Ultrasound in Med. & Biol., Vol. 49, No. 3, pp. 677–698, 2023

0301-5629/$ - see front matter

https://doi.org/10.1016/j.ultrasmedbio.2022.11.003

● *Review*

ULTRASOUND SIGNAL PROCESSING: FROM MODELS TO DEEP LEARNING

Ben Luijten,* Nishith Chennakeshava,* Yonina C. Eldar,†
Massimo Mischi,* and Ruud J.G. van Sloun*,‡

* Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands; † Faculty of Math and Computer Science, Weizmann Institute of Science, Rehovot, Israel; and ‡ Philips Research, Eindhoven, The Netherlands

(*Received* 10 *March* 2022; *revised* 2 *November* 2022; *in final form* 5 *November* 2022)

Abstract—Medical ultrasound imaging relies heavily on high-quality signal processing to provide reliable and interpretable image reconstructions. Conventionally, reconstruction algorithms have been derived from physical principles. These algorithms rely on assumptions and approximations of the underlying measurement model, limiting image quality in settings where these assumptions break down. Conversely, more sophisticated solutions based on statistical modeling or careful parameter tuning or derived from increased model complexity can be sensitive to different environments. Recently, deep learning–based methods, which are optimized in a data-driven fashion, have gained popularity. These model-agnostic techniques often rely on generic model structures and require vast training data to converge to a robust solution. A relatively new paradigm combines the power of the two: leveraging data-driven deep learning and exploiting domain knowledge. These model-based solutions yield high robustness and require fewer parameters and training data than conventional neural networks. In this work we provide an overview of these techniques from the recent literature and discuss a wide variety of ultrasound applications. We aim to inspire the reader to perform further research in this area and to address the opportunities within the field of ultrasound signal processing. We conclude with a future perspective on model-based deep learning techniques for medical ultrasound. (E-mail: w.m.b.luijten@tue.nl)

Key Words: **Ultrasound, Deep learning, Probabilistic modeling.**

INTRODUCTION

Ultrasound (US) imaging has proven itself to be an invaluable tool in medical diagnostics. Among many imaging technologies, such as X-ray, computed tomography (CT) and magnetic resonance imaging (MRI), US uniquely positions itself as an interactive diagnostic tool, providing real-time spatial and temporal information to the clinician. Combined with its relatively low cost, compact size and absence of ionizing radiation, US imaging is an increasingly popular choice in patient monitoring.

Consequently, the versatility of US imaging has spurred a wide range of applications in the field. While conventionally it is used for the acquisition of B-mode (2-D) images, more recent developments have enabled ultrafast and 3-D volumetric imaging. Additionally, US devices can be used for measuring clinically relevant features such as blood velocity (Doppler), tissue characteristics (*e.g.,* elastography maps) and perfusion through ultrasound localization microscopy (ULM). Although this wide range of applications shares the same underlying measurement steps—acquisition, reconstruction and visualisation—their signal processing pipelines are often specific for each application.

It follows that the quality of US imaging strongly depends on the implemented signal processing algorithms. The resulting demand for high-quality signal processing has pushed the reconstruction process from fixed, often hardware-based implementations to the digital domain (Kim et al., 1997, Thomenius, 1996). More recently, this has led to fully software-based algorithms, as they can open up the potential to complex measurement models and statistical signal interpretations. However, this shift has also posed a new set of challenges, as it places a significant strain on the digitisation hardware, bandwidth constrained data channels and computational

Address correspondence to: Ben Luijten, Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven 5612AZ, Netherlands. E-mail: w.m.b.luijten@tue.nl

capacity. As a result, clinical devices, where real-time imaging and robustness are of utmost importance, still rely mainly on simple hardware-based solutions.

A more recent development in this field is the utilization of deep neural networks. Such networks can provide fast approximations for signal recovery, and can often be efficiently implemented because of their exploitation of parallel processing. After training, these networks can be efficiently implemented to facilitate ultrafast signal processing. However, by inheriting generic network architectures from computer vision tasks, these approaches are highly data driven and are often over-parameterized, posing several challenges. To converge to a well-generalized solution across the full data distribution encountered in practice, large amounts of (unbiased) training data are needed, which are not always trivial to obtain. Furthermore, these models are often treated as a "black box," making it difficult to guarantee the correct behavior in a real clinical setting.

To overcome some of the challenges of purely data-driven methods, an alternative approach is to try to combine model-based and data-driven methods in an attempt to obtain the best of both worlds. The proposition here is that the design of data-driven methods for ultrasound signal processing can likely benefit from the vast amounts of research on conventional, *model-based*, reconstruction algorithms, informing, for example, specific neural network designs or hybrid processing approaches.

In this review, we aim to provide the reader with a comprehensive overview of ultrasound signal processing based on modeling, machine learning and model-based learning. To achieve this, we take a probabilistic perspective and place methods in the context of their assumptions on signal models and statistics, and training data. Although other works (Al Kassir et al., 2022, Liu et al., 2019, Monga et al., 2021, Shlezinger et al., 2020, van Sloun et al., 2019) offer an excellent overview of the different aspects of AI applied to ultrasound image processing, the focus of this review is to put the theory of both signal processing and machine learning under a unifying umbrella, rather than to showcase a general review of deep learning (DL) being applied to ultrasound-specific problems. To that end, we cover topics ranging from beamforming to post-processing and advanced applications such as super-resolution. Throughout the article we distinguish between three types of approaches that we cover in separate sections.

- *Model-based methods for US signal processing:* Conventional model-based methods derive signal processing algorithms by modeling the problem based on first principles, such as knowledge of the acquisition model, noise and signal statistics. Simple models offer analytical solutions, whereas more complex models often require iterative algorithms.
- *DL for US signal processing:* DL solutions are fully data driven and fit highly parameterized algorithms (in the form of deep neural networks) to data. DL methods are model-agnostic and thus rely on the training data to expose structure and relations between inputs and desired outputs.
- *Model-based DL for US signal processing:* Model-based DL aims at bridging the gap by deriving algorithms from first-principle models (and their assumptions) while learning parts of these models (or their analytic/iterative solutions) from data. These approaches enable incorporation of prior knowledge and structure (inductive biases) and offer tools for designing deep neural networks with architectures that are tailored to a specific problem and setting. The resulting methods resemble conventional model-based methods, but allow for overcoming mismatched or incomplete model information by learning from data.

In all cases, data are needed to test the performance of (clinical) signal processing algorithms. However, in DL-based solutions specifically, we observe an increasing need for training data when prior knowledge on the underlying signal model is not fully exploited. A schematic overview of these approaches is given in Figure 1, including examples of corresponding techniques in the case of ultrasound beamforming.

We begin by briefly explaining the probabilistic perspective and notation we adopt throughout the article in a preliminaries section, after which we provide background information on the basics of US acquisition, which can be skipped by experts in the field of ultrasound. Following this background information, we dive into model-based US signal processing, in which we derive various conventional beamforming and post-processing algorithms from their models and statistical assumptions. Next, we turn to DL methods, after which we bridge the gap between model-based and DL-based processing, identifying opportunities for data-driven enhancement of model-based methods (and their assumptions) by DL. Finally we provide a Discussion and Conclusions, where we present a future outlook and several opportunities for DL in ultrasound signal processing.

A PROBABILISTIC APPROACH TO DL IN US SIGNAL PROCESSING

In this article we use the language and tools of probability theory to seamlessly bridge the gap between conventional model-based signal processing and contemporary machine/DL approaches. As Shakir

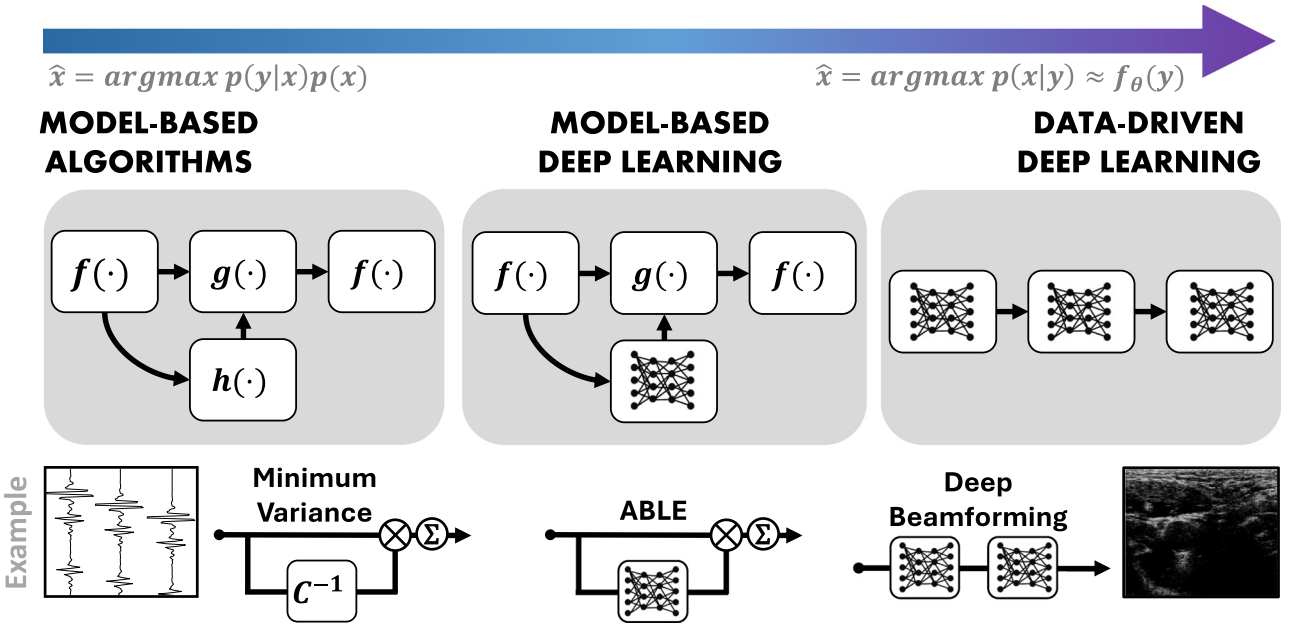

Fig. 1. Schematic overview of model-based and deep learning–based techniques in ultrasound signal processing. An example for each category is given in the context of ultrasound beamforming: minimum variance beamforming, adaptive beamforming by deep learning (ABLE) (Luijten et al., 2020) and deep beamforming (Khan et al., 2020).

Mohamed (DeepMind, https://www.shakirm.com/slides/MLSS2018-Madrid-ProbThinking.pdf) phrased it:

> "Almost all of machine learning can be viewed in probabilistic terms, making probabilistic thinking fundamental. It is, of course, not the only view. But it is through this view that we can connect what we do in machine learning to every other computational science, whether that be in stochastic optimisation, control theory, operations research, econometrics, information theory, statistical physics or bio-statistics. For this reason alone, mastery of probabilistic thinking is essential."

To that end, we begin by briefly reviewing some concepts in probabilistic signal processing based on models and then turn to recasting such problems as data-driven learning problems.

PRELIMINARIES ON MODEL-BASED PROBABILISTIC INFERENCE

Let us consider a general linear model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{y}$ is our observed signal, $\mathbf{A}$ a measurement matrix, $\mathbf{n}$ a noise vector and $\mathbf{x}$ the signal of interest. As we will see throughout the article, many problems in ultrasound signal processing can be described according to such linear models. In ultrasound beamforming for example, $\mathbf{y}$ may denote the measured (noisy) RF signals, $\mathbf{x}$ the spatial tissue reflectivity and A a matrix that transforms such a reflectivity map to channel domain signals. The goal of beamforming is then to infer $\mathbf{x}$ from $\mathbf{y}$ under the measurement model in eqn (1).

Recalling Bayes' rule, we can define the posterior probability of $\mathbf{x}$ given $\mathbf{y}$, as a product of the likelihood $p(\mathbf{y}|\mathbf{x})$ and a prior $p(\mathbf{x})$, such that

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \tag{2}$$

$$\propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \tag{3}$$

Following (3) we can define a maximum *a posteriori* (MAP) estimator for (1), given by

$$\hat{\mathbf{x}}_{\text{MAP}} := \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) = \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \tag{4}$$

which provides a single, most likely, estimate according to the posterior distribution. If we assume a Gaussian white noise vector $\mathbf{n}$ in eqn (1), that is, $\mathbf{y} \sim \mathcal{N}(\mathbf{A}\mathbf{x}|\sigma_n^2 I)$, the MAP estimator becomes

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \| \mathbf{y} - \mathbf{A}\mathbf{x} \|_2^2 - \lambda \log p(\mathbf{x}) \tag{5}$$

where λ is a scalar regularization parameter.

Evidently, the MAP estimator takes the prior density function $p(\mathbf{x})$ into account. In other words, it allows us to incorporate and exploit prior information on $\mathbf{x}$,

should this be available. Conversely, if x is assumed to be deterministic but unknown, we get the maximum likelihood (ML) estimator. The ML estimator thus assigns equal likelihood to each $\mathbf{x}$ in the absence of measurements. As such this simplifies to

$$\widehat{\mathbf{x}}_{\mathrm{ML}} := \arg\max_x p(\mathbf{y}|\mathbf{x}) \tag{6}$$

Many traditional ultrasound processing methods are in this form, where its output only depends on a set of (finely tuned) hyperparameters and the input data. This is not surprising, as deriving a strong and useful prior that generalizes well to the entire expected data distribution is challenging in its own right.

Data-driven approaches aim to overcome the challenges of accurate modeling by learning the likelihood function, the prior, the entire posterior or a direct end-to-end mapping (replacing the complete MAP estimator) from data. We will detail on these methods in the following section.

Preliminaries on DL-based inference

Fully data-driven methods aim at learning the optimal parameters θ^* of a generic parameterized mapping, $f_\theta(\cdot)$, $Y \to X$, from training data. In DL, the mapping function $f_\theta(\cdot)$ is a deep neural network. Learning itself can also be formulated as a probabilistic inference problem, where optimized parameter settings for a fixed network architecture are inferred from a data set $\mathscr{D}$. To that end we define a posterior over the parameters

$$p(\theta|\mathscr{D}) = \frac{p(\mathscr{D}|\theta)p(\theta)}{p(\mathscr{D})} \tag{7}$$

$$\propto p(\mathscr{D}|\theta)p(\theta) \tag{8}$$

where $p(\theta)$ denotes a prior over the parameters. Often $p(\theta)$ is fully factorized, that is, each parameter is assumed independent, to keep the learning problem in deep networks (with millions of parameters) tractable. Typical priors are Gaussian or Laplacian density functions.

Most DL applications rely on MAP estimation to find the set of parameters that minimize the negative log posterior:

$$\theta^* = \arg\max_\theta p(\theta|\mathscr{D}) = \arg\max_\theta \log p(\mathscr{D}|\theta)p(\theta) \tag{9}$$

$$= \arg\min_\theta -\{\log p(\mathscr{D}|\theta) + \log p(\theta)\} \tag{10}$$

Note that for measurement (input)−signal (output) training pairs, $(\mathbf{y}_i, \mathbf{x}_i) \in \mathscr{D}$, common forms of $p(\mathbf{x}|f_\theta(\mathbf{y}), \theta)$ are Gaussian, Laplacian or categorical distributions, resulting in mean squared error, mean absolute error and cross-entropy negative log-likelihood functions, respectively. Similarly, Gaussian and Laplacian priors lead to ℓ_2 and ℓ_1 regularization on the parameters, respectively.

It is worth noting that although most DL applications perform MAP estimation, there is increasing interest in so-called Bayesian DL, which aims at learning the parameters of the prior distribution $p(\theta)$ as well. This enables posterior sampling during inference (by sampling from $p(\theta)$) for (epistemic) uncertainty estimation. Again, often these distributions are fully factorized (*e.g.,* independent Gaussian or Bernoulli) to make the problem tractable (Gal and Ghahramani, 2016).

After training (*i.e.,* inferring parameter settings), we can use the network to perform MAP inference to retrieve $\mathbf{x}$ from new input measurements $\mathbf{y}$:

$$\widehat{\mathbf{x}} = \arg\max_{\mathbf{x}} p\Big(\mathbf{x}|f_\theta(\mathbf{y}), \theta\Big). \tag{11}$$

The neural network thus directly models the parameters of the posterior, and does not factorize it into a likelihood and prior term as model-based MAP inference does. Note that for Gaussian and Laplace density functions, a neural network $f_\theta(\mathbf{y})$ computes the distribution mean, $\arg\max_{\mathbf{x}} p(\mathbf{x}|f_\theta(\mathbf{y}), \theta) = f_\theta(\mathbf{y})$. For categorical distributions, $f_\theta(\mathbf{y})$ computes the probabilities for each category/class.

Typical deep neural network parameterizations $f_\theta(\cdot)$ are therefore model-agnostic, as they disregard the structure of the measurement/likelihood model and prior and offer a high degree of flexibility to fit many data distributions and problems. However, many such parameterizations do exploit specific symmetries in the expected input data. Examples are convolutional neural networks, which exploit the spatial shift-invariant structure of many image classification/regression problems through shift-equivariant convolutional layers. Similarly, many applications in which the input is temporally correlated, such as time series analysis, recurrent neural networks (RNN) are employed.

Preliminaries on model-based DL

Model-based DL aims at imposing much more structure to the network architectures and parameterizations of $f_\theta(\cdot)$. Where standard deep networks aim at fitting a broad class of problems, model-based DL offers architectures that are highly tailored to specific inference problems given in eqns (1) and (4); that is, they are aware of the model and structure of the problem. This promises to relax challenges related to generalization, robustness and interpretability in DL. It often also enables design of smaller (but more specialized) networks with a lower computational and memory footprint.

To derive a model-based DL method, one can start by deriving a MAP estimator for $\mathbf{x}$ from the model, including assumptions on likelihood models $p(\mathbf{y}|\mathbf{x})$ and priors $p(\mathbf{x})$. Generally, such estimators come in two forms: *analytic* (direct) and *iterative* solutions. The

solution structure dictates the neural network architecture. One then has to select which parts of the original model-based graph are to be replaced by learnable functions.

One of the first examples of model-based DL is the learned iterative-shrinkage and thresholding algorithm (LISTA), proposed by Gregor and LeCun (Gregor and LeCun, 2010). As the name suggests, $f_\theta(\cdot)$ is based on an iterative solution, specifically to the MAP sparse coding problem: arg $\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$, with $\mathbf{x}\text{Laplace}(0, b\mathbf{I})$, where b is a scale parameter, and $\mathbf{y}|\mathbf{x} \sim \mathcal{N}(\mathbf{A}\mathbf{x}, \sigma^2\mathbf{I})$. This iterative solution consists of two alternating steps: (i) a gradient step on $\mathbf{x}$ to maximize the log-likelihood of $\log p(\mathbf{y}|\mathbf{x})$, and (ii) a proximal step that moves the intermediate solution for $\mathbf{x}$ toward higher log-likelihood under the prior distribution $\log p(\mathbf{x})$. The model-based DL method LISTA *unfolds* or *unrolls* a limited number of algorithm iterations to form a feed-forward neural network, learning the parameters of the gradient step and the proximal step end-to-end from training examples $(\mathbf{y}_i, \mathbf{x}_i \in \mathcal{D})$, without knowledge on the underlying distribution of these parameters. Moreover, LISTA has been shown to accelerate inference compared with the iterative algorithm.

Today, model-based DL is a rapidly developing field, with many variations of such model-based DL methods being developed for various problems and applications (Monga et al., 2021). One can learn optimal gradient steps, adopt neural-network-based proximal operators and include deep generative models for iterative/unfolded inference, thereby overcoming the limitations of naive assumptions in model-based methods. Similarly, for analytical solutions, one can replace computations that rely on accurate model knowledge (that may be unavailable) or are challenging/time-consuming to compute by neural networks.

Also within the field of US imaging and signal processing, model-based DL is seeing increasing adoption for problems spanning from beamforming (Luijten et al., 2020) to clutter suppression (Solomon et al., 2019) and localization microscopy (van Sloun et al., 2019). Exact implementations of these model-based DL methods for US imaging are indeed highly application specific (which is its merit), as we discuss in a later section.

FUNDAMENTALS OF US ACQUISITION

Ultrasound imaging is based on the pulse-echo principle. First, a pressure pulse is transmitted toward a region of interest by the US transducer consisting of multiple transducer elements. Within the medium, scattering occurs because of inhomogeneities in density, speed-of-sound and non-linear behavior. The resulting back-scattered echoes are recorded using the same transducer, yielding a set of radiofrequency (RF) channel signals that can be processed. Typical ultrasound signal processing includes B-mode image reconstruction via beamforming, velocity estimation (Doppler) and additional downstream post-processing and analysis.

Although the focus of this article is on these processing methods, which we discuss in later sections, we will for the sake of completeness briefly review the basic principles of ultrasound channel signal acquisition.

Transmit schemes

Consider an ultrasound transducer with channels $c \in C$. A transmit scheme consists of a series of transmit events $e \in E$. Different transmit events can be constructed by adjusting the per channel transmit delays (focusing), the number of active channels (aperture) and, in advanced modes, also waveform parameters. We briefly list the most common transmit schemes.

Line scanning

Most commercial ultrasound devices rely on focused, line-by-line, acquisition schemes, as these yield superior resolution and contrast compared with unfocused strategies. In line scanning, a subaperture of channels focuses the acoustic energy by channel-dependent transmit delays along a single (axial) path at a set depth, maximizing the reflected echo intensity in a region of interest (Ding et al., 2014). Some transmit schemes make use of multiple foci per line. To cover the full lateral field of view, many scan lines are needed, limiting the overall frame rate.

Synthetic aperture

In synthetic aperture (SA) imaging, each channel transmit−receive pair is acquired separately (Jensen et al., 2006, Ylitalo and Ermert, 1994). To that end, each element independently fires a spherical wavefront, the reflections of which can be simultaneously recorded by all receiving elements. Typically, the number of transmit events is equal to the number of transducer elements ($E = C$). Having access to these individual transmit−receive pairs enables retrospective transmit focusing to an arbitrary set of foci (*e.g.,* each pixel). Although SA imaging offers advantages in terms of receive processing, it is time consuming, similar to line scanning. Furthermore, single elements generate low acoustic energy, which reduces the signal-to-noise ratio (SNR).

Plane and diverging wave

Recently, unfocused (parallel) acquisition schemes have become more popular, as they can drastically reduce acquisition times, yielding so-called ultrafast imaging at very high frame rates. Plane wave (PW)

imaging insonifies the entire region of interest at once through a planar wave field, by firing with all elements and placing the axial focus point at infinity. Diverging wave (DW) transmissions also insonify the entire region of interest in one shot, but generate a spherical (diverging) wavefront by placing a (virtual) focus point behind the transducer array. Especially for small transducer footprints (*e.g.,* phased array probes), DW schemes are useful to cover a large image region.

Both PW imaging and DW imaging suffer from deteriorated resolution and low contrast (high clutter) because of strong interference by scattering from all directions. Often, multiple transmits at different angles are therefore compounded to boost image quality. However, this reduces frame rate. Unfocused transmissions rely heavily on the powerful receive processing to yield an image of sufficient quality, raising computational requirements.

Doppler

Beyond positional information, ultrasound also permits measurement of velocities, useful in the context of, for example, blood flow imaging or tissue motion estimation. This imaging mode, called Doppler imaging (Chan and Perlas, 2011, Hamelmann et al., 2019, Routh, 1996), often requires dedicated transmit schemes with multiple high-rate sequential acquisitions. Continuous wave Doppler allows for simultaneous transmit and receive of acoustic waves using separate sub-apertures. Although this yields a high temporal sampling rate, and prevents aliasing, it does result in some spatial ambiguity. The entire region of overlap between the transmit and receive beams contributes to the velocity estimate. Alternatively, pulsed-wave Doppler relies on a series of snapshots of the slow-time signal, with the temporal sampling rate being equal to the frame rate. From these measurements, a more confined region of interest can be selected for improved position information, at the cost of possible aliasing.

Waveform and frequency

The resolution that can be obtained using ultrasound is in large part dependent on the frequency of the transmitted pulse. High transmit pulse frequencies and short pulse durations yield high spatial resolution, but are strongly affected by attenuation. This becomes especially problematic in deep tissue regions. As a general rule, the smallest measurable structures scale to approximately half the wavelength of the transmit frequency, that is, the diffraction limit. In practice the transmit pulse spans multiple wavelengths, which additionally limits axial resolution by half the transmit pulse length. Design choices such as transducer array aperture, element sensitivity, bandwidth of the front-end circuitry and reconstruction algorithms also play a dominant role in this.

Array designs

Depending on the application, different transducer types may be preferred. Either because of physical constraints or from having desirable imaging properties. Commonly used transducer geometries include linear, convex and phased arrays. Effectively, the transducer array, consisting of elements, spatially samples the array response. Typically, these array elements have a center-to-center spacing (pitch) of $\lambda/2$ or less, to avoid spatial aliasing. In general, a larger number of elements yields a better resolution image, but this consequently increases size, complexity and bandwidth requirements. Especially for 2-D arrays (used in 3-D imaging), the large number of transducer elements can be problematic in implantation because of the vast number of physical hardware connections. Other than translating to an increase in cost and complexity, it also raises power consumption. In those cases, often some form of micro-beamforming is applied in the front end, combining individual channel signals early in the signal chain.

Similar reductions in data rates can be achieved through sub-sampling of the receive channels. Trivial approaches include uniform or random sub-sampling, at the cost of reduced resolution, and more pronounced aliasing artifacts (grating lobes). Several works have reported that these effects can be mitigated either by principled array designs (Cohen and Eldar, 2020, Song et al., 2020) or by learning sub-sampling patterns from data in a task-adaptive fashion (Huijben et al., 2020).

Sub-Nyquist signal sampling

Digital signal processing of US signals requires sampling of the signals received by the transducer, after which the digital signal is transferred to the processing unit. To prevent frequency-aliasing artifacts, sampling at the Nyquist limit is necessary. In practice, sampling rates $4-10$ times higher are common, as they allow for a finer resolution during digital focusing. As a consequence, this leads to high bandwidth data streams, which become especially problematic for large transducer arrays (*e.g.,* 3-D probes).

Compressed sensing (CS) provides a framework that allows for reduced data rates, by sampling below the Nyquist limit, alleviating the burden on data transfer (Eldar, 2015). CS acquisition methods provide strong signal recovery guarantees when complemented with advanced processing methods for reconstruction of the signal of interest. These reconstruction methods are typically based on MAP estimation, combining likelihood models on the measured data (*i.e.,* a measurement matrix) with priors on signal structure (*e.g.,* sparsity in some basis). Many of the signal processing algorithms that we list throughout the article will find application

within a CS context, especially those methods that introduce a signal prior for reconstruction, either through models or by learning from data. The latter is especially useful for elaborate tasks where little is known about the distribution of system parameters, offering signal reconstruction beyond what is possible using conventional CS methods.

For further reading into the fundamentals of ultrasound, the reader may refer to works such as Brahme (Brahme, 2014).

Model-based US signal processing

Model-based ultrasound signal processing techniques are based on first principles such as the underlying physics of the imaging setup or knowledge of the statistical structure of the signals. We now describe some of the most commonly used model-based ultrasound signal processing techniques, building on the probabilistic perspective sketched in earlier sections. For each algorithm, we explicitly list (i) inputs and outputs (and dimensions), (ii) the assumed signal model and statistics, (iii) signal priors and (iv) the resulting ML/MAP objective and solution.

Beamforming, the act of reconstructing an image from the received raw RF channel signals, is central to ultrasound imaging and typically the first step in the signal processing pipeline. We thus start our description with beamforming methods.

Beamforming

Given an ultrasound acquisition of C transducer channels, N_t axial samples and E transmission events, we can denote $Y \in \mathrm{R}^{E\times C\times N_t}$ as the recorded RF data cube, representing backscattered echoes from each transmission event. With beamforming, we aim to transform the raw aperture domain signals Y to the spatial domain through a processing function $f(\cdot)$ such that

$$\widehat{X} = f(Y) \tag{12}$$

where $\widehat{X}$ represents the data beamformed to a set of focus points $S_{\mathbf{r}}$. As an example, in pixel-based beamforming, these focus points could be a pixel grid such that $S_{\mathbf{r}} \in \mathrm{R}^{r_x\times r_z}$, where r_x and r_y represent the lateral and axial components of the vector indicating the pixel coordinates, respectively. Note that although this example is given in Cartesian coordinates, beamforming to other coordinate systems (*e.g.,* polar coordinates) is also common.

Delay-and-sum beamforming

Delay-and-sum (DAS) beamforming has been the backbone of ultrasound image reconstruction for decades. This is driven mainly by its low computational complexity, which allows for real-time processing, and efficient hardware implementations. In DAS, to reconstruct a tissue reflectivity image (B-mode), the aperture domain signals are first migrated back to the image domain in a process called time-of-flight (TOF) correction. This transformation is based on the back-projection of the time-domain signals, and aims at aligning the received signals for a set of focus points (in pixel-based beamforming: pixels) by applying time delays. We can define the total TOF from transmission to the receiving element as

$$\boldsymbol{\tau}_{\mathbf{r}} = \boldsymbol{\tau}[r_x, r_z] = \frac{||\mathbf{r}_e - \mathbf{r}||_2 + ||\mathbf{r}_c - \mathbf{r}||_2}{v} \tag{13}$$

where $\boldsymbol{\tau}_{\mathbf{r}}$ is the channel delay required to focus to an imaging point $\mathbf{r}$, vectors $\mathbf{r}_e$ and $\mathbf{r}_c$ correspond to the origin of the transmit event e, and the position of element c, respectively, and v is the speed-of-sound in the medium. Note that the speed-of-sound is generally assumed to be constant throughout the medium. As a consequence, speed-of-sound variations can cause misalignment of the channel signals and result in aberration errors.

After TOF correction, we obtain a channel vector $\mathbf{y}_{\mathbf{r}}$ per pixel $\mathbf{r}$, for which we can define a linear forward model to recover the pixel reflectivity $x_{\mathbf{r}}$:

$$\mathbf{y}_{\mathbf{r}} = 1_{\mathbf{r}} x_{\mathbf{r}} + \mathbf{n}_{\mathbf{r}} \tag{14}$$

Here, $\mathbf{y}_{\mathbf{r}} \in \mathrm{R}^C$ is a vector containing the received aperture signals, $x_{\mathbf{r}} \in \mathrm{R}$ is the tissue reflectively at a single focus point $\mathbf{r}$ and $\mathbf{n}_{\mathbf{r}} \in \mathrm{R}^{C\times 1}$ is an additive Gaussian noise vector $\sim \mathcal{N}(0, \sigma_n^2\mathbf{I})$. In this simplified model, all interference (*e.g.,* clutter, off-axis scattering, thermal noise) is contained in $\mathbf{n}$. Note that (without loss of generality) we assume a real-valued array response in our analysis, which can be straightforwardly extended to complex values (*e.g.,* after in-phase and quadrature demodulation). Under the Gaussian noise model, eqn (14) yields the likelihood model for the channel vector

$$p(\mathbf{y}_{\mathbf{r}}|x_{\mathbf{r}}) = \frac{1}{\sqrt{(2\pi\sigma_n^2)}} \exp\left[-\frac{(\mathbf{y}_{\mathbf{r}} - 1_{\mathbf{r}} x_{\mathbf{r}})^2}{2\sigma_n^2}\right] \tag{15}$$

where σ_n^2 denotes the noise power.

The delay-and-sum beamformer is the per-pixel ML estimator of the tissue reflectively, $\widehat{x}_{\mathbf{r}}$, given by

$$\widehat{x}_{\mathbf{r}} := \arg\max_{x_{\mathbf{r}}} \log p(\mathbf{y}_{\mathbf{r}}|x_{\mathbf{r}}) \tag{16}$$

$$= \arg\max_{x_{\mathbf{r}}} (\mathbf{y}_{\mathbf{r}} - 1_{\mathbf{r}} x_{\mathbf{r}})^H (\mathbf{y}_{\mathbf{r}} - 1_{\mathbf{r}} x_{\mathbf{r}}) \tag{17}$$

Solving eqn (17) yields

$$\widehat{x}_{\mathbf{r}} = \frac{1}{C} 1^H \mathbf{y}_{\mathbf{r}} = \frac{1}{C} \sum_{c=1}^{C} y_c \tag{18}$$

where C is the number of array elements. In practice, apodization/tapering weights are included to suppress side lobes:

$$\widehat{x}_{\mathbf{r}} = \frac{1}{C}\mathbf{w}^H\mathbf{y}_{\mathbf{r}} = \frac{1}{C}\sum_{c=1}^{C} w_c y_c \tag{19}$$

This form can be recognized as the standard definition of DAS beamforming, in which the channel signals are weighed using an apodization function, $\mathbf{w}$, and subsequently summed to yield a beamformed signal.

To promote either higher resolution or contrast, further adjustments can be made to the apodization function. However, this always poses an inherent compromise between main lobe width and side lobe intensity or, equivalently, resolution and contrast. This can be attributed to the beamformer, which is a spatial filter characterized by the aperture pitch (sampling rate), aperture size (filter length) and apodization weights (filter coefficients). Typical choices for apodization weights are tapered functions, such as Hanning or Hamming windows, in which elements further away from the focus point are weighed less then the elements close by. In commercial devices, these apodization weights are often finely tuned to optimize image quality for each transducer type and imaging setting, and may vary spatially (per focus point). Similar to DAS for aggregating channels, different transmit events can be aggregated, known as coherent compounding.

Frequency domain beamforming

In the previous section, all processing was done in the time domain. Alternatively, the TOF correction can be implemented in the frequency domain. This has several benefits. First, it avoids the need for oversampling channel data at high rates to enable accurate time-domain delays. Second, it facilitates sub-Nyquist acquisition schemes that sample only the most important Fourier coefficients (*e.g.,* via sum of sinc filters). In this context, Chernyakova and Eldar (Chernyakova and Eldar, 2014) proposed a Fourier-domain beamformer (FDBF), which allows for a substantial reduction in sampling rates while maintaining image quality. Denote by $\omega[k]$ the kth Fourier series coefficient of a beamformed signal. The Fourier transform of the time-aligned signals is defined as

$$\widehat{\omega}_c[k] = \sum \omega_c[k-n]Q_{k,c,\mathbf{r}} \tag{20}$$

Here, $Q_{k,c,\mathbf{r}}$ are the Fourier coefficients of a distortion function derived from the beamforming delays at $\mathbf{r}$, as in eqn (13).

When not all Fourier coefficients are sampled (*i.e.,* in sub-Nyquist acquisition), the desired time-domain signal can be recovered using CS methods such as NESTA (Becker et al., 2011) or via DL approaches.

Advanced adaptive beamforming

The shortcomings of standard DAS beamforming have spurred the development of a wide range of adaptive beamforming algorithms. These methods aim to overcome some of the limitations that DAS faces by adaptively tuning its processing based on the input signal statistics.

Minimum variance

Delay-and-sum beamforming is the ML solution of eqn (14) under white Gaussian noise. To improve realism for more structured noise sources, such as off-axis interference, we can introduce a colored (correlated) Gaussian noise profile $\mathbf{n}\tilde{N}(0,\Gamma_n)$, with $\Gamma_{\mathbf{r}}$ being the array covariance matrix for beamforming point $\mathbf{r}$. Maximum (log) likelihood estimation for $x_{\mathbf{r}}$ then yields

$$\widehat{x}_{\mathbf{r}} = \arg\max_{x_{\mathbf{r}}} \log p(\mathbf{y}_{\mathbf{r}}|x_{\mathbf{r}},\Gamma_{\mathbf{r}}) \tag{21}$$

$$= \arg\min_{x_{\mathbf{r}}} (\mathbf{y}_{\mathbf{r}} - \mathbf{1}x_{\mathbf{r}})^H \Gamma_{\mathbf{r}}^{-1} (\mathbf{y}_{\mathbf{r}} - \mathbf{1}x_{\mathbf{r}}) \tag{22}$$

Setting the gradient of the argument in eqn (22) with respect to $\widehat{x}_{\mathbf{r}}$ equal to zero gives

$$0 = \frac{d}{d\widehat{x}_{\mathbf{r}}} (\mathbf{y}_{\mathbf{r}} - \mathbf{1}\widehat{x}_{\mathbf{r}})^H \Gamma_{\mathbf{r}}^{-1} (\mathbf{y}_{\mathbf{r}} - \mathbf{1}\widehat{x}_{\mathbf{r}}) \tag{23}$$

$$0 = -2\mathbf{1}^H \Gamma_{\mathbf{r}}^{-1} (\mathbf{y}_{\mathbf{r}} - \mathbf{1}\widehat{x}_{\mathbf{r}}) \tag{24}$$

$$\widehat{x}_{\mathbf{r}} = \left(\mathbf{1}^H \Gamma_{\mathbf{r}}^{-1} \mathbf{1}\right)^{-1} \mathbf{1}^H \Gamma_{\mathbf{r}}^{-1} \mathbf{y}_{\mathbf{r}} \tag{25}$$

It can be shown that solution (25) can also be obtained by minimizing the total output power (or variance), while maintaining unity gain in a desired direction (the foresight):

$$\min_{\mathbf{w}} \ \mathbf{w}_{\mathbf{r}}^{\mathrm{H}} \Gamma_{\mathbf{r}} \mathbf{w}_{\mathbf{r}},$$

$$\text{s.t. } \mathbf{w}_{\mathbf{r}}^{\mathrm{H}} \mathbf{1} = 1 \tag{26}$$

Solving for eqn (26) yields the closed form solution

$$\widehat{\mathbf{w}}_{\mathrm{MV}} = \frac{\mathbf{1}\Gamma_n^{-1}}{\mathbf{1}^H \Gamma_n^{-1} \mathbf{1}} \tag{27}$$

which is known as minimum variance (MV) or capon beamforming.

In practice, the noise covariance is unknown, and is instead empirically estimated from data ($\Gamma_n = E[\mathbf{y}\mathbf{y}^H]$). For stability of covariance matrix inversion, this estimation often relies on averaging multiple sub-apertures and focus points or adding a constant factor to the diagonal of the covariance matrix (diagonal loading). Note here that for $\Gamma = \sigma_n^2\mathbf{I}$ (white Gaussian noise), we get the DAS solution as in eqn (18).

Minimum variance beamforming was found to improve both resolution and contrast in ultrasound images, and has similarly found application in plane

wave compounding (Austeng et al., 2011). However, it is computationally complex because of the inversion of the covariance matrix (Raz, 2002), leading to significantly longer reconstruction times compared with DAS. To boost image quality further, eigenspace-based MV beamforming has been proposed (Deylami et al., 2016), at the expense of further increasing computational complexity. As a result of this, real-time implementations remain challenging, to an extent that MV beamforming is almost exclusively used as a research tool.

Wiener beamforming

In the previously covered methods, we considered the ML estimate of $\hat{x}$. Following eqn (4), we can extend this by including a prior probability distribution $p(x_\mathbf{r})$, such that

$$\hat{x}_\mathbf{r} = \arg\max_{x_\mathbf{r}} (\mathbf{y_r}|x_\mathbf{r})p(x_\mathbf{r}) \tag{28}$$

For a Gaussian likelihood model, the solution to this MAP estimate is equivalent to minimizing the mean squared error, such that

$$\hat{\mathbf{w}} = \arg\min_\mathbf{w} E[|x_\mathbf{r} - \mathbf{w}^H \mathbf{y_r}|^2] \tag{29}$$

also known as Wiener beamforming (Van Trees, 2004). Solving this yields

$$\hat{\mathbf{w}}_{\text{wiener}} = \frac{\sigma_x^2}{\sigma_x^2 + \mathbf{w}_{\text{MV}}^H \mathbf{\Gamma_r} \mathbf{w}_{\text{MV}}} \mathbf{w}_{\text{MV}} \tag{30}$$

with $\mathbf{\Gamma_r}$ being the array covariance for beamforming point $\mathbf{r}$, and $\mathbf{w}_{\text{MV}}$ the MV beamforming weights given by eqn (27). Wiener beamforming is therefore equivalent to MV beamforming, followed by a scaling factor based on the ratio between the signal power and total power of the output signal, which can be referred to as post-filtering. Based on this result, Nilsen and Holm (2010) observed that for any $\mathbf{w}$ that satisfies $\mathbf{w}^H 1 = 1$ (unity gain), we can find a Wiener post-filter that minimizes the MSE of the estimated signal. As such, we can write

$$H_{\text{wiener}} = \arg\min_H E[|x_\mathbf{r} - H\mathbf{w}^H \mathbf{y_r}|^2] \tag{31}$$

$$H_{\text{wiener}} = \frac{\sigma_x^2}{\sigma_x^2 + \mathbf{w}^H \mathbf{\Gamma_r} \mathbf{w}} \tag{32}$$

Assuming white Gaussian noise ($\mathbf{\Gamma_r} = \sigma_n^2 \mathbf{I}$, and $x \sim \mathcal{N}(0, \sigma_x^2)$ the Wiener beamformer is equivalent to Wiener post-filtering for DAS, given by

$$\hat{w}_{\text{wiener}} = H_{\text{wiener}} \mathbf{w}_{\text{DAS}} = \frac{\sigma_x^2}{C\sigma_x^2 + \sigma_n^2} 1 \tag{33}$$

Coherence factor weighing

The coherence factor (CF) (Mallart and Fink, 1994) aims to quantify the coherence of the backscattered echoes to improve image quality through scaling with a so-called coherence factor, defined as

$$\text{CF} = \frac{|1^H \mathbf{y_r}|^2}{C \mathbf{y_r}^H \mathbf{y_r}} \tag{34}$$

where C denotes the number of channels. Effectively, this operates as a post-filter, after beamforming, based on the ratio of coherent and incoherent energy across the array. As such, it can suppress focusing errors that may occur because of speed-of-sound inhomogeneity, given by

$$\hat{x}_{\text{CF}} = \text{CF} \cdot \hat{x}_{\text{DAS}} = \frac{\text{CF}}{C} \mathbf{1}^H \mathbf{y_r} = \frac{\sigma_x^2}{C\sigma_x^2 + C\sigma_n^2} 1^H \mathbf{y} \tag{35}$$

The CF has been reported to significantly improve contrast, especially in regions affected by phase distortions. However it also suffers from reduced brightness and speckle degradation. An explanation for this can be found when comparing eqn (35) with the Wiener post-filter for DAS in eqn (33). We can see that CF weighing is in fact a Wiener post-filter where the noise is scaled by a factor C, leading to a stronger suppression of interference, but consequently also reducing brightness. Several derivations of the CF have been proposed to overcome some of these limitations or to further improve image quality, such as the generalized CF (Li and Li, 2003) and phase coherence factor (Camacho et al., 2009).

Iterative MAP beamforming

Chernyakova et al. Chernyakova et al., 2019) proposed an iterative maximum *a posteriori* (iMAP) estimator, which provides a statistical interpretation to post-filtering. The iMAP estimator works under the assumption of knowledge on the received signal model, and treats signal of interest and interference as uncorrelated Gaussian random variables with variance σ_x^2. Given the likelihood model in eqn ((15) and $x \sim \mathcal{N}(0, \sigma_x^2)$, the MAP estimator of x is given by

$$\hat{x}_{\text{MAP}} = \frac{\sigma_x^2}{\sigma_n^2 + C\sigma_x^2} \mathbf{1}^H \mathbf{y_r} = \frac{C\sigma_x^2}{\sigma_n^2 + C\sigma_x^2} \hat{x}_{\text{DAS}} \tag{36}$$

However, the parameters σ_x^2 and σ_n^2 are unknown in practice. Instead, these can be estimated from the data at hand, leading to an iterative solution

First an initial estimate of the signal and noise variances is calculated through

$$\left\{\hat{\sigma}_x^2, \hat{\sigma}_n^2\right\}_{(t)} = \left\{\hat{x}_{(t)}^2, \frac{1}{C}||\mathbf{y_r} - \mathbf{1}\hat{x}_{(t)}||^2\right\} \tag{37}$$

and initialization with the DAS estimate $\hat{x}_{(0)} = \frac{1}{C}\mathbf{1}^{\text{H}}\mathbf{y}$. Following eqns (4) and (37), a MAP estimate of the

beamformed signal is given by

$$\widehat{x}_{\mathrm{iMAP},(t+1)} = \frac{\widehat{\sigma}^2_{x,(t)}}{C\widehat{\sigma}^2_{x,(t)} + \widehat{\sigma}^2_{n,(t)}} \mathbf{1}^{\mathrm{H}}\mathbf{y_r} \tag{38}$$

where t is an index denoting the number of iterations. Equations (37) and (38) are iterated until a stopping criterion is met. The authors suggest that two iterations yield a sufficient noise reduction (80−100 dB) for conventional ultrasound images. If we compare eqn (33) with eqn (38), we can see that for the estimates given by eqn (37), Wiener beamforming coincides with a single iteration of the iMAP algorithm.

ADMIRE

Byram and colleagues have proposed aperture domain model image reconstruction (ADMIRE) (Byram, 2017, Byram et al., 2015), which explicitly models the underlying acoustic "sources" in the scan region. Let $\mathbf{A}$ be a model matrix with predictors representing both scattering from cluttered regions and scattering from the direction of interest, and $\boldsymbol{\beta}$ a vector of model parameters. We can then write the received signal at each time instance as $\mathbf{y} = \mathbf{A}\beta$, where $\mathbf{y}$ is a single frequency bin of the received signal from a given short-time Fourier transform (STFT) window. Exact details on $\mathbf{y}$, $\boldsymbol{\beta}$ and $\mathbf{A}$ are given in Byram et al. (Byram et al., 2015). Performing MAP estimation for $\boldsymbol{\beta}$ under a (white) Gaussian measurement model and a mixed Gaussian/Laplacian prior yields

$$\widehat{\beta} = \arg\min_{\boldsymbol{\beta}}\left(\left|\left\|\mathbf{y} - \mathbf{A}\beta\right\|_2^2 + \lambda\left(\alpha\left\|\boldsymbol{\beta}\right\|_1 + (1-\alpha)\left\|\mathbf{y}\right\|_2^2/2\right)\right)\right. \tag{39}$$

where α and λ are regularization parameters. This particular form of regularization is also called elastic-net regularization. ADMIRE shows significant reduction in clutter as a result of multipath scattering and reverberation, resulting in a 10- to 20-dB improvement in contrast-to-noise ratio (CNR).

Sparse coding

Chernyakova and Eldar (Chernyakova and Eldar, 2014) proposed formulating the beamforming process as a line-by-line recovery of backscatter intensities from (potentially undersampled) Fourier coefficients. Denoting the axial fast-time intensities by $\mathbf{x} \in R^N$, and the noisy measured DFT coefficients of a scan line by $\tilde{\mathbf{y}} \in R^M$, with $M \leq N$, we can formulate the linear measurement model

$$\tilde{\mathbf{y}} = \mathbf{H}\mathbf{F}_u\mathbf{x} + \mathbf{n} = \mathbf{A}\mathbf{x} + \mathbf{n} \tag{40}$$

where $\mathbf{F}_u$ is an $M \times N$ (partial) DFT matrix, $\mathbf{H}$ is a diagonal $M \times M$ matrix representing the DFT coefficients of the transmit pulse and $\mathbf{n}$ is a white Gaussian noise vector. Recovery of $\mathbf{x}$ under a Laplace prior distribution (*i.e.,* assuming it is sparse) can again be posed as a MAP estimation problem:

$$\begin{aligned}\widehat{\mathbf{x}} &= \arg\max_{\mathbf{x}} p(\tilde{\mathbf{y}}|\mathbf{x})p(\mathbf{x}) \\ &= \arg\max_{\mathbf{x}} ||\tilde{\mathbf{y}} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda||\mathbf{x}||_1\end{aligned} \tag{41}$$

where λ is a regularization parameter. Problem (41) can be solved using the iterative shrinkage and thresholding algorithm (ISTA), a proximal gradient method:

$$\widehat{\mathbf{x}}^{k+1} = \tau_\lambda\left(\mathbf{x}^k - \mu\mathbf{A}^H\left(\mathbf{A}\mathbf{x}^k - \tilde{\mathbf{y}}\right)\right) \tag{42}$$

where $\tau_\lambda = \mathrm{sgn}(x_i)(|x_i| - \lambda)_+$ is the proximal operator of the ℓ_1 norm, μ is the gradient step size and $(\cdot)^H$ denotes the Hermitian, or conjugate transpose. It is interesting to note that the first step in the ISTA algorithm, given by $\widehat{\mathbf{x}}^1 = \mathbf{A}^H\tilde{\mathbf{y}} = \mathbf{F}_u^H\mathbf{H}^H\tilde{\mathbf{y}}$, thus maps $\tilde{\mathbf{y}}$ back to the axial/fast-time domain through the zero-filled inverse DFT.

Wavefield inversion

The previously described beamforming methods all build on measurement models that treat pixels or scan lines (or, for ADMIRE, short-time windows) independently. As a result, complex interaction of contributions and interference from the full lateral field of view are not explicitly modeled and are often approximated through some noise model. To that end, several works explore reconstruction methods that model the joint across the full field of view and its intricate behavior, at the cost of a higher computational footprint.

Such methods typically rely on some form of "wavefield inversion," that is, inverting the physical wave propagation model. One option is to pose beamforming as a MAP optimization problem through a likelihood model that relates the per-pixel backscatter intensities to the channel signals Ozkan et al., 2017, Szasz et al., 2016, Szasz et al., 2016) and some prior/regularization term on the statistics of spatial distributions of backscatter intensities in anatomical images. On the basis of the time delays given by eqn ((13) (and the Green's function of the wave equation), our typical linear forward model can again be formulated as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \tag{43}$$

where $\mathbf{x} \in \mathrm{R}^{r_x r_z}$ is a vector of beamformed data, $\mathbf{n} \in \mathrm{R}^{CN_t}$ an additive white Gaussian noise vector and $\mathbf{y} \in \mathrm{R}^{CN_t}$ the received channel data. The space−time mapping is encoded in the sparse matrix $\mathbf{A} \in \mathrm{R}$.

Solving this system of equations relies heavily on priors to yield a unique and anatomically feasible solution, and yields the MAP optimization problem

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \| \mathbf{y} - \mathbf{A}\mathbf{x} \|_2^2 - \log p_\theta(\mathbf{x}), \tag{44}$$

where $\log p_\theta(\mathbf{x})$ acts as a regularizer, with parameters θ (*e.g.,* an l_1 norm to promote a sparse solution (Combettes and Wajs, 2005)). Ozkan et al. (Ozkan et al., 2017) investigated several intuition- and physics-based regularizers and their effect on the beamformed image. The results indicate benefits for contrast and resolution for all proposed regularization methods, however each yielding different visual characteristics. This illustrates that choosing correct regularization terms and parameters that yield a robust beamformer can be challenging.

Post-processing

After the channel data are mapped to the image domain via beamforming, ultrasound systems apply several post-processing steps. Classically, this includes further image processing to boost B-mode image quality (*e.g.,* contrast, resolution, despeckling), but also spatiotemporal processing to suppress tissue clutter and to estimate motion (*e.g.,* blood flow). Beyond this, we see increasing attention for post-processing methods dedicated to advanced applications such as super-resolution ultrasound localization microscopy (ULM). We now review some of the model-based methods for post-processing, covering B-mode image quality improvement, tissue clutter filtering and ULM.

B-Mode image quality improvement

Throughout the years, many B-mode image-quality boost algorithms have been proposed with aims that can be broadly categorized into: (i) resolution enhancement, (ii) contrast enhancement and (iii) speckle suppression. Although our focus lies on model-based methods (to recall: methods that are derived from models and first principles), it is worth nothing that B-mode processing often also relies on heuristics to accommodate, for example, user preferences. These include fine-tuned brightness curves (S-curves) to improve perceived contrast.

In a method commonly used to boost image quality, multiple transmissions are coherently compounded with diverse transmit parameters. Often, a simple measurement model similar to that in DAS is assumed, where multiple transmissions are (after potential TOF alignment) assumed to measure the same tissue intensity for a given pixel, but with different Gaussian noise realizations. As for the Gaussian likelihood model for DAS, this then simply yields averaging of the individual measurements (*e.g.,* different plane wave angles, or frequencies). More advanced model-based compounding methods use MV weighting of the transmits, thus assuming a likelihood model in which multiple measurements have correlated noise:

$$\hat{x}_{\mathbf{r}} = \arg\max_{x_{\mathbf{r}}} \log p(\mathbf{y}_{\mathbf{r}} | x_{\mathbf{r}}, \Gamma_{\mathbf{r}}) \tag{45}$$

$$= \arg\min_{x_{\mathbf{r}}} (\mathbf{y}_{\mathbf{r}} - \mathbf{1}x_{\mathbf{r}})^H \Gamma_{\mathbf{r}}^{-1} (\mathbf{y}_{\mathbf{r}} - \mathbf{1}x_{\mathbf{r}}) \tag{46}$$

Note that here, unlike in MV beamforming, $\mathbf{y}_{\mathbf{r}}$ is a vector containing the beamformed pixel intensities from multiple transmits/measurements (after TOF alignment), $\hat{x}_{\mathbf{r}}$ is the compounded pixel and $\Gamma_{\mathbf{r}}$ is the auto-correlation matrix across the series of transmits to be estimated. Compounding can boost resolution and contrast and suppress speckle.

After compounding, additional processing is performed to further boost image quality. For denoising/speckle suppression, many advanced methods have been developed that all aim to reduce inter-tissue variations while preserving sharpness of edges. This includes directional median filtering (Czerwinski et al., 1995), adaptive kernel sizes (Nugroho et al., 2019) or, more recently, non-local-means (NLM) filters and their Bayesian extension, the optimized Bayesian NLM (OBNLM) filter. These methods formulate denoising as a patch-based MAP inference problem (Kervrann et al., 2007), where an unobserved image patch $\mathbf{x}$ with some unknown probability density function is estimated from its noisy observation $\mathbf{y} = f(\mathbf{x} + \mathbf{n})$, with $f(\cdot)$ denoting a function related to the image formation processes, and $\mathbf{n}$ a random noise vector with independent and identically distributed (i.i.d.) entries. The density functions are then estimated by exploiting redundancy across patches in an image, drawing samples from a local neighborhood. This probabilistic Bayesian interpretation of NLM has enabled ultrasound-specific implementations with more realistic (multiplicative) noise models (Coupe et al., 2008). Other MAP approaches pose denoising as a dictionary matching problem (Jabarulla and Lee, 2018). These methods do not explicitly estimate patch density functions from the image, but instead learn a dictionary of patches. Other approaches combine spatial filters with priors (*e.g.,* sparsity) in a transformed domain, that is, the wavelet domain (Garg and Khandelwal, 2019), or combine PCA and wavelet transforms (Jagadesh and Rani, 2016)].

To achieve a boost in image resolution, the problem can be recast as MAP estimation under a likelihood model that includes a deterministic blurring/point-spread-function matrix $\mathbf{A}_{\text{blur}}$

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \| \mathbf{y} - \mathbf{A}_{\text{blur}}\mathbf{x} \|_2^2 - \log p_\theta(\mathbf{x}) \tag{47}$$

where $\mathbf{x}$ is the (vectorized) high-resolution image to be recovered, and $\mathbf{y}$ is a (vectorized) blurred and noisy (Gaussian white) observation $\mathbf{y}$. This deconvolution problem is ill posed and requires adequate regularization via priors. As we noted before, the log-prior term can take many forms, including ℓ_1 or total variation-based regularizers.

Clutter filtering for flow

For many applications (including flow mapping and localization microscopy), suppression of reflections from tissue is of interest (Wildeboer et al., 2020). Slow-moving tissue introduces a clutter signal that introduces artifacts and obscures the feature of interest being imaged (be it blood velocity or, *e.g.,* contrast agents), and considerable effort has gone into suppressing this tissue clutter signal. Although infinite impulse response (IIR) and finite impulse response (FIR) filters have been the filters most commonly used for such tasks, it is still very difficult to separate the signals originating from slow-moving blood or fast-moving tissue. Therefore, spatiotemporal clutter filtering is receiving increasing attention. We here review some of these more advanced methods (including singular value thresholding and robust principle component analysis), again taking a probabilistic MAP perspective.

We define the spatiotemporal measured signal as a Casorati matrix, $\mathbf{Y} \in \mathrm{R}^{NM \times T}$, where N and M are spatial dimensions, and T is the time dimension, which we model as $\mathbf{Y} = \mathbf{X}_{\text{tissue}} + \mathbf{X}_{\text{blood}}$, where $\mathbf{X}_{\text{tissue}} \in \mathrm{R}^{NM \times T}$ is the tissue component, and $\mathbf{X}_{\text{blood}} \in \mathrm{R}^{NM \times T}$ is the blood/flow component. We then impose a prior on $\mathbf{X}_{\text{tissue}}$, and assume it to be *low rank*. If we additionally assume $\mathbf{X}_{\text{blood}}$ to have i.i.d. Gaussian entries, the MAP estimation problem for the tissue clutter signal becomes

$$\begin{aligned}\widehat{\mathbf{X}}_{\text{tissue}} &= \arg\max_{\mathbf{X}_{\text{tissue}}} p(\mathbf{Y}|\mathbf{X}_{\text{tissue}}) p(\mathbf{X}_{\text{tissue}}) \\ &= \arg\max_{\mathbf{X}_{\text{tissue}}} \log p(\mathbf{Y}|\mathbf{X}_{\text{tissue}}) + \log(p(\mathbf{X}_{\text{tissue}})) \\ &= \arg\min_{\mathbf{X}_{\text{tissue}}} \| \mathbf{Y} - \mathbf{X}_{\text{tissue}} \|_F + \lambda \| \mathbf{X}_{\text{tissue}} \|_*\end{aligned} \tag{48}$$

where $\| \cdot \|_F$ and $\| \cdot \|_*$ denote the Frobenius norm and the nuclear norm, respectively. The solution to eqn (48) is

$$\widehat{\mathbf{X}}_{\text{tissue}} = \mathscr{T}_{\text{SVT},\lambda}(\mathbf{Y}) \tag{49}$$

where $\mathscr{T}_{\text{SVT},\lambda}$ is the singular value thresholding function, which is the proximal operator of the nuclear norm (Cai et al., 2010).

To improve on the model in eqn (48), one can include a more specific prior on the flow components and separate them from the noise:

$$\mathbf{Y} = \mathbf{X}_{\text{tissue}} + \mathbf{X}_{\text{blood}} + \mathbf{N} \tag{50}$$

Here we place a mixed ℓ_1/ℓ_2 prior on the blood flow component $\mathbf{X}_{\text{blood}}$ and assume i.i.d. Gaussian entries in the noise matrix $\mathbf{N}$, such that

$$\begin{aligned}\widehat{\mathbf{X}}_{\text{tissue}}, \widehat{\mathbf{X}}_{\text{blood}} &= \arg\max_{\mathbf{X}_{\text{tissue}}, \mathbf{X}_{\text{blood}}} p(\mathbf{Y}|\mathbf{X}_{\text{tissue}}, \mathbf{X}_{\text{blood}}) \\ &\quad p(\mathbf{X}_{\text{tissue}})\, p(\mathbf{X}_{\text{blood}}) \\ &\quad + \log(p(\mathbf{X}_{\text{tissue}})) + \log(p(\mathbf{X}_{\text{blood}})) \\ &= \arg\min_{\mathbf{X}_{\text{tissue}}, \mathbf{X}_{\text{blood}}} \| \mathbf{Y} - \mathbf{X}_{\text{tissue}} - \mathbf{X}_{\text{blood}} \|_F \\ &\quad + \lambda_1 \| \mathbf{X}_{\text{tissue}} \|_* + \lambda_2 \| \mathbf{X}_{\text{blood}} \|_{1,2}\end{aligned}$$

where $\| \cdot \|_{1,2}$ indicates the ℓ_1 and ℓ_2 norm. This *low-rank plus sparse* optimization problem is also termed robust principle component analysis (RPCA), and can be solved through an iterative proximal gradient method

$$\mathbf{X}_{\text{tissue}}^{k+1} = \mathscr{T}_{\text{SVT},\lambda_1}\left(\mathbf{X}_{\text{tissue}}^{k} - \mu_1\left(\mathbf{Y} - \mathbf{X}_{\text{tissue}}^{k} - \mathbf{X}_{\text{blood}}^{k}\right)\right) \tag{51}$$

$$\mathbf{X}_{\text{blood}}^{k+1} = \mathscr{T}_{1,2,\lambda_2}\left(\mathbf{X}_{\text{blood}}^{k} - \mu_2\left(\mathbf{Y} - \mathbf{X}_{\text{tissue}}^{k} - \mathbf{X}_{\text{blood}}^{k}\right)\right), \tag{52}$$

where $\mathscr{T}_{\text{SVT},\lambda_1}$ is the solution of eqn (48) (*i.e.,* the proximal operator of the nuclear norm), $\mathscr{T}_{1,2,\lambda_1}$ is the mixed $\ell_1 - \ell_2$ thresholding operation and μ_1 and μ_2 are the gradient steps for the two terms.

Shen et al. (Shen et al., 2019) further augmented the RPCA formulation to boost resolution for the blood flow estimates. To that end, they add a PSF-based convolution kernel to the blood component $\mathbf{A}_r \circledast \mathbf{X}_{\text{blood}}$, casting it as a joint deblurring and signal separation problem.

Ultrasound localization microscopy

We now turn to an advanced and increasingly popular ultrasound signal processing application: ULM. Conventional ultrasound resolution is fundamentally limited by wave physics, to half the wavelength of the transmitted wave, that is, the diffraction limit. This limit is in the range of millimeters for most ultrasound probes, and is inversely proportional to the transmission frequency. However, high transmit frequencies come at the cost of lower penetration depth.

To overcome this diffraction limit, ULM adapts concepts from Nobel prize-winning superresolution fluorescence microscopy to ultrasound. Instead of localizing fluorescent blinking molecules, ULM detects and localizes ultrasound contrast agents, microbubbles (MBs), flowing through the vascular bed. These microbubbles have are similar in size to red blood cells and act as point scatterers. By accumulating precisely localized microbubbles across many frames, a super-resolution image of

the vascular bed can be obtained. In typical implementations, the localization of the MBs is performed by centroid detection (Christensen-Jeffries et al., 2020, Couture et al., 2011, Siepmann et al., 2011).

Not surprisingly, we can also pose microbubble localization as a MAP estimation problem (van Sloun et al., 2017). We define a sparse high-resolution image that is vectorized into $\mathbf{x}$, in which only few pixels have non-zero entries: those pixels that contain a microbubble. Our vectorized measurements can then be modeled as $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, where $\mathbf{A}$ is a PSF matrix and $\mathbf{n}$ is a white Gaussian noise vector. This yields the following MAP problem:

$$\begin{aligned}\hat{\mathbf{x}} &= \arg\max_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \\ &= \arg\max_{\mathbf{x}} \log p(\mathbf{y}|\mathbf{x}) + \log p(\mathbf{x}) \\ &= \arg\min_{\mathbf{x}} \| \mathbf{y} - \mathbf{A}\mathbf{x} \|_2^2 + \lambda \| \mathbf{x} \|_1 \end{aligned} \quad (53)$$

van Sloun et al. van Sloun et al., 2017) proposed solving this sparse coding problem using ISTA, similar to the formulation in eqn ((42).

Instead of processing each image frame independently, Bar-Zion et al. (Bar-Zion et al., 2018) exploited sparse structure in the temporal correlation domain; that is, $\mathbf{y}$ is a correlation image, leading to the SUSHI method. Later, Solomon et al. (Solomon et al., 2019) combined MAP estimation across the spatial dimensions with MAP estimation in time, by complementing the spatial sparse coding problem with a Kalman filter that places a prior on future microbubble locations according to a motion model.

DL FOR US SIGNAL PROCESSING

Deep learning–based ultrasound signal processing offers a highly flexible framework for learning a desired input–output mapping $\hat{X} = f_\theta(Y)$ from training data, overcoming the need for explicit modeling and derivation of solutions. This can be especially advantageous for complex problems in which models fall short (*e.g.,* incomplete, with naive assumptions) or their solutions are demanding or even intractable. We now review some emerging applications of DL in the ultrasound signal processing pipeline. As in the previous section, we first cover advanced methods for beamforming and then turn to downstream post-processing such as B-mode image quality improvement, clutter suppression and ULM.

Beamforming

We discern two categories of approaches: neural networks that replace the entire mapping from channel data to images, and those that replace only the beamsumming operation, that is, after TOF correction. Hyun et al. (Hyun et al., 2021) and Bell et al. (Bell et al., 2019, Bell et al., 2020) recently organized the Challenge on Ultrasound Beamforming with Deep Learning (CUBDL) to incentivize new research in this area. For a more in-depth survey on DL for ultrasound beamforming, including common training strategies and loss functions, we refer the reader to van Sloun et al. (van Sloun et al., 2021).

Direct channel to image transformation

Nair et al. (Nair et al., 2018, Nair et al., 2020) introduced a method that learns a direct convolutional neural network-based transformation between the channel signals and a target B-mode image. The proposed U-Net architecture thus has to learn both the (geometry-based) time-of-flight correction and the subsequent beamsumming. The inputs and outputs of the network comprise IQ-demodulated channel data (separate I and Q inputs) $\mathbf{Y} \in \mathrm{R}^{C \times N_t \times 2}$ and a beamformed image $\mathbf{X} \in \mathrm{R}^{R_x \times R_y}$, respectively. The network additionally outputs direct segmentations of anechoic regions in the image. Training is done using ultrasound simulations of a variety of anechoic cysts.

Replacing the entire beamforming pipeline is an unconventional application of U-Net–style architectures, which were originally designed for image-to-image translation (segmentation) problems. Instead, the U-Net here also performs time-to-space migration. It is worth noting that much of the efficiency of convolutional networks comes from their ability to exploit spatial symmetries (*i.e.,* translation/shift equivariance): their operations do not depend on the position in the data; that is, they are spatially invariant. In contrast, TOF correction is based on the geometry of the array and the scan region, and its operation varies depending on the focus point. As such, learning time–space migration through standard convolutional architectures is not trivial. Most beamforming approaches thus benefit from traditional alignment of the channel data before processing.

Beam summing after TOF correction

Although the solution proposed by Nair et al. (Nair et al., 2018, Nair et al., 2020) replaces the entire beamformer with a convolutional neural network (including the time–space migration), most works confine the problem to only the beamsumming and compounding steps, leaving the TOF correction based on geometry. As such it replaces the model-based probabilistic beamforming/summing methods, mapping TOF-corrected channel data $\mathbf{Y} \in \mathrm{R}^{C \times R_x \times R_y}$, to an image $\mathbf{X} \in \mathrm{R}^{R_x \times R_y}$. In advanced adaptive beamforming with intricate models, this step is often the most time consuming and complex, in some cases preventing real-time implementation. Often, an important goal of DL-based beamsumming is

therefore to accelerate these advanced beamsumming methods by learning fast neural network approximations. Many of the methods we list here make use of deep convolutional architectures, which have a significant spatial receptive field: the summed output at a given pixel is based on a large spatial input support. This spatial "awareness" contrasts with most of the model-based beamforming methods that operate on a per-pixel or sometimes per-line basis.

In the work by Khan et al. (Khan et al., 2019, Khan et al., 2020), a DL-based beamforming network was proposed that replaces the conventional summation of channel signals and envelope detection with a deep convolutional neural network (CNN). Additionally, Khan et al. reported that their method enables image reconstruction from sub-sampled radiofrequency (RF) data, in which a set of receive elements is inactive, thereby reducing the required bandwidth. In an extension of this work, Khan et al. (Khan et al., 2021, Khan et al., 2021) found that such a beamformer can be optimized for different imaging settings and controlled through the introduction of a style code vector in the latent space and training on a corresponding image target with a given style/setting. Such an approach avoids the need to store separate models for each setting.

Vignon et al. (Vignon et al., 2020) proposed a solution similar to that of Khan et al. (Khan et al., 2021, Khan et al., 2021) in which linewise channel signals are beamsummed by a CNN. It is worth noting that in this work, training data are generated using purely simulations, in which targets correspond to DAS beamformed images obtained with larger simulated array aperture to yield better imaging resolution.

Mamistvalov et al. (Mamistvalov et al., 2021) proposed a U-Net architecture for beamsumming after Fourier-domain TOF correction. This enables sub-Nyquist acquisition by Xampling (Chernyakova and Eldar, 2014), reducing data rates. The authors report that their convolutional U-Net architecture enables reconstruction of high-quality images from fast-time sub-Nyquist acquisitions acquired with a sparse array (channel subsampling) (Cohen and Eldar, 2018), suppressing the aliasing artifacts.

Similarly, Huijben et al. (Huijben et al., 2020) performed image reconstruction from sparse arrays of undersampled channels using a convolutional network architecture. In addition, they provided a mechanism for jointly learning optimal channel selection/sparse array design via a technique dubbed *deep probabilistic subsampling*.

Although most beamforming methods aim at boosting resolution and contrast, Hyun et al. (Hyun et al., 2019) argued that beamformers should accurately estimate the true tissue backscatter map and, thus, also target speckle reduction. The authors trained their beamformer on ultrasound simulations of a large variety of artificial tissue backscatter maps derived from natural images.

Post-processing

Application of DL methods to general image processing/restoration problems has seen a surge of interest in recent years, resulting in remarkable performance across a range of applications. Naturally, these pure image processing methods are being explored for ultrasound post-processing as well. In this section we treat the same topics as in the previous section, but focus on recent DL methods.

B-Mode image quality improvement

A common means to boost ultrasound image quality is compounding data from different transmits. Although model-based methods offer simple pixel-based compounding strategies (either by simply summing or via MV processing), several groups have investigated the use of deep CNNs for improved compounding (Guo et al., 2020, Khan et al., 2019, Lu et al., 2019). Jansen et al. (Jansen et al., 2021) propose performing this compounding step in the Radon domain.

In addition to compounding multiple transmits, many DL methods aim at single-image enhancement, including resolution/contrast boost, but also speckle suppression. Gasse et al. (Gasse et al., 2017) explored mapping a single PW image to an image that was compounded using multiple PWs. Zhou et al. (Zhou et al., 2018) pursued a similar goal, but introduced a multibranch CNN with an additional wavelet-based post-processing step. Qi et al. (Qi et al., 2020) performed processing in the Fourier domain, using pixelwise fully connected neural networks. Rather than compounded PWs, the authors use a focused line-scan image as a target.

Chang et al. (Chang et al., 2019) posed denoising as a signal decomposition problem. To that end they proposed a two-stage CNN that simultaneously models the image and noise, where the noise estimates, in turn, inform the image estimates to cope with various noise distributions.

Temiz and Bilge (Temiz and Bilge, 2020) aimed at single-image super-resolution, that is, to achieve a B-mode image with a higher pixel resolution. The authors achieve this by training a deep CNN with a dataset containing B-mode US images across a range of scale/zoom factors. A similar approach was taken by Choi et al. (Choi et al., 2018), who proposed a deep CNN called SRGAN with the aim of mapping low-resolution images to a high-resolution domain.

Both Vedula et al. (Vedula et al., 2017) and Ando et al. (Ando et al., 2020) approached the issue of speckle reduction in similar ways, by using a CNN. However,

they used different forms of input and target data, while Vedula et al. (Vedula et al., 2017) used IQ data, Ando et al. (Ando et al., 2020) used B-mode images. Similarly, Dietrichson et al. (Dietrichson et al., 2018) performed this task by using a CNN, albeit with a more complex training strategy, by employing a GAN-based structure. Karaoglu et al. (2021) compared many approaches to this problem from the perspective of neural networks, and detailed the effectiveness of the many architectures. They found that the GAN and the U-Net−like algorithms in their study performed the best. Although all the works cited on this topic so far deal with 2-D US scans, Li et al. (Li et al., 2020) attempted to extend it to 3-D imaging using a 3-D U-Net model. It is an interesting point to note that Vedula et al. (Vedula et al., 2017) and Ando et al. (Ando et al., 2020) used simulated data, whereas the other works on speckle reduction used *in vivo* data gathered from volunteers. However, there is uniformity in how these works created their target images through model based speckle reduction algorithms.

Most DL methods for image quality improvement rely on supervised learning, requiring ground truth targets, which are often difficult to obtain. As an alternative, Huh et al. (Huh et al., 2022) presented a self-supervised method based on the cycle-GAN architecture, originally developed for unpaired (cycle-consistent) style transfer (Huh et al. 2021). This approach aims at transferring the features of a high-quality target distribution of images to a given low-quality image, which the authors leverage to improve elevational image quality in 3-D ultrasound.

Clutter filtering for flow

Brown et al. (Brown et al., 2020) described a 3-D (2-D + time) CNN-based spatiotemporal filtering scheme for tissue clutter suppression, aimed mostly at accelerating the Singular Vector Decomposition (SVD) algorithm for real-time use. To that end, they used SVD-processed *in vivo* and *in vitro* images as targets. Similarly, Wang et al. (Wang et al., 2021) aimed at replacing SVD thresholding, but rather than a 3-D CNN, the authors adopted a 2-D (spatial) CNN, aggregating temporal information in the feature space through a recurrent neural network.

Tabassian et al. (Tabassian et al., 2019) used a deep 3-D CNN (2-D + time) to suppress clutter and reverberation artifacts that plague echocardiographic imaging. Their deep network was trained on realistic simulations of echocardiographic exams, with simulated superimposed artifacts.

Ultrasound localization microscopy

van Sloun et al. (van Sloun et al., 2019, van Sloun et al., 2020) proposed a DL method based on an encoder−decoder architecture that aims to replace costly iterative techniques for sparse coding to obtain super-resolution vascular images from high-density contrast-enhanced ultrasound data. Later, Liu et al. (Liu et al., 2020) proposed a similar approach, but used a sub-pixel CNN. Brown et al. (Brown et al., 2021) proposed to perform jointly tissue clutter filtering and localization by a 3-D CNN to further boost processing rates. Notably, Youn et al. (Youn et al., 2020) performed localization directly from channel data.

MODEL-BASED DL FOR US SIGNAL PROCESSING

We now highlight several works that incorporate signal processing knowledge in their DL approaches to improve performance, reduce network complexity and provide reliable inference models. Generally, these models retain a large part of the conventional signal processing pipeline intact, and replace critical points in the processing with neural networks so as to provide robust inference as a result. We discuss methods ranging from iterative solvers to unfolded fixed complexity solutions.

Beamforming

Model-based pre-focusing using DL. Pre-focusing (or TOF correction) is conventionally done deterministically based on the array geometry and assuming a constant speed-of-sound. Instead, data-adaptive focusing, by calculating delays based on the recorded data, facilitates correction for speed-of-sound mismatches. The work by Nair et al. (Nair et al., 2018, Nair et al., 2020) does this implicitly, by finding a direct mapping from the time domain to an output image, using DL. However, this yields a black-box solution, which can be difficult to interpret.

Kim et al. (Kim et al., 2021) adheres more strictly to a conventional beamforming structure and tackled this problem in two steps: first, the estimation of a local speed-of-sound map, and second, the calculation of the corresponding beamforming delays. The speed-of-sound image is predicted from multi-angled plane wave transmissions using SQI-net (Oh et al., 2021), a type of U-Net. One then needs to find the propagation path and travel time of the transmitted pulse, that is, the delay matrix, between each imaging point and transducer element. For a uniform speed-of-sound this is trivial, as the shortest distance between a point and element corresponds to the fastest path. For a non-uniform speed-of-sound, this is more challenging and requires a path-finding algorithm that adds to the computational complexity. The Dijkstra algorithm (Dijkstra, 1959), for instance, which is commonly used to find the fastest path, has a complexity of $\mathcal{O}(n^2 \log n)$, where n is the number of

nodes in the graph or, equivalently, the density of the local speed-of-sound grid.

As such, the authors proposed a second U-Net−style neural network, referred to as DelayNet, for estimating these delay times. The network comprises 3 × 3 locally masked convolutions, such that no filter weights are assigned in the direction opposite the direction of wave propagation. Intuitively, this can be understood as enforcing an increasing delay time the further we get from the transducer; that is, the wave does not move in reverse direction. Furthermore, the reduced filter count improves computational efficiency by ∼33%.

Finally, the predicted delay matrix is used to focus the RF data using the corrected delays, after which it is beamsummed to yield a beamformed output signal. As such, DelayNet does not have to be trained directly on a target delay matrix, but instead can be trained end-to-end on the desired beamformed targets. Note that in this method, the speed-of-sound is estimated in a purely data-driven fashion. However, the pre-focusing itself inherits a model-based structure by constraining the problem to learning time shifts from the aforementioned speed-of-sound map.

Model-based beamsumming using DL

Luijten et al. [(Luijten et al., 2019), 2020] proposed adaptive beamforming by DL (ABLE), a DL-based beamsumming approach that inherits its structure from adaptive beamforming algorithms, specifically minimum variance (MV) beamforming. ABLE specifically aims to overcome the most computationally complex part of the beamforming, the calculation of the adaptive apodization weights, replacing this with a neural network f_θ. The step from the model-based MAP estimator to ABLE is then given by

$$\widehat{x}_{\mathbf{r}} = \arg\,\max_{x_{\mathbf{r}}} p(\mathbf{y}_{\mathbf{r}}|x_{\mathbf{r}}) = \left(\mathbf{1}^H \Gamma_{\mathbf{r}}^{-1} \mathbf{1}\right)^{-1} \mathbf{1}^H \Gamma_{\mathbf{r}}^{-1} \mathbf{y}_{\mathbf{r}} \tag{54}$$

$$= \arg\,\max_{x_{\mathbf{r}}} p(\mathbf{y}_{\mathbf{r}}|x_{\mathbf{r}}) \approx f_\theta(\mathbf{y}_{\mathbf{r}})^H \mathbf{y}_{\mathbf{r}} \tag{55}$$

where θ comprises the neural network weights, and $\mathbf{y}_{\mathbf{r}}$ the TOF-corrected RF data. Multiplying the predicted weights by the TOF-corrected data and summing the result yield a beamformed output signal.

Note that for training, we do not need access to the apodization weights as in MV beamforming. Instead, this is done end-to-end toward an MV-generated target, given by

$$\underset{\theta}{\text{argmin}}\ \mathscr{L}\left(f_\theta(\mathbf{y}_{\mathbf{r}})^H \mathbf{y}_{\mathbf{r}},\, \widehat{\mathrm{x}}_{\mathrm{MV}}\right) \tag{56}$$

where $\widehat{\mathbf{x}}_{\mathrm{MV}}$ is an MV training target, and $\mathscr{L}$ is a loss function. As the network operates directly on RF data, which have positive and negative signal components, as well as a high dynamic range, the authors propose an anti-rectifier as an activation function. The anti-rectifier introduces a non-linearity while preserving the sign information and dynamic range, unlike the rectified linear unit, or hyperbolic tangent. Similarly, a signed mean squared logarithmic error (SMSLE) loss function is introduced that ensures that errors in the RF domain reflect the errors in the log-compressed output image. The authors illustrate that a relatively small network, comprising four fully connected layers, can solve this task, and is able to generalize well to different data sets. They report an increase in resolution and contrast, while reducing computational complexity by two to three orders of magnitude.

Wiacek et al. (Wiacek et al., 2020) similarly exploited DNNs as a function approximator to accelerate the calculation of the short-lag spatial coherence (SLSC). Specifically the authors applied their method to SLSC beamforming, which displays the spatial coherence of backscattered echoes across the transducer array. This contrasts with conventional DAS beamforming in which the recorded pressures are visualized. The authors report a 3.4 times faster computation compared with the standard CPU-based approach, corresponding to a frame rate of 11 frames/s.

Luchies and Byram (Luchies and Byram, 2018) proposed a wideband DNN for suppressing off-axis scattering, which operates in the frequency domain, similar to ADMIRE discussed earlier. After focusing on an axially gated section of channel data, the RF signals undergo a discrete Fourier transform (DFT), mapping the signal into different frequency bins. The neural network operates specifically on these frequency bins, after which the data are transformed back to the time domain using the inverse discrete Fourier transform (IDFT) and summed to yield a beamformed signal. The same fully connected network structure was used for different center frequencies, only retraining the weights.

An extension of this work is described in Khan et al. Khan et al., 2021, Khan et al., 2021), in which the neural network itself is replaced by a model-based network architecture. The estimation of model parameter β, as formulated in eqn ((39), can be seen as a sparse coding problem $\mathbf{y} = \mathbf{A}\beta$ (where β is a sparse vector) which can be solved by using an iterative algorithm such as ISTA. This yields

$$\widehat{\beta}^{k+1} = \tau_\lambda\left(\boldsymbol{\beta}^k - \mu \mathbf{A}^T\left(\mathbf{A}\boldsymbol{\beta}^k - \mathbf{y}\right)\right) \tag{57}$$

where $\tau_\lambda(\cdot)$ is the soft-thresholding function parameterized by λ.

To derive a model-based network architecture, eqn (57) is unfolded as a feed-forward neural network with

input $\mathbf{A}^T\mathbf{y}$ and output $\widehat{\beta}$, the predicted model coefficients. For each iteration, or fold, we can then learn the weight matrice, and the soft-thresholding parameter λ trainable. This then leads to a learned ISTA algorithm (LISTA)

$$\widehat{\beta}^{k+1} = \tau_{\lambda^k}\left(\mathbf{W}^k\boldsymbol{\beta}^k + \mathbf{A}^T\mathbf{y}\right) \tag{58}$$

where $\mathbf{W}^k$ represents a trainable fully connected layer and λ^k is a (per fold) trainable thresholding parameter. When contrasted with its model-based iterative counterpart ISTA, LISTA is a fixed complexity solution that tailors its processing to a given data set using DL. Compared with conventional deep neural networks, however, LISTA has a lower number of trainable parameters.

The authors illustrate that LISTA can be trained on model fits of ADMIRE or even simulation data containing targets without off-axis scattering, thereby potentially outperforming the fully model-based algorithm ADMIRE, because of its ability to learn optimal regularization parameters from data.

Mamistvalov and Eldar (Mamistvalov and Eldar, 2021) take a similar avenue and recast their model-based solution for reconstructing images from sub-Nyquist acquisitions (across channels and time) into an unfolded LISTA architecture. Although the aforementioned method based on ADMIRE learns to estimate sparse codes in the aperture/channel-domain dictionary, the method of Mamistvalov and Eldar (Mamistvalov and Eldar, 2021) learns to sparsely encode (fast-time) RF lines. The original ISTA-based algorithm is recast as a fixed-length feed-forward model in which the matrix operations are replaced with learned convolutional layers. By training the network on pairs of sub-Nyquist− and full Nyquist−rate data, the authors illustrate that their approach enables a reduction in data rates as high as 88% without significantly compromising image quality.

Model-based wavefield inversion using DL

Reconstruction techniques based on the inversion of a (non-)linear measurement model are often very computationally intensive and require careful tuning of hyperparameters to ensure robust inference. Alternatively, Almansouri et al. (Almansouri et al., 2018, Almansouri et al., 2018) proposed a two-step approach that leverages a simple linear model to obtain an initial estimation, after which further refinement is done through a CNN. As such, the neural network can account for non-linear, and space-varying, artifacts in the measurement model.

The ultrasound forward model is based on a set of differential equations and depends mainly on three parameters: the acoustic velocity c_0, the density ρ_0 and the attenuation α_0. Such a model could abstractly be defined as

$$y = f(c_0, \rho_0, \alpha_0) \tag{59}$$

However, because of the complex non-linear nature of this forward model, a simplified linear model was developed by Almansouri et al. (Almansouri et al., 2018, Almansouri et al., 2018) that yields the estimator

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \| \mathbf{y} - \mathbf{A}\mathbf{x} \|_2^2 - \log p_\theta(\mathbf{x}) \tag{60}$$

where A is a matrix that accounts for time shifting and attenuation of the transmit pulse. The adjoint operator operator of the linearized model gives an approximate estimator for x, given by $\sim\mathbf{x} = \mathbf{A}^T\mathbf{y}$. The authors adopt a U-Net architecture to compensate for artifacts caused by non-linearities. Effectively the the network finds a mapping from a relatively simple estimate, yet based on the physical measurement model, and maps it to a desired high-quality image such that

$$\widehat{\mathbf{x}} \approx f_\theta\left(\mathbf{A}^T\mathbf{y}\right) \tag{61}$$

where $f(\cdot)_\theta$ denotes the neural network, and $\hat{x}$ is the high-quality estimate.

Post-processing and interpretation

Deep unfolding for B-mode IQ enhancement/PW compounding/compressed acquisition. Chennakeshava et al. (Chennakeshava et al., 2020, Chennakeshava et al., 2021) proposed a plane wave compounding and deconvolution method based on deep unfolding. Their architecture is based on a proximal gradient descent algorithm derived from a model-based MAP optimization problem, which is subsequently unfolded and trained to compound three plane wave images, gathered at low frequency, into an image gathered using 75 compounded plane wave transmissions at a higher frequency. This encourages a learned proximal operator that maps low-resolution, low-contrast input images onto a manifold of images with better spatial resolution and contrast.

Denote $\mathbf{x} \in \mathbb{R}^N$ as the vectorized high-resolution beamformed RF image and $\mathbf{y} \in \mathbb{R}^{NM}$ as the vectorized measurement of low-resolution beamformed RF images from $M = 3$ transmitted plane waves. The authors assume the acquisition model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \tag{62}$$

where

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_M \end{pmatrix} \tag{63}$$

and

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{pmatrix} \tag{64}$$

where $\mathbf{y}_m$ is the vectorized, beamformed RF image belonging to the mth steered plane wave transmission, $\mathbf{n} \in \mathbb{R}^{NM}$ is a noise vector that is assumed to follow a Gaussian distribution with zero mean and diagonal covariance and $\mathbf{A} \in \mathbb{R}^{NM \times N}$ is a block matrix, with its blocks $\mathbf{A}_1$, $\mathbf{A}_2$,..., $\mathbf{A}_M$ being the measurement matrices of individual PW acquisitions. The authors assume that the measurement matrices (which capture the system PSF for each PW) follow a convolutional Toeplitz structure.

On the basis of this model, each fold in the unfolded proximal gradient algorithm aimed at recovering the high-resolution image $\mathbf{x}$ is written as

$$\widehat{\mathbf{x}}^{(k+1)} = \mathscr{P}_\theta^{(k)}\left(\mathbf{W}^{(k)}\mathbf{y} + \mathbf{V}^{(k)}\widehat{\mathbf{x}}^{(k)}\right) \tag{65}$$

where $\mathscr{P}_\theta$ is a U-Net−style neural network replacing the generalised proximal operator, and

$$\mathbf{W}^{(k)}\mathbf{y} := \mathbf{w}_1^{(k)} \circledast \mathbf{y}_1 + \mathbf{w}_2^{(k)} \circledast \mathbf{y}_2 + \ldots + \mathbf{w}_M^{(k)} \circledast \mathbf{y}_m \tag{66}$$

$$\mathbf{V}^{(k)}\widehat{\mathbf{x}}^{(k)} = \mathbf{v}^{(k)} \circledast \widehat{\mathbf{x}}^{(k)} \tag{67}$$

where $\circledast$ denotes a convolutional operation, and $\{\mathbf{w}_1^{(k)}, \ldots, \mathbf{w}_m^{(k)}\}$ and $\mathbf{v}^{(k)}$ are learned convolutional kernels. The authors illustrate that their model-based DL architecture outperforms model-agnostic DL methods, yielding high-contrast and high-resolution outputs.

Deep unfolding for clutter filtering. Solomon et al. Solomon et al., 2019) proposed deep unfolded convolutional robust RPCA for ultrasound clutter suppression. The approach is derived from the RPCA algorithm, given by ((51) and (52), but unfolds it and learns all the parameters (gradient projection and regularization weights) from data. Each network layer in the unfolded architecture takes the form

$$\widehat{\mathbf{X}}_{\text{tissue}}^{(k+1)} = \mathscr{T}_{\text{SVT}}^{(k)}(\mathbf{W}_1^{(k)} \circledast \mathbf{Y} + \mathbf{W}_3^{(k)} \circledast \mathbf{X}_{\text{blood}}^{(k)} + \mathbf{W}_5^{(k)} \circledast \mathbf{X}_{\text{tissue}}^{(k)}) \tag{68}$$

and

$$\widehat{\mathbf{X}}_{\text{blood}}^{(k+1)} = \mathscr{T}_{\lambda}^{(k)}(\mathbf{W}_2^{(k)} \circledast \mathbf{Y} + \mathbf{W}_4^{(k)} \circledast \mathbf{X}_{\text{blood}}^{(k)} + \mathbf{W}_6^{(k)} \circledast \mathbf{X}_{\text{tissue}}^{(k)}) \tag{69}$$

where $\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \mathbf{W}_4, \mathbf{W}_5$ and $\mathbf{W}_6$ are trainable convolutional kernels. The resulting deep network has two distinct (model-based) non-linearities/activations per layer: the mixed $\ell_{1,2}$ thresholding and singular value thresholding. The authors trained the architecture end to end on a combination of simulations and RPCA results on real data, and determined that it outperforms a strong non-model-based deep network (a ResNet).

Deep unfolding for ultrasound localization microscopy. In the spirit of unfolding, van Sloun et al. (van Sloun et al., 2019) proposed unfolding their sparse recovery algorithm for ULM to enable accurate localization even for high concentrations of microbubbles. Similar to the previous examples of unfolding, each of the layers $\boldsymbol{k}$ in the resulting architecture takes the form

$$\mathbf{x}^{(k+1)} = \mathscr{T}_{\lambda^{(k)}}\left(\mathbf{W}_1^{(k)}\mathbf{y} + \mathbf{W}_2^{(k)}\mathbf{x}^{(k)}\right) \tag{70}$$

with $\mathbf{W}_1^{(k)}$ and $\mathbf{W}_2^{(k)}$ being trainable convolutional kernels. The authors trained this convolutional LISTA architecture on simulated envelope-detected RF US data comprising microbubbles under a distribution of point-spread functions. This method was later adopted by Bar-Shira et al. (Bar-Shira et al., 2021), who used it to perform localization microscopy in *in vivo* breast lesions.

Youn et al. (Youn et al., 2021) took this one step further and combined the image-domain LISTA architecture by van Sloun et al. (van Sloun et al., 2021) with the ABLE beamforming architecture of Luijten et al. (Luijten et al., 2020), training the joint network end-to-end. The authors found that this outperforms non-joint optimization and that ABLE learns to accommodate the downstream localization problem addressed by LISTA.

Speckle suppression using deep generative priors

Van de Schaft and van Sloun (Van de Schaft and van Sloun, 2021) formulate the task of speckle suppression as a MAP problem in which a clean image $\mathbf{x}$ is recovered from a measured, speckle-corrupted, image $\mathbf{y}$:

$$\widehat{\mathbf{x}} = \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) = \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}).$$

The authors proposed modeling the clean image prior $p(\mathbf{x})$ using a deep generative model (a normalizing flow) trained on magnetic resonance images, which naturally have no speckle but display similar anatomical structure. Under such a deep generative normalizing flow prior (with normalized hidden space $\mathbf{z}$), optimization is then performed in z-space:

$$\widehat{\mathbf{z}} = \arg\max_{\mathbf{z}} \log p\left(\mathbf{y}|f_\theta^{-1}(\mathbf{z})\right) + \log p(\mathbf{z}),$$

Here, $f_\theta^{-1}(\mathbf{z})$ is the inverse transformation of the normalizing flow $f_\theta(\mathbf{z})$, that is, the generative direction. By assuming a simple Gaussian likelihood model for the log-compressed envelope detected US images, this can be rewritten as

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}} \left\| f_\theta^{-1}(\mathbf{z}) - \mathbf{y} \right\|_2^2 + \lambda \left\| \mathbf{z} \right\|_2^2 \tag{71}$$

where λ is a parameter that depends on the assumed noise variance. Iterative optimization of eqn (71) was performed using gradient descent, and the recovered clean image is given by $\hat{\mathbf{x}} = f_\theta^{-1}(\hat{\mathbf{z}})$.

DISCUSSION

Over the past decade, the field of ultrasound signal processing has seen a large transformation, with the development of novel algorithms and processing methods. This development is driven in large part by the move from hardware- to software-based reconstruction. In this review, we have showcased several works, from conventional algorithms to full DL-based approaches, each having their own strengths and weaknesses.

Conventional model-based algorithms are based on first principles and offer a great amount of interpretability, which is relevant in clinical settings. However, as we have illustrated in this article, these methods rely on estimations and often simplifications of the underlying physics model, which result in sub-optimal signal reconstructions. For example, DAS beamforming assumes a linear measurement model and a Gaussian noise profile, both of which are very crude approximations of a realistic ultrasound measurement. In contrast, adaptive methods (*e.g.,* MV beamforming) that aim at modeling the signal statistics more accurately are often computationally expensive to implement in real-time applications.

Spurred by the need to overcome these limitations, we see a shift in research toward data-driven signal processing methods (mostly based on DL), a trend that started around 2014 (Zhang et al., 2021), which sees a significant increase in the number of peer-reviewed artificial intelligence (AI) publications. This can be explained by two significant factors: (i) the availability of high compute-power GPUs, and (ii) the availability of easy-to-use machine learning frameworks such as TensorFlow (Abadi et al., 2015) and PyTorch (Paszke et al., 2019), which have significantly lowered the threshold of entry into the field of AI for ultrasound researchers. However, the performance of data-driven and, more specifically, DL algorithms is inherently bounded by the availability of large amounts of high-quality training data. Acquiring ground truth data is not trivial in ultrasound beamforming and signal processing applications, and thus, simulations or the outputs of advanced yet slow model-based algorithms are often considered as training targets. Moreover, the lack of clear understanding of the behavior of learned models (*i.e.,* the black box model) and ability to predict their performance "in the wild" make implementations in clinical devices challenging.

These general challenges associated with fully data-driven DL methods have in turn spurred research in the field of "model-based deep learning." Model-based DL combines the model-based and data-driven paradigms and offers a robust signal processing framework. It enables learning those aspects of full models from data for which no adequate first-principles derivation is available or complementing/augmenting partial model knowledge. Compared with conventional deep neural networks, these systems often require a smaller number of parameters, and fewer training data, to learn an accurate input−output mapping.

Similar to model-based methods, we can broadly categorize model-based DL methods into (1) algorithms based on iterative solutions and (2) algorithms based on analytic solutions. Algorithms based on iterative solutions can be further split into (1a) truncated algorithms (fixed number of iterations) the parameters of which (possibly a subset of) are fine-tuned end-to-end (termed *deep unfolding/unrolling*), and (1b) iterative algorithms with data-driven priors (*e.g., plug-and-play optimization*). Examples of (1a) are deep unfolded ULM and deep unfolded robust PCA, but also unfolded ADMIRE beamforming. A recent example of (1b) is the work by Van de Schaft and van Sloun Van de Schaft and van Sloun, 2021), in which MRI-based image priors learned with normalizing flows are used for ultrasound speckle suppression. An example of method ((2) is ABLE, in which the analytic ML solution for beamforming under unknown non-diagonal covariance Gaussian channel noise is augmented with a neural network, and the entire hybrid solution is optimized end-to-end. The methods covered here aim to achieve a better imaging quality, for example, temporal or spatial resolution, ultimately aiding in the diagnosis process. Although a deeper analysis of the clinical relevance is a crucial and interesting topic, it is beyond the scope of this work.

CONCLUSIONS

In this review, we outlined the development of signal processing methods in US, from classic model-based algorithms to fully data-driven DL-based methods. We also discussed methods that lie at the intersection of these two approaches, using neural architectures inspired by model-based algorithms and derived from probabilistic inference problems. We take a probabilistic perspective, offering a generalized framework with which we can describe the multitude of approaches described in this article, all under the same umbrella. This provides us insight into the demarcation between components

derived from first principles and the components derived from data. This also affords us the ability to combine these components in a unique combination to derive architectures that integrate multiple classes of signal processing algorithms.

The application of such novel, DL-based reconstruction methods requires the next generation of US devices to be equipped accordingly: either by fast networking and on-device encoding, or by fully arming them with sufficient and appropriate processing power (GPUs and TPUs), which allows for flexible and real-time deployment of AI algorithms.

CONFLICT OF INTEREST DISCLOSURE

The authors declare that they have no conflicts of interest with respect to this work.

Acknowledgments—This work was supported in part by the Dutch Research Council (NWO) and Philips Research through the research program High Tech Systems and Materials (HTSM) under Project 17144.

REFERENCES

Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mane D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viegas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorow.org.

Al Kassir H, Zaharis ZD, Lazaridis PI, Kantartzis NV, Yioultsis TV, Xenos TD. A review of the state of the art and future challenges of deep learning–based beamforming. IEEE Access 2022;10:80869–80882.

Almansouri H, Venkatakrishnan S, Clayton D, Polsky Y, Bouman C, Santos-Villalobos H. Anisotropic modeling and joint-map stitching for improved ultrasound model-based iterative reconstruction of large and thick specimens. AIP Conf Proc 2018;1949 030002 a.

Almansouri H, Venkatakrishnan SV, Buzzard GT, Bouman CA, Santos-Villalobos H. Deep neural networks for non-linear model-based ultrasound reconstruction. 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP). New York. : IEEE; 2018. p. 6–10. b.

Ando K, Nagaoka R, Hasegawa H. Speckle reduction of medical ultrasound images using deep learning with fully convolutional network. Jón J Appl Phys 2020;59:SKKE06.

Austeng A, Nilsen CIC, Jensen AC, Nsholm SP, Holm S. Coherent planewave compounding and minimum variance beamforming. Proc IEEE Int Ultrason Symp 2011;2448–2451.

Bar-Shira O, Grubstein A, Rapson Y, Suhami D, Atar E, Peri-Hanania K, Rosen R, Eldar YC. Learned super resolution ultrasound for improved breast lesion characterization. International Conference on Medical Image Computing and Computer-Assisted Intervention MICCAI 2021). Lecture Notes in Computer Science, vol. 12907. Cham. : Springer; 2021. p. 109–118.

Bar-Zion A, Solomon O, Tremblay-Darveau C, Adam D, Eldar YC. SUSHI: Sparsity-based ultrasound super-resolution hemodynamic imaging. IEEE Trans Ultrason Ferroelectr Freq Control 2018;65:2365–2380.

Becker S, Bobin J, Cands EJ. NESTA: A fast and accurate first-order method for sparse recovery. SIAM J Imaging Sci 2011;4:1–39.

Bell M, Huang J, Wiacek A, Gong P, Chen S, Ramalli A, Tortoli P, Luijten B, Mischi M, Rindal OMH, Perrot V, Liebgott H, Zhang X, Luo J, Oluyemi E, Ambinder E. Challenge on ultrasound beamforming with deep learning (CUBDL) datasets. IEEE Dataport 2019; Available at: doi.org/10.21227/f0hn-8f92. Accessed July 10, 2022.

Bell MAL, Huang J, Hyun D, Eldar YC, van Sloun R, Mischi M. Challenge on ultrasound beamforming with deep learning (CUBDL). Proc IEEE Int Ultrason Symp 2020;1–5.

Brahme A. Comprehensive biomedical physics. 2. Amsterdam: Elsevier; 2014.

Brown KG, Ghosh D, Hoyt K. Deep learning of spatiotemporal filtering for fast super-resolution ultrasound imaging. IEEE Trans Ultrason Ferroelectr Freq Control 2020;67:1820–1829.

Brown KG, Waggener SC, Redfern AD, Hoyt K. Faster super-resolution ultrasound imaging with a deep learning model for tissue decluttering and contrast agent localization. Biomed Phys Eng Express 2021;7 065035.

Byram B, Dei K, Tierney J, Dumont D. A model and regularization scheme for ultrasonic beamforming clutter reduction. IEEE Trans Ultrason Ferroelectr Freq Control 2015;62:1913–1927.

Byram B. Aperture domain model image reconstruction (ADMIRE) for improved ultrasound imaging. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New York. : IEEE; 2017. p. 6250–6253.

Cai JF, Candes EJ, Shen Z. A singular value thresholding algorithm for matrix completion. SIAM J Optimization 2010;20:1956–1982.

Camacho J, Parrilla M, Fritsch C. Phase coherence imaging. IEEE Trans Ultrason Ferroelectr Freq Control 2009;56:958–974.

Chan V, Perlas A. Basics of ultrasound imaging. In: Narouze S, (ed). Atlas of ultrasound-guided procedures in interventional pain management. New York: Springer; 2011. p. 13–19.

Chang Y, Yan L, Chen M, Fang H, Zhong S. Two-stage convolutional neural network for medical noise removal via image decomposition. IEEE Trans Instrum Meas 2019;69:2707–2721.

Chennakeshava N, Luijten B, Drori O, Mischi M, Eldar YC, van Sloun RJ. High resolution plane wave compounding through deep proximal learning. Proc IEEE Int Ultrason Symp 2020;1–4.

Chennakeshava N, Luijten B, Mischi M, Eldar YC, van Sloun RJ. Deep proximal learning for high-resolution plane wave compounding. arXiv preprint arXiv:2112.12410, 2021.

Chernyakova T, Cohen D, Shoham M, Eldar YC. iMAP beamforming for high quality high frame rate imaging. IEEE Trans Ultrason Ferroelectr Freq Control 2019;66:1830–1844.

Chernyakova T, Eldar YC. Fourier-domain beamforming: The path to compressed ultrasound imaging. IEEE Trans Ultrason Ferroelectr Freq Control 2014;61:1252–1267.

Choi W, Kim M, HakLee J, Kim J, BeomRa J. Deep CNN-based ultrasound super-resolution for high-speed high-resolution b-mode imaging. Proc IEEE Int Ultrasonics Symp 2018;1–4.

Christensen-Jeffries K, Couture O, Dayton PA, Eldar YC, Hynynen K, Kiessling F, O'Reilly M, Pinton GF, Schmitz G, Tang MX, Tanter M, van Sloun RJC. Super-resolution ultrasound imaging. Ultrasound Med Biol 2020;46:865–891.

Cohen R, Eldar YC. Sparse convolutional beamforming for ultrasound imaging. IEEE Trans Ultrason Ferroelectr Freq Control 2018;65:2390–2406.

Cohen R, Eldar YC. Sparse array design via fractal geometries. IEEE Trans Signal Process 2020;68:4797–4812.

Combettes PL, Wajs VR. Signal recovery by proximal forward-backward splitting. Multiscale Model Simul 2005;4:1168–1200.

Coupe P, Hellier P, Kervrann C, Barillot C. Bayesian non local means-based speckle filtering. 5th IEEE International Symposium on Biomedical Imaging: From nano to macro. New York. : IEEE; 2008. p. 1291–1294.

Couture O, Besson B, Montaldo G, Fink M, Tanter M. Microbubble ultrasound super-localization imaging (MUSLI). Proc IEEE Int Ultrason Symp 2011;1285–1287.

Czerwinski RN, Jones DL, O'Brien WD. Ultrasound speckle reduction by directional median filtering. In: Proceedings, International Conference on Image Processing. Vol. 1. New York. : IEEE; 1995. p. 358–361.

Deylami AM, Jensen JA, Asl BM. An improved minimum variance beamforming applied to plane-wave imaging in medical ultrasound. Proc IEEE Int Ultrason Symp 2016;1–4.
Dietrichson F, Smistad E, Ostvik A, Lovstakken L. Ultrasound speckle reduction using generative adversial networks. 2018 IEEE International Ultrasonics Symposium 2018;1–4.
Dijkstra EW. A note on two problems in connexion with graphs. Numer Math 1959;1:269–271.
Ding T, Zhang S, Fu Q, Xu Z, Wan M. Ultrasound line-by-line scanning method of spatial−temporal active cavitation mapping for high-intensity focused ultrasound. Ultrasonics 2014;54:147–155.
Eldar Y. Sampling theory: Beyond bandlimited systems. Cambridge: Cambridge University Press; 2015.
Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. international conference on machine learning. PMLR 2016;48:1050–1059.
Garg A, Khandelwal V. Despeckling of medical ultrasound images using fast bilateral filter and NeighShrinkSure filter in wavelet domain. In: Rawat B, Trivedi A, Manhas S, Karwal V, (eds). Advances in Signal Processing and Communication. Lecture Notes in Electrical Engineering, vol. 526. Singapore: Springer; 2019. p. 271–280.
Gasse M, Millioz F, Roux E, Garcia D, Liebgott H, Friboulet D. High-quality plane wave compounding using convolutional neural networks. IEEE Trans Ultrason Ferroelectr Freq Control 2017;64:1637–1639.
Gregor K, LeCun Y. Learning fast approximations of sparse coding. In: Proceedings, 27th International Conference on International Conference on Machine Learning. Madison, WI. : Omnipress; 2010. p. 399–406.
Guo B, Zhang B, Ma Z, Li N, Bao Y, Yu D. High-quality plane wave compounding using deep learning for hand-held ultrasound devices. International Conference on Advanced Data Mining and Applications (ADMA 2020). Cham. : Springer; 2020. p. 547–559.
Hamelmann P, Vullings R, Kolen AF, Bergmans JW, van Laar JO, Tortoli P, Mischi M. Doppler ultrasound technology for fetal heart rate monitoring: A review. IEEE Trans Ultrason Ferroelectr Freq Control 2019;67:226–238.
Huh J, Khan S, Choi S, Shin D, Lee JE, Lee ES, Ye JC. Tunable image quality control of 3-D ultrasound using switchable CycleGAN. Med Image Anal 2022;83 102651.
Huijben IA, Veeling BS, Janse K, Mischi M, van Sloun RJ. Learning sub-sampling and signal recovery with applications in ultrasound imaging. IEEE Trans Med Imaging 2020;39:3955–3966.
Hyun D, Brickson LL, Looby KT, Dahl JJ. Beamforming and speckle reduction using neural networks. IEEE Trans Ultrason Ferroelectr Freq Control 2019;66:898–910.
Hyun D, Wiacek A, Goudarzi S, Rothlubbers S, Asif A, Eickel K, Eldar YC, Huang J, Mischi M, Rivaz H, Sinden D, van Sloun RJG, Strohm H, Bell MAL. Deep learning for ultrasound image formation: CUBDL evaluation framework and open datasets. IEEE Trans Ultrason Ferroelectr Freq Control 2021;68:3466–3483.
Jabarulla MY, Lee HN. Speckle reduction on ultrasound liver images based on a sparse representation over a learned dictionary. Appl Sci 2018;8:903.
Jagadesh T, Rani RJ. A novel speckle noise reduction in biomedical images using PCA and wavelet transform. 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). New York. : IEEE; 2016. p. 1335–1340.
Jansen G, Awasthi N, Schwab HM, Lopata R. Enhanced radon domain beamforming using deep-learning-based plane wave compounding. Proc IEEE Int Ultrason Symp 2021;1–4.
Jensen JA, Nikolov SI, Gammelmark KL, Pedersen MH. Synthetic aperture ultrasound imaging. Ultrasonics 2006;44:e5–e15.
Karaoglu O, Bilge HS. Uluer I. Removal of speckle noises from ultrasound images using five different deep learning networks. Eng Sci Technol Int J 2021; 101082.
Kervrann C, Boulanger J, Coupe P. Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal. In: Sgallari F, Murli A, Paragios N, (eds). Scale space and variational methods in computer vision. Lecture Notes in Computer Science, Vol. 4485. : Berlin/Heidelberg: Springer; 2007. p. 520–532.
Khan C, van Sloun RJ, Byram B. Performing aperture domain model image reconstruction (ADMIRE) using a deep neural network sparse encoder. Proc IEEE Int Ultrason Symp 2021;1–4 a.
Khan S, Huh J, Ye JC. Universal plane-wave compounding for high quality us imaging using deep learning. Proc IEEE Int Ultrason Symp 2019;2345–2347.
Khan S, Huh J, Ye JC. Adaptive and compressive beamforming using deep learning for medical ultrasound. IEEE Trans Ultrason Ferroelectr Freq Control 2020;67:1558–1572.
Khan S, Huh J, Ye JC. Switchable deep beamformer for ultrasound imaging using adain. 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). New York:. : IEEE; 2021. p. 677–680. b.
Kim Y, Kim J, Basoglu C, Winter T. Programmable ultrasound imaging using multimedia technologies: a next-generation ultrasound machine. IEEE Trans Inf Technol Biomed 1997;1:19–29.
Kim YM, Kim MG, Oh SH, Jung GI, Bae HM. Learning based approach for speed-of-sound adaptive rx beamforming. Proc IEEE Int Ultrason Symp 2021;1–4.
Li H, Mezheritsky T, Romaguera LV, Kadoury S. 3D B-mode ultrasound speckle reduction using deep learning for 3D registration applications. arXiv 2008.01147, 2020.
Li PC, Li ML. Adaptive imaging using the generalized coherence factor. IEEE Trans Ultrason Ferroelectr Freq Control 2003;50:128–141.
Liu S, Wang Y, Yang X, Lei B, Liu L, Li SX, Ni D, Wang T. Deep learning in medical ultrasound analysis: A review. Engineering 2019;5:261–275.
Liu X, Zhou T, Lu M, Yang Y, He Q, Luo J. Deep learning for ultrasound localization microscopy. IEEE Trans Med Imaging 2020;39:3064–3078.
Lu J, Millioz F, Garcia D, Salles S, Friboulet D. Fast diverging wave imaging using deep-learning-based compounding. Proc IEEE Int Ultrason Symp 2019;2341–2344.
Luchies AC, Byram BC. Deep neural networks for ultrasound beamforming. IEEE Trans Med Imaging 2018;37:2010–2021.
Luijten B, Cohen R, De Bruijn FJ, Schmeitz HA, Mischi M, Eldar YC, van Sloun RJ. Deep learning for fast adaptive beamforming. ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New York. : IEEE; 2019. p. 1333–1337.
Luijten B, Cohen R, De Bruijn FJ, Schmeitz HA, Mischi M, Eldar YC, van Sloun RJ. Adaptive ultrasound beamforming using deep learning. IEEE Trans Med Imaging 2020;39:3967–3978.
Mallart R, Fink M. Adaptive focusing in scattering media through soundspeed inhomogeneities: The Van Cittert Zernike approach and focusing criterion. J Acoust Soc Am 1994;96:3721–3732.
Mamistvalov A, Amar A, Kessler N, Eldar YC. Deep-learning based adaptive ultrasound imaging from sub-Nyquist channel data. IEEE Trans Ultrason Ferroelectr Freq Control 2021;69:1638–1648.
Mamistvalov A, Eldar YC. Deep unfolded recovery of sub-Nyquist sampled ultrasound image. IEEE Trans Ultrason Ferroelectr Freq Control 2021;68:3484–3496.
Monga V, Li Y, Eldar YC. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. IEEE Signal Process Mag 2021;38:18–44.
Nair AA, Tran TD, Reiter A, Bell MAL. A deep learning based alternative to beamforming ultrasound images. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New York. : IEEE; 2018. p. 3359–3363.
Nair AA, Washington KN, Tran TD, Reiter A, Lediju Bell MA. 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;67:2493–2509.
Nilsen CI, Holm S. Wiener beamforming and the coherence factor in ultrasound imaging. IEEE transactions on ultrasonics, ferroelectrics, and frequency control 2010;57(6):1329–1346 https://doi.org/10.1109/TUFFC.2010.1553.
Nugroho A, Hidayat R, Nugroho HA. Artifact removal in radiological ultrasound images using selective and adaptive median filter. In: Proceedings, 3rd International Conference on Cryptography, Security and Privacy. New York. : ACM; 2019. p. 237–241.
Oh S, Kim MG, Kim Y, Bae HM. A learned representation for multivariable ultrasonic lesion quantification. 2021 IEEE 18th

International Symposium on Biomedical Imaging (ISBI). New York. : IEEE; 2021. p. 1177–1181.
Ozkan E, Vishnevsky V, Goksel O. Inverse problem of ultrasound beam-forming with sparsity constraints and regularization. IEEE Trans Ultrason Ferroelectr Freq Control 2017;65:356–365.
Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S. Pytorch: An imperative style, high-performance deep learning library. In: Wallach H, Larochelle H, Bergelmir A, d'Alche-Buc F, Fox E, Garnett R, (eds). Advances in neural information processing systems 32. Red Hook, NY: Curran Associates; 2019. p. 8024–8035.
Qi Y, Guo Y, Wang Y. Image quality enhancement using a deep neural network for plane wave medical ultrasound imaging. IEEE Trans Ultrason Ferroelectr Freq Control 2020;68:926–934.
Raz R. On the complexity of matrix product. In: Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing: STOC '02. Red Hook, NY. : ACM; 2002. p. 144–151.
Routh HF. Doppler ultrasound. IEEE Eng Med Biol Mag 1996;15:31–40.
Shen H, Barthélémy C, Khoury E, Zemmoura Y, Reméenieras JP, Basarab A, Kouamé D. High-resolution and high-sensitivity blood ow estimation using optimization approaches with application to vascularization imaging. Proc IEEE Int Ultrason Symp 2019;467–470.
Shlezinger N, Whang J, Eldar YC, Dimakis AG. Model-based deep learning. arXiv preprint arXiv:2012.08405, 2020.
Siepmann M, Schmitz G, Bzyl J, Palmowski M, Kiessling F. Imaging tumor vascularity by tracing single microbubbles. Proc IEEE Int Ultrason Symp 2011;1906–1909.
Solomon O, Cohen R, Zhang Y, Yang Y, He Q, Luo J, van Sloun RJ, Eldar YC. Deep unfolded robust PCA with application to clutter suppression in ultrasound. IEEE Trans Med Imaging 2019;39:1051–1063 a.
Solomon O, van Sloun RJ, Wijkstra H, Mischi M, Eldar YC. Exploiting flow dynamics for superresolution in contrast-enhanced ultrasound. IEEE Trans Ultrason Ferroelectr Freq Control 2019;66:1573–1586 b.
Song JH, Lee J, Yeo S, Kim GD, Song TK. An analytical approach to designing optimal sparse 1-D phased arrays for handheld ultrasound imaging. IEEE Trans Ultrason Ferroelectr Freq Control 2020;67:1354–1365.
Szasz T, Basarab A, Kouamé D. Beamforming through regularized inverse problems in ultrasound medical imaging. IEEE Trans Ultrason Ferroelectr Freq Control 2016;63:2031–2044 a.
Szasz T, Basarab A, Vaida MF, Kouamé D. Elastic-Net based beamforming in medical ultrasound imaging. 13th IEEE International Symposium on Biomedical Imaging, ISBI 2016. Prague, Czech Republic, April 13−16, 2016. New York. : IEEE; 2016. p. 477–480. b.
Tabassian M, Hu X, Chakraborty B, Dhooge J. Clutter filtering using a 3D deep convolutional neural network. Proc IEEE Int Ultrason Symp 2019;2114–2117.
Temiz H, Bilge HS. Super resolution of B-mode ultrasound images with deep learning. IEEE Access 2020;8:78808–78820.
Thomenius K. Evolution of ultrasound beamformers. Proc IEEE Ultrason Symp 1996;2:1615–1622.
Van de Schaft V, van Sloun RJ. Ultrasound speckle suppression and denoising using MRI-derived normalizing flow priors. arXiv preprint arXiv:2112.13110, 2021.
van Sloun RJ, Cohen R, Eldar YC. Deep learning in ultrasound imaging. Proc IEEE 2019;108:11–29 a.
van Sloun RJ, Solomon O, Bruce M, Khaing ZZ, Eldar YC, Mischi M. Deep learning for super-resolution vascular ultrasound imaging. In: ICASSP 2019−2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New York: IEEE, 2019b:1055−1059.
van Sloun RJ, Solomon O, Bruce M, Khaing ZZ, Wijkstra H, Eldar YC, Mischi M. Super-resolution ultrasound localization microscopy through deep learning. IEEE Trans Med Imaging 2020;40:829–839.
van Sloun RJ, Solomon O, Eldar YC, Wijkstra H, Mischi M. Sparsity-driven super-resolution in clinical contrast-enhanced ultrasound. Proc IEEE Int Ultrason Symp 2017;1–4.
van Sloun RJ, Ye JC, Eldar YC. Deep learning for ultrasound beamforming. arXiv preprint arXiv:2109.11431, 2021.
Van Trees H. Detection, estimation, and modulation theory: Part I. Detection, estimation, and linear modulation theory. New York: Wiley; 2004.
Vedula S, Senouf O, Bronstein AM, Michailovich OV, Zibulevsky M. Towards CT-quality ultrasound imaging using deep learning. arXiv preprint arXiv:1710.06304, 2017.
Vignon F, Shin JS, Meral FC, Apostolakis I, Huang SW, Robert JL. Resolution improvement with a fully convolutional neural network applied to aligned per-channel data. Proc IEEE Int Ultrason Symp 2020;1–4.
Wang H, Gao S, Mozumi M, Omura M, Nagaoka R, Hasegawa H. Preliminary investigation on clutter filtering based on deep learning. Jón J Appl Phys 2021;60:SDDE21.
Wiacek A, González E, Bell MAL. Coherenet: A deep learning architecture for ultrasound spatial correlation estimation and coherence-based beamforming. IEEE Trans Ultrason Ferroelectr Freq Control 2020;67:2574–2583.
Wildeboer RR, Sammali F, van Sloun RJG, Huang Y, Chen P, Bruce M, Rabotti C, Shulepov S, Salomon G, Schoot BC, Wijkstra H, Mischi M. Blind source separation for clutter and noise suppression in ultrasound imaging: Review for different applications. IEEE Trans Ultrason Ferroelectr Freq Control 2020;67:1497–1512.
Ylitalo JT, Ermert H. Ultrasound synthetic aperture imaging: monostatic approach. IEEE Trans Ultrason Ferroelectr Freq Control 1994;41:333–339.
Youn J, Luijten B, Schou M, Stuart MB, Eldar YC, van Sloun RJ, Jensen JA. Model-based deep learning on ultrasound channel data for fast ultrasound localization microscopy. Proc IEEE Int Ultrason Symp 2021;1–4.
Youn J, Ommen ML, Stuart MB, Thomsen EV, Larsen NB, Jensen JA. Detection and localization of ultrasound scatterers using convolutional neural networks. IEEE Trans Med Imaging 2020;39:3855–3867.
Zhang D, Mishra S, Brynjolfsson E, Etchemendy J, Ganguli D, Grosz B, Lyons T, Manyika J, Niebles JC, Sellitto M, Shoham Y, Clark J, Perrault R. The AI index 2021 annual report. arXiv 2103.06312, 2021.
Zhou Z, Wang Y, Yu J, Guo Y, Guo W, Qi Y. High spatial−temporal resolution reconstruction of plane-wave ultrasound images with a multichannel multiscale convolutional neural network. IEEE Trans Ultrason Ferroelectr Freq Control 2018;65:1983–1996.