlated activation maps through GLM analysis as the ground truth, confirming that NOfUS replicates spatially specific task-correlated activity. However, we caution against directly using NO-fUS for mapping behavior, such as through GLM analysis, due to the presence of noise and the potential for signal hallucination, as shown in (fig. 5d). Despite this limitation, NO-fUS is a robust method for constructing DOP images from less data (reducing both the collection wait time and the original frame rate) without significantly compromising critical behavioral information. These findings demonstrate that NOfUS has potential to improve upon real-time fUS-based BCI's by providing increased imaging frame rate and increased image samples for decoder training over the same period of time as traditional fUS using fewer data frames.

It is important to note that NO-fUS does not enhance image quality or decodability relative to the standard being currently used but is rather a supplementary tool that can be used to improve the computational efficiency of fUS acquisition and processing by reducing the frame rate and collection wait time. NO-fUS offers an approximated reconstruction of the ground-truth DOP image but is not a perfect replication. Similar to any data-driven approach, it also has the potential to introduce artifacts in signal over time, as seen during GLM analysis. Additionally, NO-fUS is a deep-learning framework; hence, it requires a training stage and a dataset consisting of IQ/DOP pairs. In the case, where we guide the NO-fUS using an approximated input DOP, NO-fUS may show poor performance if the quality of the approximated DOP is low. Despite these limitations, NOfUS still presents large potential in advancing fUS neurotechnology as a supplementary package.

To conclude, NO-fUS' ability to construct high-quality DOP images from reduced compounded IQ frames and a reduced frame rate allows researchers to use less resources to provide similar quality fUS images. This presents a crucial advancement for the future of fUS neurotechnology. While current parameters for fUS imaging operate at an acceptable temporal resolution and safety index, as fUS requires more data such as for volumetric imaging, computational and storage efficiency will become a major limiting factor. The large number of pulses used during 3D imaging may put subjects at risk of probe heating, and the vastly large data will be a huge challenge to store. The proposed NO-fUS architecture has the potential to be adapted for such advancements to reduce the number and rate of pulses and data needed per DOP image, reducing storage and heating concerns. As such, NO-fUS not only can be applied to advance current fUS neuroimaging but, in fact, has an even larger potential for improving fUS imaging in the future.

References

[1] J. L. Collinger, B. Wodlinger, J. E. Downey, W. Wang, E. C. Tyler-Kabara, D. J. Weber, A. J. C. McMorland, M. Velliste, M. L. Boninger, and A. B. Schwartz, "High-performance neuroprosthetic control by an individual with tetraplegia," vol. 381, no. 9866, pp. 557–564.

[2] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," vol. 442, no. 7099, pp. 164–171.

[3] D. A. Bjånes, S. Kellis, R. Nickl, B. Baker, T. Aflalo, L. Bashford, S. Chivukula, M. S. Fifer, L. E. Osborn, B. Christie, B. A. Wester, P. A. Celnik, D. Kramer, K. Pejsa, N. E. Crone, W. S. Anderson, N. Pouratian, B. Lee, C. Y. Liu, F. Tenore, L. Rieth, and R. A. Andersen, "Quantifying physical degradation alongside recording and stimulation performance of 980 intracortical microelectrodes chronically implanted in three humans for 956–2246 days," p. 2024.09.09.24313281.

[4] S. Suner, M. Fellows, C. Vargas-Irwin, G. Nakata, and J. Donoghue, "Reliability of signals from a chronically implanted, silicon-based electrode array in non-human primate primary motor cortex," vol. 13, no. 4, pp. 524–541. Conference Name: IEEE Transactions on Neural Systems and Rehabilitation Engineering.

[5] J. E. Downey, N. Schwed, S. M. Chase, A. B. Schwartz, and J. L. Collinger, "Intracortical recording stability in human brain–computer interface users," vol. 15, no. 4, p. 046016. Publisher: IOP Publishing.

[6] E. Macé, G. Montaldo, I. Cohen, M. Baulac, M. Fink, and M. Tanter, "Functional ultrasound imaging of the brain," *Nature methods*, vol. 8, no. 8, pp. 662–664, 2011.

[7] T. Deffieux, C. Demené, and M. Tanter, "Functional ultrasound imaging: A new imaging modality for neuroscience," *Neuroscience*, vol. 474, pp. 110–121, 2021. Brain imaging.

[8] S. Soloukey, A. J. P. E. Vincent, D. D. Satoer, F. Mastik, M. Smits, C. M. F. Dirven, C. Strydis, J. G. Bosch, A. F. W van der Steen, C. I. De Zeeuw, S. K. E. Koekkoek, and P. Kruizinga, "Functional ultrasound (fus) during awake brain surgery: The clinical potential of intra-operative functional and vascular brain mapping," *Frontiers in Neuroscience*, vol. 13, p. 1384, 2020.

[9] C. Demene, J. Baranger, M. Bernal, C. Delanoe, S. Auvin, V. Biran, M. Alison, J. Mairesse, E. Harribaud, M. Pernot, M. Tanter, and O. Baud, "Functional ultrasound imaging of brain activity in human newborns," vol. 9, no. 411, p. eaah6756.

[10] J. Baranger, C. Demene, A. Frerot, F. Faure, C. Delanoë, H. Serroune, A. Houdouin, J. Mairesse, V. Biran, O. Baud, and M. Tanter, "Bedside functional monitoring of the dynamic brain connectivity in human neonates," vol. 12, no. 1, p. 1080. Publisher: Nature Publishing Group.

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

[12] S. L. Norman, D. Maresca, V. Christopoulos, W. S. Griggs, C. Demene, M. Tanter, M. G. Shapiro, *et al.*, "Single-trial decoding of movement intentions using functional ultrasound neuroimaging," *Neuron*, vol. 109, pp. 1554–1566.e4, 2021.

[13] W. S. Griggs, S. L. Norman, T. Deffieux, F. Segura, B.-F. Osmanski, G. Chau, V. Christopoulos, C. Liu, M. Tanter, M. G. Shapiro, *et al.*, "Decoding motor plans using a closed-loop ultrasonic brain–machine interface," *Nature Neuroscience*, vol. 27, no. 1, pp. 196–207, 2024.

[14] C. Rabut, S. L. Norman, W. S. Griggs, J. J. Russin, K. Jann, V. Christopoulos, C. Liu, R. A. Andersen, and M. G. Shapiro, "Functional ultrasound imaging of human brain activity through an acoustically transparent cranial window," *Science Translational Medicine*, vol. 16, no. 749, p. eadj3143, 2024.

[15] H. Zheng, L. Niu, W. Qiu, D. Liang, X. Long, G. Li, Z. Liu, and L. Meng, "The emergence of functional ultrasound for noninvasive brain–computer interface," *Research*, vol. 6, p. 0200, 2023.

[16] A. Fort, C. Manfredi, and S. Rocchi, "Adaptive svd-based ar model order determination for time-frequency analysis of doppler ultrasound signals," *Ultrasound in medicine & biology*, vol. 21, no. 6, pp. 793–805, 1995.

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

[18] G. Montaldo, M. Tanter, J. Bercoff, N. Benech, and M. Fink, "Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, no. 3, pp. 489–506, 2009.

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

[20] T. Di Ianni and R. D. Airan, "Deep-fus: A deep learning platform for functional ultrasound imaging of the brain using sparse data," *IEEE transactions on medical imaging*, vol. 41, no. 7, pp. 1813–1825, 2022.

[21] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, "Neural operator: Graph kernel network for partial differential equations," *arXiv preprint arXiv:2003.03485*, 2020.

[22] Z. Li, N. B. Kovachki, K. Azizzadenesheli, B. liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, "Fourier neural operator for parametric partial differential equations," in *International Conference on Learning Representations*, 2021.

[23] N. B. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. M. Stuart, and A. Anandkumar, "Neural operator: Learning maps between function spaces with applications to pdes.," *J. Mach. Learn. Res.*, vol. 24, no. 89, pp. 1–97, 2023.

[24] K. Azizzadenesheli, N. Kovachki, Z. Li, M. Liu-Schiaffini, J. Kossaifi, and A. Anandkumar, "Neural operators for accelerating scientific simulations and design," *Nature Reviews Physics*, pp. 1–9, 2024.

[25] Z. Li, N. Kovachki, C. Choy, B. Li, J. Kossaifi, S. Otta, M. A. Nabian, M. Stadler, C. Hundt, K. Azizzadenesheli, *et al.*, "Geometry-informed neural operator for large-scale 3d pdes," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[26] T. Kurth, S. Subramanian, P. Harrington, J. Pathak, M. Mardani, D. Hall, A. Miele, K. Kashinath, and A. Anandkumar, "Fourcastnet: Accelerating global high-resolution weather forecasting using adaptive fourier neural operators," in *Proceedings of the Platform for Advanced Scientific Computing Conference*, pp. 1–11, 2023.

[27] Y. Yang, A. F. Gao, J. C. Castellanos, Z. E. Ross, K. Azizzadenesheli, and R. W. Clayton, "Seismic wave propagation and inversion with neural operators," *The Seismic Record*, vol. 1, no. 3, pp. 126–134, 2021.

[28] H. Yang, Z. Li, K. Sastry, S. Mukhopadhyay, M. Kilgard, A. Anandkumar, B. Khailany, V. Singh, and H. Ren, "Generic lithography modeling with dual-band optics-inspired neural networks," in *Proceedings of the 59th ACM/IEEE Design Automation Conference*, pp. 973–978, 2022.

[29] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pp. 234–241, Springer, 2015.

[30] M. A. Rahman, Z. E. Ross, and K. Azizzadenesheli, "U-no: U-shaped neural operators," *Transactions on Machine Learning Research*, 2023.

[31] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE transactions on image processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[32] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 1–37, 2011.

[33] O. Solomon, R. Cohen, Y. Zhang, Y. Yang, Q. He, J. Luo, R. J. van Sloun, and Y. C. Eldar, "Deep unfolded robust pca with application to clutter suppression in ultrasound," *IEEE transactions on medical imaging*, vol. 39, no. 4, pp. 1051–1063, 2019.

[34] J. S. Yoon, K. Oh, Y. Shin, M. A. Mazurowski, and H.-I. Suk, "Domain generalization for medical image analysis: A review," *Proceedings of the IEEE*, 2024.

[35] J. Liu, Z. Shen, Y. He, X. Zhang, R. Xu, H. Yu, and P. Cui, "Towards out-of-distribution generalization: A survey," *arXiv preprint arXiv:2108.13624*, 2021.

[36] L. Lu, P. Jin, and G. E. Karniadakis, "Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators," *arXiv preprint arXiv:1910.03193*, 2019.

[37] H.-S. Choi, J.-H. Kim, J. Huh, A. Kim, J.-W. Ha, and K. Lee, "Phase-aware speech enhancement with deep complex u-net," in *International Conference on Learning Representations*, 2018.

[38] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," *Advances in neural information processing systems*, vol. 32, 2019.

[39] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A general u-shaped transformer for image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 17683–17693, 2022.

[40] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (red)," *SIAM Journal on Imaging Sciences*, vol. 10, no. 4, pp. 1804–1844, 2017.

[41] S. Bjaerum, H. Torp, and K. Kristoffersen, "Clutter filters adapted to tissue motion in ultrasound color flow imaging," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 49, no. 6, pp. 693–704, 2002.

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