Published in final edited form as:

IEEE Trans Med Imaging. 2020 October ; 39(10): 3079–3088. doi:10.1109/TMI.2020.2986762.

Nondestructive Detection of Targeted Microbubbles Using Dual-Mode Data and Deep Learning for Real-Time Ultrasound Molecular Imaging

Dongwoon Hyun [Member, IEEE],

Department of Radiology, Stanford University, Stanford, CA 94305 USA

Lotfi Abou-Elkacem,

Stanford University, Stanford, CA 94305 USA.

Rakesh Bam,

Department of Radiology, Stanford University, Stanford, CA 94305 USA

Leandra L. Brickson

Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA.

Carl D. Herickhoff [Member, IEEE], Jeremy J. Dahl [Member, IEEE]

Department of Radiology, Stanford University, Stanford, CA 94305 USA

Abstract

Ultrasound molecular imaging (UMI) is enabled by targeted microbubbles (MBs), which are highly reflective ultrasound contrast agents that bind to specific biomarkers. Distinguishing between adherent MBs and background signals can be challenging *in vivo*. The preferred preclinical technique is differential targeted enhancement (DTE), wherein a strong acoustic pulse is used to destroy MBs to verify their locations. However, DTE intrinsically cannot be used for real-time imaging and may cause undesirable bioeffects. In this work, we propose a simple 4-layer convolutional neural network to *nondestructively* detect adherent MB signatures. We investigated several types of input data to the network: “anatomy-mode” (fundamental frequency), “contrast-mode” (pulse-inversion harmonic frequency), or both, i.e., “dual-mode”, using IQ channel signals, the channel sum, or the channel sum magnitude. Training and evaluation were performed on *in vivo* mouse tumor data and microvessel phantoms. The dual-mode channel signals yielded optimal performance, achieving a soft Dice coefficient of 0.45 and AUC of 0.91 in two test images. In a volumetric acquisition, the network best detected a breast cancer tumor, resulting in a generalized contrast-to-noise ratio (GCNR) of 0.93 and Kolmogorov-Smirnov statistic (KSS) of 0.86, outperforming both regular contrast mode imaging (GCNR=0.76, KSS=0.53) and DTE imaging (GCNR=0.81, KSS=0.62). Further development of the methodology is necessary to distinguish free from adherent MBs. These results demonstrate that neural networks can be trained to detect

Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

The data is available upon request to the corresponding author. (dongwoon.hyun@stanford.edu).

L. Abou-Elkacem is now with Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030 USA.

targeted MBs with DTE-like quality using nondestructive dual-mode data, and can be used to facilitate the safe and real-time translation of UMI to clinical applications.

Index Terms—

Image enhancement/restoration (noise and artifact reduction); Machine learning; Molecular and cellular imaging; Neural network; Ultrasound

I. Introduction

ULTRASOUND is a widely used medical imaging modality that provides anatomical and functional information about soft tissue. Ultrasound images are typically obtained using endogenous mechanisms of contrast: for example, B-mode, Doppler, and elastography images display the anatomy, motion, and stiffness of the underlying tissue, respectively. Conventional ultrasound can be augmented to achieve ultrasound molecular imaging (UMI) by introducing exogenous contrast agents such as targeted microbubbles (MBs) [1]–[6]. MBs, which produce strong reflections to ultrasound, are gas bubbles with lipid or polymer shells that are several micrometers in diameter. The shells can be chemically engineered to adhere to specific molecules of interest, producing *targeted* MBs. A specialized harmonic pulse-echo sequence is then used to detect any adherent targeted MBs, which would imply the presence of the targeted molecules [7], [8].

UMI has been used to image diseases such as inflammatory bowel disease, myocardial ischemia, and atherosclerosis [9] as well as to detect precursors of diabetes [10]. UMI is also being investigated as a tool for early cancer detection. A common target is VEGFR-2, a biomarker associated with the development of tumor neovasculature, which has been applied towards the detection of breast, prostate, and ovarian cancers in animal models [11]–[16] and more recently, in humans [17], [18]. Another target of interest is B7-H3, a biomarker demonstrated to distinguish between normal, benign, precursor, and malignant breast pathologies for UMI [19]. For each desired imaging target, a rigorous biochemical process is used to verify binding specificity of the targeted MBs. These efforts must be complemented by a highly sensitive and specific imaging technique that can detect the adherent targeted MBs.

A major challenge in UMI is the isolation of the adherent targeted MB signal from background signals due to tissue, noise, and circulating MBs. Current approaches can be loosely grouped into four major categories: spectral analysis, temporal analysis, beamforming, and physical perturbation. Spectral analyses use the broadband acoustic backscatter of targeted MBs to isolate them from the comparatively narrowband background tissue signal. Examples include pulse-inversion harmonic imaging [20], “contrast pulse sequences” [7], subharmonic imaging [21], and acoustic angiography [22]. Temporal analyses isolate adherent targeted MBs based on so-called “slow-time” observations acquired over multiple frames. Examples include frame averaging, dwell-time imaging [23], minimum intensity projection [24], and singular value decomposition-based techniques [25], [26]. Beamforming methods combine the “channel data” acquired by the different transducer array elements into images. Examples include short-lag spatial coherence [27], which

identifies MBs according to second-order statistics across the channels, and transmit aperture synthesis [8], [28]. A fourth approach is to physically perturb the MBs, e.g., via acoustic radiation force to enhance binding and/or to help differentiate between tissue, adherent MBs, and free MBs [29]–[32].

A widely-used technique for targeted MB detection is differential targeted enhancement (DTE) imaging [1], [9], wherein a strong destructive pulse is used to burst the targeted MBs and images acquired post-burst are subtracted from images acquired pre-burst, thus removing the background signals and leaving behind only the targeted MB signal (see Fig. 1). DTE imaging is commonly employed in preclinical UMI studies; however, bursting of the targeted MBs can damage the vasculature and surrounding tissue [33], and may have additional bioeffects that are yet undiscovered. In first-in-human UMI studies [17], [18], destructive pulses were not used due to concerns for patient safety, leading to poor tissue background suppression. Moreover, destructive pulses intrinsically cannot be used for real-time imaging. Each time the targeted MBs are destroyed, they must be replenished and given time to bind to the biomarkers (often upwards of 5 min.), leading to long examination times.

A *nondestructive* UMI technique would allow the clinician to freely interrogate the tissue for MBs in real time until they can arrive at a diagnosis. The normalized singular spectrum area (NSSA) has been proposed as a nondestructive metric that combines spectral analysis, temporal analysis, and physical perturbation methods to classify between tissue, adherent targeted MBs, and free circulating MBs with performance comparable to DTE [26], suggesting that the nondestructive data contain enough information to achieve high-quality UMI, provided it can be extracted properly.

Artificial neural networks have emerged as powerful tools for extracting desired information from specified inputs. In the field of computer vision, deep learning methods have recently outperformed traditional human-engineered algorithms in a variety of image classification tasks [34], [35]. In ultrasound image reconstruction, deep learning has been applied to techniques that utilize pre-image data (e.g., complex in-phase and quadrature (IQ) channel data, rather than the magnitude of the IQ channel sum). In [36], a neural network was trained to estimate the echogenicity of backscatter from channel signals, resulting in speckle-reduced B-mode imaging. Neural networks have been used to reduce clutter in ultrasound channel signals using a frequency-domain [37] or convolutional [38] approach. Other neural networks have attempted to reproduce a high quality plane wave synthetic aperture image using a reduced number of transmissions [39], [40]. Neural networks have also been used for flow and elastography applications [41], [42].

Previously, we demonstrated a proof of concept of a nondestructive deep learning beamformer for UMI using a limited dataset of dual-frequency channel signals [43]. In this work, we expand on the methodology and compare this approach to other configurations, including those readily available for real-time UMI. Here, we use a fully convolutional neural network to transform various forms of anatomy and contrast mode data into estimates of MB probability, and apply the network in an *in vivo* study of UMI in a mouse model of early cancer detection.

II. Methods

A. Overview of Proposed Detection Methodology

The proposed MB detection methodology is illustrated in Fig. 2 and is summarized as follows. Consider a $P \times Q$ grid of image points, referred to as “pixels”, and denote the ground truth of MB presence at each pixel as a binary mask $\mathbf{y} \in \{0, 1\}^{PQ}$. Focused ultrasound data, denoted as $\mathbf{X} \in \mathcal{T}^{PQ}$, is acquired at each pixel with data type $\mathcal{T}$, where $\mathcal{T}$ is a real or complex vector of signals (see Sec. II-D). A MB detector uses $\mathbf{X}$ to estimate the MB probability $\hat{\mathbf{y}} \in [0, 1]^{PQ}$. The quality of the estimate is quantified by a loss function $\mathcal{L}(\hat{\mathbf{y}}, \mathbf{y})$.

In this study, MBs were detected using a neural network f with parameters θ , i.e., $f(\mathbf{X}; \theta) = \hat{\mathbf{y}}$. Gradient descent was used to find the θ that minimized a training loss function measured over training and validation datasets. MB detection performance of the trained network was then tested in previously unobserved UMI data. In particular, we investigated how the choice of $\mathcal{T}$ affected the trainability and the test performance of the neural network.

B. Ultrasound Molecular Imaging Protocol

In vivo UMI data were acquired in two mouse models of cancer. All *in vivo* experiments were performed with prior approval from the Institutional Administrative Panel on Laboratory Animal Care at Stanford University. The first mouse model was hepatocellular carcinoma (HCC) in xenografted subcutaneous tumors, where imaging was performed using VEGFR-2-targeted BR55 MBs (Bracco, Milan, Italy). Human HCC xenografts were established on the mouse flanks by subcutaneously injecting 5×10^6 HepG2 cells mixed in 50 μL low growth factor matrigel membrane matrix (BD Biosciences, Billerica, MA). Tumors were allowed to grow for 2–3 weeks until they reached a mean maximum diameter of 8 mm (range, 4–8 mm). The second model was a transgenic mouse model for breast cancer development [44]–[46], where imaging was performed using B7-H3-targeted MBs developed in-house [47]. MBs were injected via the tail vein and allowed to circulate for 7 min. prior to imaging to provide sufficient time for targeted MBs to bind and for free MBs to be cleared. Acquisitions were performed in mouse tumors 7 min. post-injection, as well as in a tissue-mimicking microvessel phantom (positive control) and a mouse abdomen prior to MB injection (negative control). A subset of the acquisitions were performed with a degassed 1 cm layer of porcine abdominal tissue to mimic clinical imaging conditions.

A Verasonics Vantage 256 research scanner and an L12–3v transducer were used to transmit a sequence of angled plane wave transmissions and to obtain radiofrequency (RF) signals from 128 transducer elements. Each acquisition was composed of 25 plane wave steering angles uniformly distributed over a 10° angular span. Plane wave synthetic transmit aperture was used to insonify the entire field of view and achieve dynamic transmit and receive focusing while avoiding forming a physical focus [8], [28]. Two imaging modes were employed: “anatomy” and “contrast”. Anatomy mode imaging was performed using a 2-cycle 10 MHz transmit pulse (i.e., fundamental frequency imaging). Contrast mode was performed using two 2-cycle 5 MHz transmit pulses with opposite polarity whose received signals were subsequently summed (i.e., pulse inversion harmonic imaging). The transmit voltage was selected so as to maximize the MB harmonic signal while suppressing the tissue

harmonic signal; the selected voltages produced pulses with mechanical index (MI) of < 0.1 . The three transmit pulses were emitted at each steering angle and the resulting echoes were recorded on all receive channels before proceeding to the next angle. The pulse repetition frequency was 10 kHz, and the acquisition frame rate was 100 frames per second. All received signals were bandpass filtered at 10 MHz with 80% fractional bandwidth. The RF signals were demodulated and focused (i.e., delayed but not summed) into a 240×180 pixel grid (axial-by-azimuth) with a density of 3 pixels per wavelength, corresponding to an approximately 12 mm depth-by-9 mm width box. Every pixel in the pixel grid was associated with 128 channels of complex IQ signals corresponding to the 128 transducer elements.

C. Ground Truth

The ground truth (y) was obtained by manual segmentation of DTE on contrast-mode image data. After the MBs were allowed to bind to the imaging target, 100 frames of raw channel data were acquired. The transmit voltage was increased until MBs were eliminated from the field of view, and a second 100 frames of raw channel data were acquired. Pre- and post-destruction images were reconstructed by selecting 30 contiguous frames with minimal motion artifacts, averaging across the frames, and beamforming. The post-destruction images were subtracted from the pre-destruction images, leaving behind only the MB signal in the DTE image (see Fig. 1). DTE images were formed using both the delay-and-sum (DAS) and short-lag spatial coherence (SLSC) [27] beamformers. SLSC has been shown to yield higher SNR of targeted MBs through the suppression of background noise [27], and is therefore expected to yield a higher quality ground truth. For the SLSC images, the maximum lag threshold was set to 12 elements of separation. Each resulting image was manually segmented using a custom semi-automated MATLAB script to highlight MBs and eliminate artifacts, where the segmentation depended on both signal amplitude and distance from user-input seed positions. Finally, the ground truth MB signals were selected as the intersection of the DAS and SLSC segmentations so as to include only the high-amplitude and high-coherence targets. This intersection rejected low-amplitude and high-coherence targets or high-amplitude and low-coherence targets that may be attributed to false-positives caused by tissue or noise, respectively.

D. Ultrasound Pixel Data Types

Raw ultrasound data undergoes several processing steps before arriving at the final image. We investigated the impact of using focused pixel data from three different points within the processing workflow: 1) the complex IQ channel signals; 2) the complex IQ channel sum (i.e. RF beamsum); and 3) the magnitude of the complex IQ channel sum. Furthermore, we varied the input data between anatomy-mode only, contrast-mode only, or a combination of both (i.e., “dual-mode”). Dual-mode data was formed by concatenating the anatomy and contrast mode data in the channel dimension. The real and imaginary components of complex values were treated as distinct inputs, such that N complex channels were treated as $2N$ real inputs. Analysis was limited to a maximum of $N=16$ complex channels, which were obtained by subdividing the aperture into subapertures of equal size and obtaining their channel sums. Table I lists the nine different pixel data types used along with the actual number of real scalar inputs per pixel that were input into the network.

E. Training and Validation Datasets

A total of 29 UMI acquisitions were split into training and validation sets of 24 and 5 acquisitions, respectively. Acquisitions were performed in different locations and tumors to avoid inadvertent reuse of highly correlated data between training and validation datasets. For each acquisition, two frames of pre-destruction data were selected at random to obtain two realizations of thermal noise with the same ground truth. The datasets were then augmented another two-fold by using a constant $\pi/3$ radian complex phase rotation over the entire dataset while utilizing the same ground truth. This trained the neural network to be invariant to complex phase rotations. Finally, the datasets were augmented another two-fold by applying a flip in both the azimuth and channel dimensions, taking advantage of the inherent symmetry in azimuth. (Note that the depth dimension is fundamentally asymmetric due to effects such as attenuation and refraction, and was thus not used for data augmentation.) In all, the training and validation sets were augmented 8-fold, yielding a total of 192 training samples and 40 validation samples. Each sample was then converted into the nine pixel data types described above in Table I.

F. Neural Network Architecture and Metrics

Fully convolutional networks were implemented in Tensor-Flow Keras and were used to estimate pixel-wise probability of MB presence. The network consisted of M repeated “convolution blocks”, wherein each block was comprised of a 2D convolution layer over the image pixel dimensions, a rectified linear unit (ReLU) activation layer, and a batch normalization layer. In this analysis, $M=4$ was selected. Each convolution layer was composed of 32 kernels of size $3 \times 3 \times N_m$, where N_m was the number of inputs to the m -th layer. The last convolution block was followed by a final convolution layer with 2 kernels of size $1 \times 1 \times 32$, followed by a softmax operator applied to the outputs to obtain $\hat{y}$. The architecture is depicted in the inset of Fig. 2. The focus of this study was to investigate how the choice of data type affected the training and performance of a simple fixed neural network. As such, the architecture parameters were selected heuristically with the intention of using a simple convolutional neural network that could be trained rapidly on a single GPU and eventually deployed in real-time. Default values were used for hyperparameters, such as initializers.

Detection performance was quantified using the area under the receiver operating characteristic curve (AUC) and the soft Dice coefficient (SDC), defined as

$$\text{SDC}(\hat{y}, y) = \frac{\sum_p^{PQ} 2\hat{y}_p y_p + \epsilon}{\sum_p^{PQ} \hat{y}_p + y_p + \epsilon}, \quad (1)$$

where $\epsilon = 10^{-10}$ was used as a smoothing parameter. A higher AUC and SDC indicate better detection performance. We used SDC rather than the standard Dice coefficient (which compares two binary images) because like AUC, SDC does not depend on the choice of segmentation threshold.

G. Training and Hyperparameter Optimization

Networks were trained to minimize a mixture of the soft Dice loss and cross-entropy loss functions:

$$\mathcal{L}_{\text{Train}}(\hat{\mathbf{y}}, \mathbf{y}) = (1 - \beta)\mathcal{L}_{\text{XEnt}}(\hat{\mathbf{y}}, \mathbf{y}) + \beta\mathcal{L}_{\text{Dice}}(\hat{\mathbf{y}}, \mathbf{y}), \quad (2)$$

where $\beta = 0.3$ was selected heuristically, and the component loss functions were defined as

$$\mathcal{L}_{\text{Dice}}(\hat{\mathbf{y}}, \mathbf{y}) = 1 - \text{SDC}(\hat{\mathbf{y}}, \mathbf{y}) \quad (3)$$

$$\mathcal{L}_{\text{XEnt}}(\hat{\mathbf{y}}, \mathbf{y}) = - \sum_p^{PQ} y_p \log \hat{y}_p + (1 - y_p) \log (1 - \hat{y}_p). \quad (4)$$

Each network was trained for up to 125 epochs, with early termination in the case of no improvement in validation loss for 30 epochs.

For each of the nine pixel data types, a total of 100 training runs were performed. A different random seed was used for each run. Bayesian hyperparameter optimization was performed to find the optimal learning rate in a range from 10^{-5} to 10^0 using the “hyperopt” software package [48]. Bayesian optimization utilizes the accumulated history of training results to model the hyperparameter space and to interrogate hyperparameters that are expected to yield the greatest improvements. The optimal network was selected as the network (out of 100) that maximized the SDC in the validation dataset.

H. Testing of Optimal Networks

The optimal networks for each of the nine pixel data types were tested in two previously unseen datasets: an acquisition in a subcutaneous HCC tumor with VEGFR-2-targeted MBs and an acquisition in a breast cancer tumor with B7-H3-targeted MBs. Both acquisitions were acquired with the same protocol as the training and validation data.

Additionally, a volumetric UMI scan was acquired across the abdomen of a transgenic mouse with a breast tumor by affixing the L12-3v transducer to a motor-controlled linear translation stage (UniSlide MA25 and VXM Controller, Velmex Inc., Bloomfield, NY). Seven minutes after B7-H3-targeted MBs were injected via the tail vein, the transducer was swept a distance of 1 cm in the elevation dimension. A total of 250 frames of anatomy and contrast mode channel data were acquired. Afterwards, the transducer was returned to its initial position and was swept across the same volume while emitting destructive pulses. Finally, the transducer was returned to its initial position once more and swept across the same volume to acquire a post-destruction dataset. Four images were reconstructed: 1) a regular B-mode volume using the 10 MHz fundamental data; 2) a nondestructive pulse-inversion harmonic image; 3) a DTE image with DAS beamforming; and 4) a nondestructive UMI volume using the optimal network with the dual-mode IQ channel sum pixel data type. For the DTE image, rigid volume registration (translation only) was performed to align the post-destruction volume with the pre-destruction volume via the `imregister` function in MATLAB. We selected rigid volume registration (as opposed to frame-by-frame image

registration) to correct for global displacements caused by our translation stage setup and to reduce the impact of uncorrectable out-of-plane artifacts due to respiratory and cardiac motion. Unlike the training data, which were averaged over 30 frames, the DTE images in the swept acquisition were formed from single frames.

Because the image volumes were reconstructed using different methods and in different units (i.e., magnitudes by DAS, probabilities by the network), MB detection performance was quantified and compared using metrics that are fundamentally invariant to the underlying image value distributions: the generalized contrast-to-noise ratio (GCNR) [49] and the Kolmogorov-Smirnov statistic (KSS) [50], [51], defined as

$$\text{GCNR} = 1 - \int \min\{f_{\text{tumor}}(x), f_{\text{tissue}}(x)\} dx \quad (5)$$

$$\text{KSS} = \sup_x |F_{\text{tumor}}(x) - F_{\text{tissue}}(x)|, \quad (6)$$

where f_a and F_a respectively are the probability distribution (i.e. histogram) and cumulative probability distribution of region of interest (ROI) a . The tumor ROI was selected as a sphere in the center of tumor, and the background tissue ROI was selected as a rectangular block in the tissue. The GCNR provides a measure of histogram overlap between the ROIs, where 0 indicates complete overlap (i.e. indistinguishable) and 1 indicates no overlap (i.e., completely different). Similarly, the KSS measures the maximum difference in the cumulative distribution functions of the two ROIs and is a statistical test of whether samples in the two ROIs are drawn from the same underlying probability distribution. Both metrics provide an indication of how easily samples can be classified as originating from the tumor or tissue ROIs, regardless of any monotonic transformation applied to the entire image (e.g., grayscale remapping or histogram matching) [49]–[51]. We avoided transformation-sensitive metrics such as the signal-to-background ratio (the root-mean-square of the tumor ROI divided by the root-mean-square of the tissue ROI), which are more appropriate for comparing images with the same units.

III. Results

A. Training and Hyperparameter Optimization

Fig. 3 plots the results of hyperparameter optimization as a function of input pixel data type. The columns from left to right show the anatomy, contrast, and dual modes, and the rows from top to bottom show the 16 channel, channel sum, and magnitude of channel sum data types. The SDC is plotted as a function of learning rate. For each configuration, 100 markers are plotted, each representing the maximum achieved validation SDC over a single 125 epoch training session along with its corresponding learning rate. Using the anatomy-mode data, the SDC did not exceed 0.3 for any training run. Using 16 channels of contrast-mode data, the SDC was approximately 0.4 for learning rates between 10^{-4} to 10^{-2} ; lower values were reported when using the contrast-mode channel sum or magnitude. Using dual-mode data achieved the highest SDCs, with the 16 channel and channel sum data types yielding consistent performance for learning rates from 10^{-4} to 10^{-2} over all random seeds, and the

magnitude dual-mode data achieving high SDCs for a subset of the random seeds at learning rates around 10^{-3} .

B. Evaluation of Optimal Networks on Test Data

The SDCs and AUCs measured over the two test cases are reported in Tab. II for the validation-optimal networks. When using anatomy-mode data only, the network performed no better than random chance ($AUC = .50$). Better detection performance was achieved when using only contrast-mode data, especially when 16 channels were employed. The top performing networks (denoted in boldface) utilized dual-mode 16 channel and channel sum data.

Example images from a test sample of VEGFR-2-targeted MBs in a HCC tumor are displayed in Fig. 4. Figs. 4a and 4b show the anatomy and contrast images formed prior to MB destruction, respectively. The anatomy image clearly visualized the skin surrounding the tumor in the center of the image as well as the layer of porcine tissue at the top of image. The contrast image darkened the skin and porcine tissue and enhanced the inside of the tumor. Fig. 4c shows the DTE image, where a post-destruction contrast image (not pictured) was subtracted from the pre-destruction contrast image. The MBs were significantly enhanced and the tissue background suppressed. Fig. 4d depicts the result of the subsequent manual segmentation that was used as the ground truth. Fig. 4e shows the test predictions produced by the validation-optimal network for each of the nine pixel data types. Using anatomy-mode data, the network failed to detect MBs in all cases. The 16 channel contrast-mode data resulted in some MB detection, but also resulted in false positives in the porcine tissue (top of image), which did not contain any MBs. The dual-mode data produced images with the most visual similarity to the DTE image, particularly when using the 16 channel and channel sum data.

Fig. 5 shows example images from another test sample of B7-H3-targeted MBs in a breast tumor. As in Fig. 4, Figs. 5a–5d show the anatomy, contrast, DTE SLSC, and manually-segmented ground truth images, respectively. Fig. 5e shows the predictions of the nine different pixel data types within the tumor. The anatomy-mode data produced predictions that did not resemble the underlying MB ground truth image and contained many false positives and negatives. The contrast-mode data produced some accurate predictions of MB positions, but also introduced many false positives, indicating poor specificity. The dual-mode data produced predictions that closely resembled the DTE image for all three pixel data types (16 channel, channel sum, and magnitude).

C. Volumetric UMI

Fig. 6 displays maximum intensity projection renderings of a volumetric acquisition of a breast tumor in a transgenic mouse. In the video (available online), the volumes rotate to provide 3D perspective. Fig. 6a shows the anatomy-mode reconstruction, with the tumor appearing as a dark protuberance through the skin. Fig. 6b overlays the nondestructive contrast-mode volume, where some enhancement within the tumor was visible but was accompanied by significant nonspecific enhancement of the surrounding tissue, particularly in the highly echogenic tissue near the bottom of the volume. Compared to the contrast-

mode volume, the DTE image (Fig. 6c) better suppressed background tissue while enhancing MB signals inside the tumor. The tissue below the tumor was also enhanced, but it is unclear whether this signal is due to actual B7-H3 expression or to the hyperechogenicity of the tissue. Fig. 6d displays the volume reconstructed by the nondestructive neural network using dual-mode channel sum data, and shows strong and localized MB enhancement within the tumor.

The GCNR and KSS are listed in Table III and indicate the how well the tumor and tissue ROIs can be distinguished. The anatomy-mode weakly distinguished between the tumor and tissue ROIs, although negative contrast was observed (i.e., the tissue was brighter on average than the tumor). The contrast-mode produced positive contrast (i.e. brighter tissue than tumor), with slightly higher GCNR and KSS than the anatomy-mode. DTE further increased GCNR and KSS by 7% and 18% over contrast-mode, respectively. The dual-mode neural network produced the best separation of tumor and background tissue, increasing GCNR and KSS by 22% and 62% over contrast-mode, respectively.

IV. Discussion

In this study, nine different pixel data types were tested. The results suggest that contrast-mode data is essential to MB detection. Through 100 training attempts during hyperparameter optimization (Fig. 3), MBs were detected only when contrast mode data was included. The results also showed that dual-mode data yielded better performance than contrast-mode alone, indicating that although the anatomy data was not useful for MB detection by itself, it was useful when used jointly with contrast data. Figure 3 also indicated that training was more robust when using more raw forms of contrast-mode and dual-mode data. Interestingly, the detection performance of the 16 channel contrast-mode data was surpassed by the channel-sum dual-mode data, suggesting that the addition of the anatomy pixel data outweighed the benefit of using more raw forms of contrast-mode data. In additional experiments not presented here, we observed the same trends even when replacing the 10 MHz anatomy-mode data with the 5 MHz positive polarity pulse data of the contrast-mode acquisition (i.e., dual-mode data wherein one of the two contrast pulses is also used as the anatomy-mode data).

An intuitive explanation for the superiority of dual-mode data is that the anatomy and contrast modes exhibit differential behavior for tissue and MBs: the fundamental frequency of anatomy mode enhances tissue more strongly than MBs, whereas the harmonic frequency of contrast mode enhances nonlinear MB responses more strongly than those of tissue. It is difficult to design an algorithm that can leverage this knowledge using traditional methods. For instance, the tissue signal cannot be removed by simply subtracting the anatomy image from the contrast image due to the mismatch in speckle patterns and tissue echogenicities in the two images. Deep learning is an excellent tool for these types of applications, where the primary challenge lies in the extraction of patterns known *a priori* to be present within the data.

The trained network significantly outperformed conventional contrast-mode imaging and produced images qualitatively similar to DTE imaging. In Figs. 4–6, a significant amount of

background tissue was visible in the contrast image, whereas both DTE and the dual-mode network suppressed the background tissue and enhanced only the MBs. In the volumetric acquisition (Fig. 6), the network better enhanced the tumor and suppressed signal outside as compared to DTE. DTE was likely degraded because it required three sweeps of the transducer, possibly leading to reperfusion of free MBs and imprecise alignment due to nonrigid deformations. The DTE volume might be improved by using respiratory-gated acquisitions and a faster and more precise translation stage. By contrast, the nondestructive nature of the neural network enabled its UMI volume to be produced from a single sweep, reducing motion artifacts, allowing for repeated acquisitions, and alleviating concerns about potential bioeffects.

The channel count in this study was limited due to data size and computational constraints. As a consequence, each of the 16 complex channels in the dual-mode configuration corresponded to a subaperture of 16 transducer elements. Previously-reported correlation lengths of MB signals across the aperture *in vivo* [27] suggest that smaller subapertures (and consequently a higher channel count) should be used to exploit channel-based information for MB detection in the future. Currently, the channel sum remains a more practical implementation because of the difficulty of accessing channel data in real-time on the majority of commercial ultrasound systems.

We did not observe any benefit to utilizing more complex network architectures than the simple one used in this study. Furthermore, we observed in an omitted retrospective analysis of the hyperparameter optimization in Fig. 3 that the optimal network in validation generally did not produce the optimal network in testing, indicating that overfitting was taking place. We hypothesize that there are substantial errors in the ground truth DTE images that must be addressed prior to introducing more sensitive network architectures, which are further prone to overfitting. Incorrect and noisy ground truth labels negatively impact neural network training to a greater extent than noisy inputs, and are currently an active topic of research in the machine learning community [52]. Despite the excellent MB enhancement and tissue suppression achieved by DTE (see Figs. 1, 4c, and 5c), image labeling was challenging; many DTE signals were ambiguous and could have been considered as MB, tissue, and/or noise. The DTE images here were constructed using a rudimentary 30-frame time average (corresponding to a sub-second interval), and would likely benefit from more sophisticated filtering. A more controlled *in vitro* setup could also improve our knowledge of the ground truth, but may introduce a statistical mismatch in the distribution of training samples versus the testing environment.

Further development of the methodology is required to assess the detection of adherent MBs. Previously, *adherent* MBs have been distinguished from free MBs using multi-frame methods such as dwell-time imaging [23] and minimum intensity projection [24], and more recently, in combination with acoustic radiation force via NSSA [25], [26]. These nondestructive methods highlight the value of temporal information that is currently unused by the proposed neural network, which uses only a single frame of data per image. In particular, NSSA highlights how ultrasound can be used to actively interrogate the medium to access information about MB binding states. Neural networks provide a flexible framework that can incorporate these additional sources of information with the appropriate

architecture and training. We hypothesize that incorporating temporal information and refining the ground truth will enhance the sensitivity and specificity of adherent MB detection. Contrary to recent trends in computer vision where deep learning has largely replaced traditional feature-engineering, the presented deep learning approach strongly motivates the need for more feature-engineering and a more precise ground truth.

V. Conclusion

Neural networks were developed for the purpose of achieving safe and real-time UMI. Networks were designed to utilize nondestructive anatomy-mode and/or contrast-mode data. For each case, 16 complex channels, the complex channel sum, or the magnitude of the channel sum magnitude was used as pixel data. A fully convolutional neural network was used to produce pixel-wise estimates of MB probability. The network was trained to minimize mixture of the $\mathcal{L}_{\text{XEnt}}$ and $\mathcal{L}_{\text{Dice}}$ loss functions. A manually segmented DTE image served as the ground truth. In hyperparameter optimization over 100 training runs for the nine different pixel data types, the network learned to detect MBs only when provided with contrast-mode data. Specifically, dual-mode (both anatomy and contrast) data resulted in the highest validation SDC. When using the validation-optimal network in previously unobserved test data, a similar trend was observed: the anatomy-only modes failed to detect MBs; the contrast-only mode detected MBs but gave many false positives; and the dual-mode detected MBs achieved the highest AUC and SDC. In particular, the 16 channel dual-mode data resulted in SDC=.45 and AUC=.91. In a volumetric acquisition, a network using channel sum dual-mode data detected MB signals within a known tumor (GCNR=.93, KSS=.86) better than both contrast-mode and DTE. These results demonstrate highly accurate nondestructive detection of MBs using neural networks, and can potentially be used to enable safe and real-time translation of UMI to clinical applications.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by National Cancer Institute grant R01-CA218204 and by a seed grant from the Stanford Cancer Institute, a NCI-designated Comprehensive Cancer Center.

References

- [1]. Lindner JR, Song J, Xu F, Klivanov AL, Singbartl K, Ley K, and Kaul S, "Noninvasive ultrasound imaging of inflammation using microbubbles targeted to activated leukocytes," *Circulation*, vol. 102, no. 22, pp. 2745–2750, 2000. [PubMed: 11094042]
- [2]. Lindner JR, "Molecular imaging with contrast ultrasound and targeted microbubbles," *Journal of Nuclear Cardiology*, vol. 11, no. 2, pp. 215–221, 2004. [PubMed: 15052252]
- [3]. Klivanov AL, "Ligand-carrying gas-filled microbubbles: ultrasound contrast agents for targeted molecular imaging," *Bioconjug Chem*, vol. 16, no. 1, pp. 9–17, 2005. [PubMed: 15656569]
- [4]. Dayton PA and Rychak JJ, "Molecular ultrasound imaging using microbubble contrast agents," *Frontiers in Bioscience*, vol. 12, pp. 5124–5142, 2007. [PubMed: 17569635]

- [5]. Lindner JR, "Molecular imaging of cardiovascular disease with contrast-enhanced ultrasonography," *Nature Reviews Cardiology*, vol. 6, no. 7, pp. 475–481, 2009. [PubMed: 19506587]
- [6]. Deshpande N, Needles A, and Willmann JK, "Molecular ultrasound imaging: current status and future directions," *Clin Radiol*, vol. 65, no. 7, pp. 567–581, 2010. [PubMed: 20541656]
- [7]. Phillips P, "Contrast pulse sequences (CPS): imaging nonlinear microbubbles," 2001 IEEE Ultrasonics Symposium. Proceedings. An International Symposium (Cat. No.01CH37263), vol. 2, pp. 1739–1745, 2001.
- [8]. Couture O, Fink M, and Tanter M, "Ultrasound contrast plane wave imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 59, no. 12, pp. 2676–2683, 2012.
- [9]. Abou-Elkacem L, Bachawal SV, and Willmann JK, "Ultrasound molecular imaging: Moving toward clinical translation," *Eur J Rad*, vol. 84, no. 9, pp. 1685–1693, 2015.
- [10]. Chadderdon SM, Belcik JT, Bader L, Kirigiti MA, Peters DM, Kievit P, Grove KL, and Lindner JR, "Proinflammatory endothelial activation detected by molecular imaging in obese nonhuman primates coincides with onset of insulin resistance and progressively increases with duration of insulin resistance," *Circulation*, vol. 129, no. 4, pp. 471–478, 2014. [PubMed: 24163066]
- [11]. Willmann JK, Paulmurugan R, Chen K, Gheysens O, Rodriguez-Porcel M, Lutz AM, Chen IY, Chen X, and Gambhir SS, "US imaging of tumor angiogenesis with microbubbles targeted to vascular endothelial growth factor receptor type 2 in mice." *Radiology*, vol. 246, no. 2, pp. 508–518, 2008. [PubMed: 18180339]
- [12]. Fischer T, Thomas A, Tardy I, Schneider M, Hünigen H, Custodis P, Beyersdorff D, Plendl J, Schnorr J, Diekmann F, and Gemeinhardt O, "Vascular endothelial growth factor receptor 2-specific microbubbles for molecular ultrasound detection of prostate cancer in a rat model," *Investigative Radiology*, vol. 45, no. 10, pp. 675–684, 2010. [PubMed: 20733504]
- [13]. Pochon S, Tardy I, Bussat P, Bettinger T, Brochot J, von Wronski M, Passantino L, and Schneider M, "BR55: a lipopeptide-based VEGFR2-targeted ultrasound contrast agent for molecular imaging of angiogenesis," *Invest Radiol*, vol. 45, no. 2, pp. 89–95, 2010. [PubMed: 20027118]
- [14]. Bzyl J, Lederle W, Rix A, Grouls C, Tardy I, Pochon S, Siepmann M, Penzkofer T, Schneider M, Kiessling F, and Palmowski M, "Molecular and functional ultrasound imaging in differently aggressive breast cancer xenografts using two novel ultrasound contrast agents (BR55 and BR38)," *Eur Radiol*, vol. 21, no. 9, pp. 1988–1995, 2011. [PubMed: 21562807]
- [15]. Bachawal SV, Jensen KC, Lutz AM, Gambhir SS, Tranquart F, Tian L, and Willmann JK, "Earlier detection of breast cancer with ultrasound molecular imaging in a transgenic mouse model," *Cancer Res*, vol. 73, no. 6, pp. 1689–1698, 2013. [PubMed: 23328585]
- [16]. Barua A, Yellapa A, Bahr JM, Machado SA, Bitterman P, Basu S, Sharma S, and Abramowicz JS, "VEGFR2-Targeted Ultrasound Imaging Agent Enhances the Detection of Ovarian Tumors at Early Stage in Laying Hens, a Preclinical Model of Spontaneous Ovarian Cancer," *Ultrasonic Imaging*, vol. 37, no. 3, pp. 224–237, 2015. [PubMed: 25294846]
- [17]. Willmann JK, Bonomo L, Carla Testa A, Rinaldi P, Rindi G, Valluru KS, Petrone G, Martini M, Lutz AM, and Gambhir SS, "Ultrasound Molecular Imaging With BR55 in Patients With Breast and Ovarian Lesions: First-in-Human Results," *Journal of Clinical Oncology*, vol. 35, no. 19, pp. 2133–2140, 2017. [PubMed: 28291391]
- [18]. Smeenge M, Tranquart F, Mannaerts CK, De Reijke TM, Van De Vijver MJ, Laguna MP, Pochon S, De La Rosette JJ, and Wijkstra H, "First-in-human ultrasound molecular imaging with a VEGFR2-specific ultrasound molecular contrast agent (BR55) in prostate cancer a safety and feasibility pilot study," *Investigative Radiology*, vol. 52, no. 7, pp. 419–427, 2017. [PubMed: 28257340]
- [19]. Bachawal SV, Jensen KC, Wilson KE, Tian L, Lutz AM, and Willmann JK, "Breast cancer detection by B7-H3-targeted ultrasound molecular imaging," *Cancer Res*, vol. 75, no. 12, pp. 2501–2509, 2015. [PubMed: 25899053]
- [20]. Simpson DH, Chin CT, and Burns PN, "Pulse inversion Doppler: a new method for detecting nonlinear echoes from microbubble contrast agents," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 46, no. 2, pp. 372–382, 1999.

- [21]. Forsberg F, Shi WT, and Goldberg BB, "Subharmonic imaging of contrast agents," *Ultrasonics*, vol. 38, no. 1, pp. 93–98, 2000. [PubMed: 10829636]
- [22]. Gessner RC, Frederick CB, Foster FS, and Dayton PA, "Acoustic angiography: A new imaging modality for assessing microvasculature architecture," *International Journal of Biomedical Imaging*, vol. 2013, 2013.
- [23]. Pysz MA, Guracar I, Tian L, and Willmann JK, "Fast microbubble dwell-time based ultrasonic molecular imaging approach for quantification and monitoring of angiogenesis in cancer," *Quant Imaging Med Surg*, vol. 2, no. 2, pp. 68–80, 2012. [PubMed: 22943043]
- [24]. Daeichin V, Akkus Z, Skachkov I, Kooiman K, Needles A, Sluimer J, Janssen B, Daemen MJ, Van Der Steen AF, De Jong N, and Bosch JG, "Quantification of bound microbubbles in ultrasound molecular imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 62, no. 6, pp. 1190–1200, 2015.
- [25]. Mauldin FW, Dhanaliwala AH, Patil AV, and Hossack JA, "Real-time targeted molecular imaging using singular value spectra properties to isolate the adherent microbubble signal," *Physics in medicine and biology*, vol. 57, no. 16, pp. 5275–5293, 2012. [PubMed: 22853933]
- [26]. Herbst EB, Unnikrishnan S, Klibanov AL, Mauldin FW, and Hossack JA, "Validation of Normalized Singular Spectrum Area as a Classifier for Molecularly Targeted Microbubble Adherence," *Ultrasound in Medicine and Biology*, vol. 45, no. 9, pp. 2493–2501, 2019. [PubMed: 31227262]
- [27]. Hyun D, Abou-Elkacem L, Perez VA, Chowdhury SM, Willmann JK, and Dahl JJ, "Improved sensitivity in ultrasound molecular imaging with coherence-based beamforming," *IEEE Transactions on Medical Imaging*, vol. 37, no. 1, pp. 241–250, 2018. [PubMed: 29293430]
- [28]. Montaldo G, Tanter M, Bercoff J, Benech N, and Fink M, "Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 56, no. 3, pp. 489–506, 2009.
- [29]. Dayton P, Klibanov A, Brandenburger G, and Ferrara K, "Acoustic radiation force in vivo: A mechanism to assist targeting of microbubbles," *Ultrasound in Medicine and Biology*, vol. 25, no. 8, pp. 1195–1201, 1999. [PubMed: 10576262]
- [30]. Zhao S, Borden M, Bloch SH, Kruse D, Ferrara KW, and Dayton PA, "Radiation-force assisted targeting facilitates ultrasonic molecular imaging," *Molecular Imaging*, vol. 3, no. 3, pp. 135–148, 2004. [PubMed: 15530249]
- [31]. Rychak JJ, Klibanov AL, Ley KF, and Hossack JA, "Enhanced Targeting of Ultrasound Contrast Agents Using Acoustic Radiation Force," *Ultrasound in Medicine and Biology*, vol. 33, no. 7, pp. 1132–1139, 2007. [PubMed: 17445966]
- [32]. Herbst E, Unnikrishnan S, Wang S, Klibanov AL, Hossack JA, and Mauldin FW Jr, "The use of acoustic radiation force decorrelation-weighted pulse inversion (adw-pi) for enhanced ultrasound contrast imaging," *Investigative radiology*, vol. 52, no. 2, p. 95, 2017. [PubMed: 27495188]
- [33]. Miller DL, Averkiou MA, Brayman AA, Everbach EC, Holland CK, Jr JHW, and Wu J, "Bioeffects considerations for diagnostic ultrasound contrast agents," *Journal of Ultrasound in Medicine*, vol. 27, no. 4, pp. 611–632, 2008. [PubMed: 18359911]
- [34]. Krizhevsky A, Sutskever I, and Hinton GE, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, 2012, pp. 1097–1105.
- [35]. LeCun Y, Bengio Y, and Hinton G, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. [PubMed: 26017442]
- [36]. Hyun D, Brickson LL, Looby KT, and Dahl JJ, "Beamforming and speckle reduction using neural networks," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 5, pp. 898–910, 2019.
- [37]. Luchies AC and Byram BC, "Deep Neural Networks for Ultrasound Beamforming," *IEEE Transactions on Medical Imaging*, vol. 37, no. 9, pp. 2010–2021, 2018. [PubMed: 29994441]
- [38]. Brickson LL, Hyun D, and Dahl JJ, "Reverberation Noise Suppression in the Aperture Domain Using 3D Fully Convolutional Neural Networks," *IEEE International Ultrasonics Symposium, IUS*, vol. 2018-10, pp. 1–4, 2018.

- [39]. Perdios D, Besson A, Arditi M, and Thiran JP, "A deep learning approach to ultrasound image recovery," IEEE International Ultrasonics Symposium, IUS, 2017.
- [40]. Gasse M, Millioz F, Roux E, Garcia D, Liebgott H, and Friboulet D, "High-quality plane wave compounding using convolutional neural networks," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 64, no. 10, pp. 1637–1639, 2017.
- [41]. Peng B, Xian Y, and Jiang J, "A Convolution Neural Network-Based Speckle Tracking Method for Ultrasound Elastography," IEEE International Ultrasonics Symposium, IUS, vol. 2018-10, pp. 206–212, 2018.
- [42]. Van Sloun RJ, Belt H, Janse K, and Mischi M, "Learning Doppler with Deep Neural Networks and its Application to Intra-Cardiac Echography," IEEE International Ultrasonics Symposium, IUS, vol. 2018-1, pp. 1–4, 2018.
- [43]. Hyun D, Brickson LL, Abou-Elkacem L, Bam R, and Dahl JJ, "Nondestructive targeted microbubble detection using a dual-frequency beamforming deep neural network," in 2018 IEEE International Ultrasonics Symposium (IUS), 10 2018, pp. 1–9.
- [44]. Fluck MM and Schaffhausen BS, "Lessons in signaling and tumorigenesis from polyomavirus middle T antigen," Microbiol Mol Biol Rev, vol. 73, pp. 542–563, 2009. [PubMed: 19721090]
- [45]. Guy CT, Cardiff RD, and Muller WJ, "Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease," Mol Cell Biol, vol. 12, pp. 954–961, 1992. [PubMed: 1312220]
- [46]. Lin EY, Jones JG, Li P, Zhu L, Whitney KD, Muller WJ, and Pollard JW, "Progression to malignancy in the polyoma middle T oncoprotein mouse breast cancer model provides a reliable model for human diseases," Am J Pathol, vol. 163, pp. 2113–2126, 2003. [PubMed: 14578209]
- [47]. Bam R, Lown P, Stern L, Sharma K, Wilson K, Lutz A, Paulmurugan R, Hackel B, Dahl J, and Abou-Elkacem L, "Affibody Ligand Targeting Vascular B7-H3 for Early Detection of Breast Cancer by Ultrasound Molecular Imaging," Clinical Cancer Research, (submitted).
- [48]. Bergstra J, Yamins D, and Cox D, "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures," in Proc. 30th International Conference on Machine Learning (ICML-13), 2013.
- [49]. Rodriguez-Molares A, Rindal OMH, Dhooge J, Måsøy S-E, Austeng A, Bell MAL, and Torp H, "The generalized contrast- to-noise ratio: a formal definition for lesion detectability," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 2019.
- [50]. Geman D, Geman S, Graffigne C, and Dong P, "Boundary detection by constrained optimization," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 7, pp. 609–628, 7 1990.
- [51]. Muzzolini R, Yang Y, and Pierson R, "Multiresolution texture segmentation with application to diagnostic ultrasound images," IEEE Transactions on Medical Imaging, vol. 12, no. 1, pp. 108–123, 3 1993. [PubMed: 18218399]
- [52]. Frénay B and Verleysen M, "Classification in the presence of label noise: A survey," IEEE Transactions on Neural Networks and Learning Systems, vol. 25, no. 5, pp. 845–869, 2014. [PubMed: 24808033]

Fig. 1. Destruction-subtraction imaging is depicted. Images are acquired (a) before and (b) after a strong destructive pulse. The image in (b) is subtracted from the image in (a) to produce the (c) DTE image.

Fig. 2.

The proposed detection methodology (Sec. II-A) and the architecture of the neural network (Sec. II-F) are depicted.

Fig. 3.

The SDCs measured in the validation dataset are plotted as a function of learning rates for 100 independent training runs for each of the 9 pixel data types. Each marker denotes a single training run (up to 125 epochs) for each configuration. Networks trained using anatomy-only pixel data resulted in $SDC < 0.3$. The contrast-only networks resulted in a maximum SDC of approximately 0.4 when using the 16 channel configuration. The dual-mode networks frequently yielded maximum $SDC > 0.5$ with 16 channel and channel sum data and less consistently with the magnitude data.

Fig. 4.

An example from the test dataset. The (a) anatomy and (b) contrast image before MB destruction. The complex IQ signals used to form these images were utilized as input to the network. (c) The DTE image formed after MB destruction using SLSC beamforming. This image was used for manual segmentation to form the (d) ground truth. The white bar denotes 1 mm. (e) The network images for the nine different pixel data types, shown using a linear grayscale with a range of 0 to 1, where the grayscale value denotes the probability of the presence of MBs. The networks using dual-mode channel signal ($SDC=0.48$) and channel sum ($SDC=0.39$) yielded the best agreement with the ground truth. The contrast-mode channel signal ($SDC=0.37$) also agreed well, although some false positives are visible at the top of the image. Networks using only anatomy-mode data or magnitude data failed to reliably detect MBs.

Fig. 5.

A second example from the test dataset. The white bars indicate 1 mm. (a) The anatomy-mode image of the full tumor before MB destruction. Subsequent images are restricted to the rectangle of interest. (b) The contrast-mode image before MB destruction. The complex IQ signals used to form (a) and (b) were used as input to the networks. (c) A DTE image formed after MB destruction using SLSC beamforming. (d) Manual segmentation of the DTE image used as ground truth. (e) The network predictions are shown in grayscale as a range of probabilities from 0 to 1 for the nine different input pixel data types. Networks using dual-mode data performed best (SDC=0.37, 0.27, 0.20 for 16 channels, channel sum, and magnitude, respectively). Networks with contrast-mode data detected some MBs but resulted in many false positives throughout. Networks using anatomy-mode failed to detect MBs.

Fig. 6.

Maximum intensity projections of a volumetric UMI acquisition in a mouse with a breast tumor are shown (video available online). In the video, the volumes rotate to provide 3D perspective. (a) The anatomy-mode volume is shown, with the tumor visible as a dark protuberance on the skin. The spherical tumor ROI (magenta) and the rectangular tissue ROI (green) are also plotted. (b) The contrast-mode volume is overlaid. (c) DTE UMI with DAS beamforming is overlaid. (d) The nondestructive neural network predictions (thresholded at MB probability > 0.95) are overlaid with the B-mode volume. The nondestructive neural network best localizes the UMI signal inside the tumor.

TABLE I

Ultrasound Pixel Data Types

Data Type ($\mathcal{T}$)	Anatomy	Contrast	Dual
16 Channels	$\mathbb{C}^{16} \rightarrow \mathbb{R}^{32}$	$\mathbb{C}^{16} \rightarrow \mathbb{R}^{32}$	$\mathbb{C}^{8 \times 2} \rightarrow \mathbb{R}^{32}$
Channel Sum	$\mathbb{C}^1 \rightarrow \mathbb{R}^2$	$\mathbb{C}^1 \rightarrow \mathbb{R}^2$	$\mathbb{C}^{1 \times 2} \rightarrow \mathbb{R}^4$
Magnitude	$\mathbb{R}^1$	$\mathbb{R}^1$	$\mathbb{R}^2$

TABLE II

Average Metrics Over In Vivo Test Cases

Input Type	Anatomy		Contrast		Dual	
	SDC	AUC	SDC	AUC	SDC	AUC
16 Channels	0.07	0.43	0.29	0.82	0.45	0.91
Channel sum	0.10	0.50	0.16	0.76	0.35	0.89
Magnitude	0.07	0.42	0.08	0.61	0.19	0.84

TABLE III

Volumetric UMI Metrics

MB Detection Method	GCNR	KSS
Anatomy-Mode Volume	0.71	0.39
Contrast-Mode Volume	0.76	0.53
DTE Volume	0.81	0.62
Dual-Mode Network Volume	0.93	0.86