Deep-Learning-Enabled Microwave-Induced Thermoacoustic Tomography Based on ResAttU-Net for Transcranial Brain Hemorrhage Detection

Chenzhe Li, *Graduate Student Member, IEEE*, Zijun Xi, Gaofei Jin, Weichao Jiang, Baosheng Wang, Xiran Cai, and Xiong Wang, *Member, IEEE*

***Abstract*—*Objective:* Hemorrhagic stroke is a leading threat to human's health. The fast-developing microwave-induced thermoacoustic tomography (MITAT) technique holds potential to do brain imaging. However, transcranial brain imaging based on MITAT is still challenging due to the involved huge heterogeneity in speed of sound and acoustic attenuation of human skull. This work aims to address the adverse effect of the acoustic heterogeneity using a deep-learning-based MITAT (DL-MITAT) approach for transcranial brain hemorrhage detection. *Methods:* We establish a new network structure, a residual attention U-Net (ResAttU-Net), for the proposed DL-MITAT technique, which exhibits improved performance as compared to some traditionally used networks. We use simulation method to build training sets and take images obtained by traditional imaging algorithms as the input of the network. *Results:* We present *ex-vivo* transcranial brain hemorrhage detection as a proof-of-concept validation. By using an 8.1-mm thick bovine skull and porcine brain tissues to perform *ex-vivo* experiments, we demonstrate that the trained ResAttU-Net is capable of efficiently eliminating image artifacts and accurately restoring the hemorrhage spot. It is proved that the DL-MITAT method can reliably suppress false positive rate and detect a hemorrhage spot as small as 3 mm. We also study effects of several factors of the DL-MITAT technique to further reveal its robustness and limitations. *Conclusion:* The proposed ResAttU-Net-based DL-MITAT method is promising for mitigating the acoustic inhomogeneity issue and performing transcranial brain hemorrhage detection. *Significance:* This work provides a novel ResAttU-Net-based DL-MITAT paradigm and paves a compelling route for transcranial brain hemorrhage detection as well as other transcranial brain imaging applications.**

***Index Terms*—Acoustic inhomogeneity, brain hemorrhage, deep learning (DL), microwave imaging, microwave-induced thermoacoustic tomography, transcranial brain imaging.**

I. Introduction

MICROWAVE-induced thermoacoustic tomography (MITAT) has experienced remarkable progression and fostered a broad research interest in biomedical applications since its debut in 1981 [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], MITAT felicitously integrates several intriguing features, such as excellent image resolution, rich microwave absorption contrast, nonionizing irradiation, deep penetration in tissues, and cost-effective implementation.

Brain stroke is a major threat to human's health, which can be classified into ischemic stroke and hemorrhagic stroke [16], [17]. The former is featured by restricted/blocked blood flow due to blood clots that results in decreased dielectric properties; the latter is caused by rupture/burst brain vessel that results in increased dielectric properties. According to literature, almost 80% of brain stroke cases are ischemic and the remaining are hemorrhagic [17]. Although hemorrhagic stroke only accounts for a minority of the total cases, it results in higher mortality rates than ischemic stroke. This is because progression with bleeding and expansion of hematoma of hemorrhagic stroke is associated with worse outcomes [18].

Timely detection of hemorrhagic stroke is of paramount significance to reduce the mortality and disability. Current brain imaging techniques mainly include MRI and CT [19] However, these techniques still suffer from some drawbacks. MRI systems are costly and its availability is low in undeveloped regions. CT utilizes ionizing radiation and is not suitable for long-time imaging and monitoring. In addition, MRI and CT equipment cannot be made portable to be carried by an ambulance to offer timely detection.

In view of the unique merits, MITAT has been explored as an emerging brain imaging modality [7], [20], [21], [22], [23]. Huang et al. first studied the feasibility of MITAT for

Manuscript received 18 April 2022; revised 18 August 2022, 28 October 2022, and 18 December 2022; accepted 1 January 2023. Date of publication 10 February 2023; date of current version 19 July 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 61971287, in part by the Open Project of State Key Laboratory of Millimeter Waves under Grant K202207, and in part by the Double First-Class Initiative Fund of ShanghaiTech University under Grant SYLDX035 2022. *(Corresponding author: Xiong Wang.)*

Chenzhe Li, Zijun Xi, Gaofei Jin, Weichao Jiang, Baosheng Wang, and Xiran Cai are with the School of Information Science and Technology, ShanghaiTech University, China.

Xiong Wang is with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China (e-mail: wangxiong@shanghaitech.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/TBME.2023.3243491, provided by the authors.

Digital Object Identifier 10.1109/TBME.2023.3243491

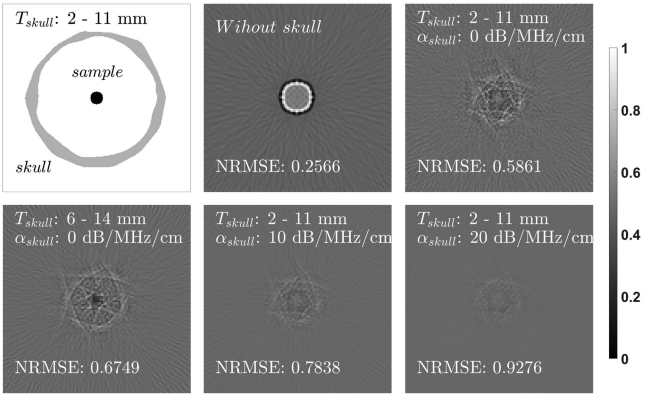

Fig. 1. Illustration of the adverse effects of a skull on the reconstructed images by MITAT. The first subfigure is the schematic of the used irregular-shaped skull with a round sample to be imaged. The rest subfigures are DAS images with different skull configurations. T_{skull} denotes the thickness variation range of the skull and α_{skull} denotes the attenuation coefficient of the skull.

hemorrhagic stroke imaging [24]. However, the major challenge for transcranial brain hemorrhage detection is the big acoustic inhomogeneity induced by the skull. To be specific, the speed of sound (SOS) of soft human tissues (such as skin and brain tissues) ranges from 1350 to 1700 m/s [25], while the SOS of skulls can vary from 2500 to 2900 m/s [26]. The acoustic attenuation in skulls ranges from 17 to 38 dB/MHz/cm [27], while the attenuation of brain tissues is around 0.59 dB/MHz/cm. These can cause strong refraction, phase distortion and attenuation in thermoacoustic signals. Such signal deterioration renders the conventional MITAT algorithms infeasible for transcranial brain hemorrhage detection as illustrated in Fig. 1, including back-projection [28], [29], delay-and-sum (DAS) [30], compressive sensing [31], [32], etc. In addition, the shapes of human skulls are irregular and the nonuniform thicknesses vary with age and gender [33], which further complicates the problem.

Although some methods have been proposed to address the acoustic inhomogeneity issue in MITAT, they still suffer from some drawbacks. The techniques requiring the prior knowledge of the skull shape [7], SOS and attenuation distributions are not practical since such information is usually unknown in clinical applications [20]. Even if it is possible to measure the SOS and attenuation distributions in advance [27], [34], [35], it significantly increases the system complexity and time consumption. A new method based on SOS autofocus has been reported to solve this problem without extra system cost [21]. However, it is only applicable to samples with moderate variations in SOS and not able to handle transcranial brain imaging. Therefore, there is a pressing need in exploring novel methods to deal with the high acoustic SOS and attenuation and enable transcranial brain hemorrhage detection by MITAT.

In the past three years, successful implementations of deep learning (DL) in diversified imaging and image processing related areas have been reported, including MRI [36], [37], CT [38], [39], PET [40], photoacoustic imaging [41], [42], [43], and microwave imaging [44], [45], [46]. In this work, we propose a novel DL-enabled MITAT (DL-MITAT) technique based on a new network structure to mitigate the adverse effects of acoustic heterogeneities in transcranial brain hemorrhage detection. The exercised network is a residual attention U-Net (ResAttU-Net), which is an improved version of the conventional U-Net. We provide detailed information of the training process of the network and implementation of the DL-MITAT technique. We test the network by two kinds of skull via *ex-vivo* experiments, a 3-D printed polymer skull and an 8.1-mm thick bovine skull, both containing a porcine brain sample containing a hemorrhage spot phantom. The obtained results corroborate that the images offered by the trained network can reveal the hemorrhage spot with high reliability even in the presence of the skull, which are in great contrast to the severely distorted DAS images. The DL-MITAT technique can suppress false positive rate and detect a hemorrhage spot as small as 3 mm. We further compare the image quality using different networks, evaluate the effect of number of training sets, and study some cases with mismatched training and testing sets. We also test the method by an ischemic stroke sample and find that it is able to distinguish between the hemorrhage and ischemia. This work provides a paradigm for the novel DL-MITAT modality. It also paves the way for practical applications of transcranial brain hemorrhage detection (and hopefully brain ischemia) in a cost-effective and portable manner.

II. Rationale

A. Fundamentals of MITAT

Time-resolved thermoacoustic signal emanated from a sample upon pulsed microwave irradiation is [15], [28]

$$p\left(\vec{r},t\right)=\frac{\beta}{4\pi C_p}\iiint\frac{A\left(\vec{r}_0\right)}{\left|\vec{r}-\vec{r}_0\right|}\left[I'\left(t\right)\right]_{t=t-\frac{\left|\vec{r}-\vec{r}_0\right|}{c}}d\vec{r}_0, \tag{1}$$

where $p(\vec{r},t)$ (Pa) denotes the acoustic pressure, β (K^{-1}) represents the volumetric thermal expansion coefficient, C_p ($\mathrm{J{\cdot}K^{-1}{\cdot}kg^{-1}}$) denotes the specific heat capacity, c ($\mathrm{m{\cdot}s^{-1}}$) is the SOS, $A(\vec{r})$ is the spatial distribution of microwave power, and $I(t)$ represents the envelope of the power of the microwave pulse [47]. The imaging algorithms of MITAT essentially make use of the measured $p(\vec{r},t)$ at different transducer locations to solve for or reconstruct $A(\vec{r})$, which is referred to as an image. For a spherical measurement setup in a homogeneous acoustic environment, the solution of $A(\vec{r})$ can be expressed as [28]

$$A\left(\vec{r}\right)=-\frac{r^2C_p}{2\pi\beta c^4}\iint_W\frac{1}{t}\frac{\partial p\left(\vec{r},t\right)}{\partial t}\bigg|_{t=\frac{\left|\vec{r}_0-\vec{r}\right|}{c}}dW. \tag{2}$$

But for an inhomogeneous acoustic environment, the traditional methods inevitably incur artifacts as shown in Fig. 1.

B. Structure of Applied Network

U-Net is one of the most widely explored multi-stage cascaded end-to-end fully convolutional networks that has performed favorably in diversified imaging related disciplines [48]. Residual U-Net, a modified version of U-Net, establishes a skip connection between the input and output data to render the network with faster convergence and higher robustness [49].

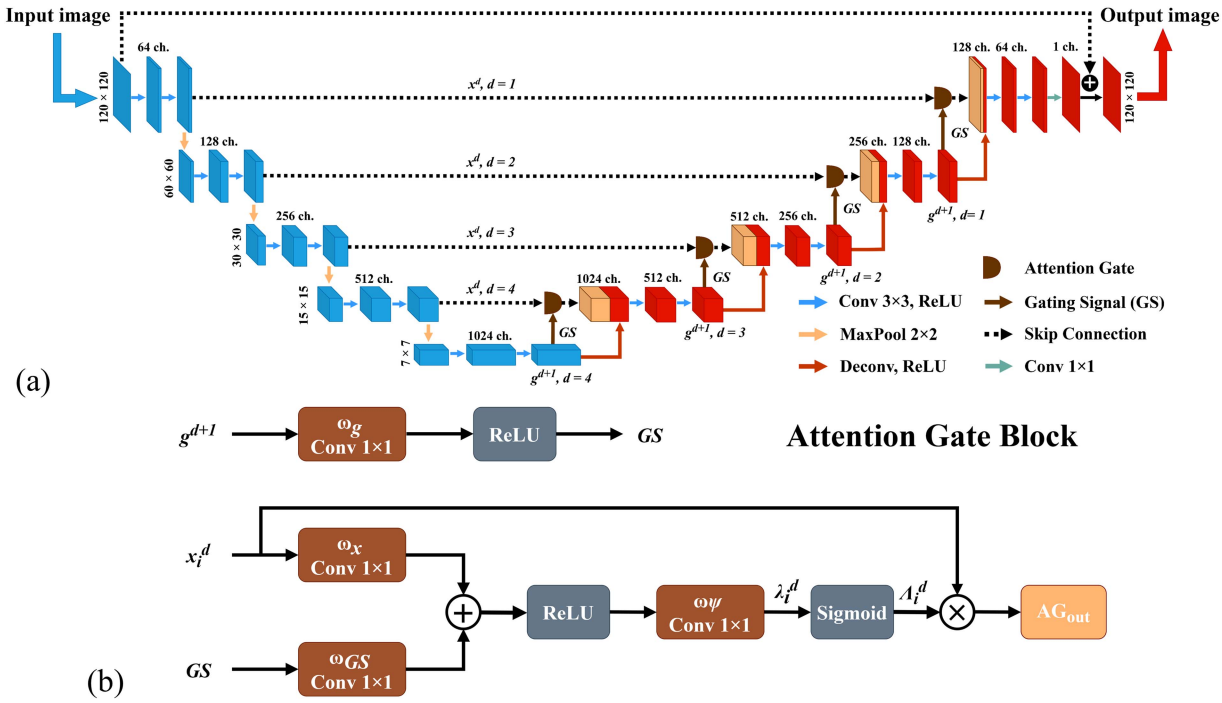

Fig. 2. (a) Structure of the proposed ResAttU-Net. (b) Schematic of the attention gate (AG).

ResU-Net has been successfully applied in imaging [50], [51], and image segmentation [52].

For conventional U-Net, feature maps extracted at multiple layers of encoder and decoder sections are directly concatenated through skip connections, which aims to mitigate the loss in information and resolution caused by the down-sampling operations in the encoder section. However, this leads to excessive and redundant information of similar features repeatedly extracted. To address this issue, attention gates (AG) have been proposed to improve the performance by learning to focus on targeted region automatically and suppress feature extraction in irrelevant regions [53]. Attention-based methods have been implemented successfully in semantic segmentation and segmentation of medical images [54], [55].

In this work, we integrate AGs in ResU-Net to form a new network structure named ResAttU-Net. AGs are used right before the concatenation operation to filter and highlight the salient features propagating through the skip connections as shown in Fig. 2(a). To be specific, each AG block consists of an attention gating signal (GS) taken from the feature map g^{d+1} in the deeper decoding path. Since g^{d+1} contains information of the finest features in the layer of depth d compared to the output of the upper encoder branches, it can help the AG block to focus relevant features on a global scale. Up-sampling and padding of g^{d+1} is performed to make its spatial dimension the same as the other input feature x_i^d that taken from the upper encoder branch, where i denotes spatial dimension. Then g^{d+1} passes through a convolution operation (ω_g) and rectified linear unit (ReLU) function for spatial feature extraction. The resulting signal is the GS as presented in (3) with b_g as the bias.

$$\mathrm{GS} = \mathrm{ReLU}\left(\omega_{\mathrm{g}}\left(g^{\mathrm{d}+1}\right) + b_g\right) \tag{3}$$

Both GS and x_i^d pass through a series of convolution layers and ReLU activation function in the AG block to generate the pre-attention coefficient λ_i^d, which is given as

$$\lambda_i^d = \omega_\psi\left(\mathrm{ReLU}\left(\omega_{GS}\left(GS\right) + \omega_x\left(x_i^d\right) + b_\lambda\right) + b_\psi\right) \tag{4}$$

The convolution layer ψ is used to reduce the dimensionality for later weighting multiplication. An attention coefficient Λ_i^d (within the range of [0, 1]) is obtained after being filtered by the Sigmoid activation function defined in (5), which is chosen to normalize the attention coefficient so as to prevent the gradient exploding problem from training.

$$\Lambda_i^d = \mathrm{Sigmoid}\left(\lambda_i^d\right) \tag{5}$$

Λ_i^d is then multiplied by the original input feature x_i^d to weight the final output of the AG block that is formulated as

$$AG_{\mathrm{out}} = \hat{x}_i^d = \Lambda_i^d \times x_i^d = \Lambda_i^d \cdot f\left(x_i^{d-1}, \Phi^{d-1}\right) \tag{6}$$

where f is the function characterized by trainable convolutional parameters Φ^d that applied in convolution layer of depth d to generate the feature map x_i^d. The attention coefficient Λ_i^d can suppress the irrelevant features and identify the salient features of the task-oriented data, such as the location of the hemorrhage spot in this work. This can help to improve the overall performance of the network.

Finally, the output of the AG block is concatenated with up-sampled feature maps in the decoder path. Benefited from the AGs, convolutional parameters in shallower layers can be updated largely based on the relevant region corresponding to the given task. The attention coefficients affect the gradient term during both the forward and backward passes. The update mechanism for convolutional parameters in the layer of depth

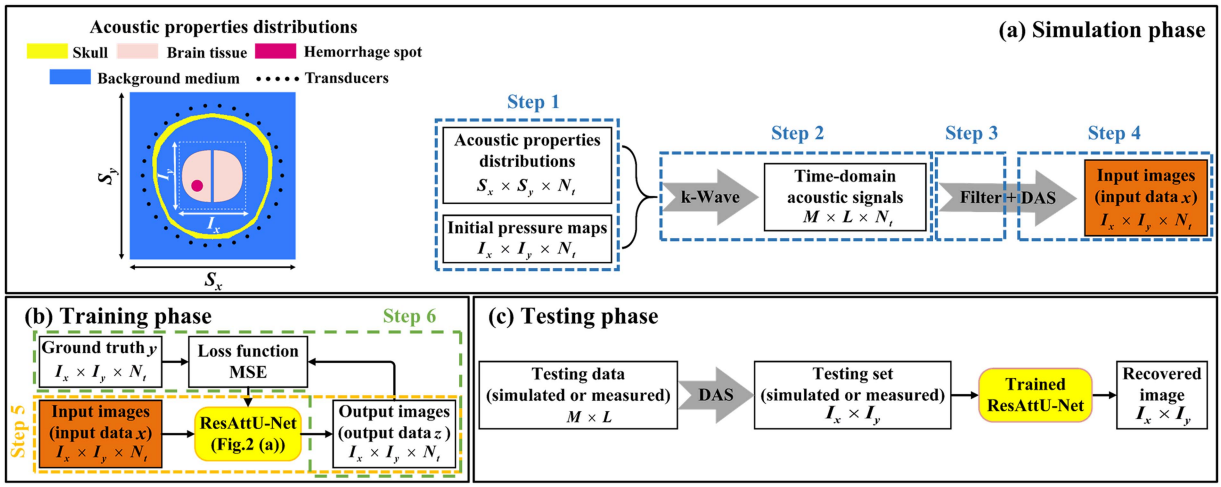

Fig. 3. Framework of the entire process of the proposed ResAttU-Net-based DL-MITAT technique for transcranial brain hemorrhage detection.

$d-1$ can be derived as

$$\frac{\partial\left(\hat{x}_i^d\right)}{\partial\left(\Phi^{d-1}\right)}=\frac{\partial\left(\Lambda_i^d\cdot f\left(x_i^{d-1},\Phi^{d-1}\right)\right)}{\partial\left(\Phi^{d-1}\right)}=\Lambda_i^d\frac{\partial\left(f\left(x_i^{d-1},\Phi^{d-1}\right)\right)}{\partial\left(\Phi^{d-1}\right)}+\frac{\partial\left(\Lambda_i^d\right)}{\partial\left(\Phi^{d-1}\right)}x_i^d. \tag{7}$$

C. Framework of DL-MITAT Method

Transcranial detection of brain hemorrhage in 2-D environment is investigated in this work as a proof-of-concept demonstration of the proposed DL-MITAT protocol. As can be seen from Sections III and IV, simplifying the problems to 2-D scenarios does not compromise the major innovation of this work, i.e., using the DL technique to address the acoustic inhomogeneity issue incurred by the skull. Accordingly, all the involved simulations and experiments are performed in 2-D conditions. We apply acoustic numerical simulations carried out by *k-Wave* MATLAB Toolbox [56], to generate the training sets including the ground truth and input data.

Fig. 3 provides the framework of the DL-MITAT technique for transcranial brain hemorrhage detection, which consists of training phase and testing phase. Detailed procedure of the training phase is described as follows.

Step 1: We first need to configure the *k-Wave* acoustic simulation. The entire simulation region has a size of $S_x \times S_y$, as shown in Fig. 3(a). Three acoustic properties are needed in the simulations, including SOS, acoustic attenuation and mass density, each of which is set to be a 2-D distribution map with a size of $S_x \times S_y$. An initial 2-D pressure map with a size of $I_x \times I_y$ serves as the source for acoustic waves, as shown in Fig. 3(a). We deploy in total M transducers on a circle enclosing the sample within the $I_x \times I_y$ region. We construct in total N_t distinct 2-D samples by varying some parameters of the samples to do N_t sets of simulations. Specific simulation setups for different investigated cases are detailed in Sections III and IV.

Step 2: We then conduct the N_t sets of *k-Wave* simulation. The final simulation results are time-domain thermoacoustic signals (having a dimension of $M \times L \times N_t$, with L denoting the number of time points) captured by the transducers.

Step 3: We filter the simulated acoustic signal raw data and add white Gaussian noises to it to improve the resemblance between the simulated and experimental signals.

Step 4: We employ the conventional DAS algorithm to reconstruct an image of the sample based on a homogeneous lossless acoustic environment, which is treated as the input image. The size of the input image is set to be the same as the initial pressure map $I_x \times I_y$, which is also named as the imaging region. It is expected that the input image is unavoidably marred with lots of artifacts.

Step 5: The input image serves as the input data for the network to achieve an output image (with a size of $I_x \times I_y$), as shown in Fig. 3(b).

Step 6: In previously reported DL-MITAT approach [57], the initial pressure map is also taken as the ground truth image or artifact-free image. For the current work aiming at the detection of brain hemorrhage, we define an image only bearing the hemorrhage spot as the ground truth image as depicted in Fig. 3(b). By this manner, reliability of the detection can be enhanced. The input image and ground truth image together form the training set. Finally, we train the network by iteratively minimizing the loss function given in (8).

$$MSE=\frac{1}{B}\sum_{i=1}^{B}\left(y_i-z_i\right)^2 \tag{8}$$

Here, y_i and z_i respectively denote the ground truth image and output image, $B = 40$ denotes the batch size for each training. Meanwhile, the training process continuously updates the adapted parameters in the network. Table I gives the setup of the training [58]. We use Python to implement the training process, which is based on the PyTorch framework running on a PC with a Tesla V100 32GB GPU. Sufficient training of the ResAttU-Net network eventually renders an output image with high resemblance to the ground truth image. Then we name the final version as a trained ResAttU-Net network and its associated output as a recovered image.

TABLE I
SETUP OF NETWORK TRAINING

Parameter	Value/Type
Batch size	40
Epoch	50
Initial Learning Rate	1e−4
Optimizer	Adam
Learning Rate Scheduler	StepLR

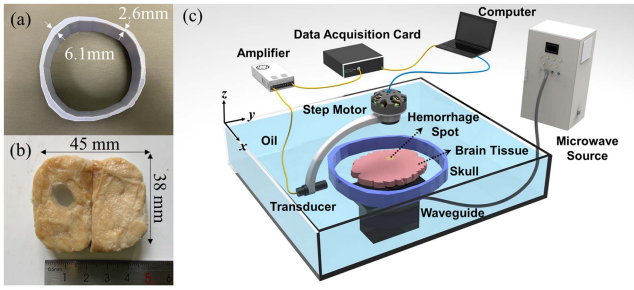

Fig. 4. Experimental setup of the *ex-vivo* testing using a 3D-printed skull. (a) Photo of the printed skull. (b) Photo of the porcine brain sample with a 10-mm-diameter hemorrhage spot. (c) Schematic of the experimental system.

In the testing phase in Fig. 3(c), the testing data, i.e., the time-resolved thermoacoustic pressure signals with a dimension of $M \times L$, can be established by either simulation or experiment. We then use the testing data to generate a testing set image (with a size of $I_x \times I_y$) based on the DAS method. By feeding this image to the trained network, we get the final output image (with a size of $I_x \times I_y$) and name it recovered image, in which the artifacts are efficiently eliminated.

III. Ex-Vivo Testing Using a 3D-Printed Skull

A. Sample Preparation and Experimental Setup

To verify the proposed ResAttU-Net-based DL-MITAT method for mitigating the adverse effects caused by human skull in transcranial brain hemorrhage detection, we first use a 3D-printed ring as the skull to conduct *ex-vivo* experimental testing. As shown in Fig. 4(a), its geometry is segmented from a 3D numerical model of a realistic adult human skull and scaled down by a factor of 2. Its outer diameter is from 105 to 113 mm and the thickness ranges from 2.6 to 6.1 mm, which is comparable to the thickness of a child's skull. Its height is 30 mm. The 3D printing material has an SOS of 2500 m/s, which falls into the range of the SOS of human skulls. Because the printing material is acoustically lossless, it cannot mimic the severe acoustic attenuation in human skulls. But the study based on such a lossless skull is still meaningful since it can show the performance of the DL-MITAT technique when only SOS inhomogeneity exists, which is the case for many application scenarios. Furthermore, comparison of the results of this section and Section IV employing a real bovine skull is also very informative.

We apply porcine brain shown in Fig. 4(b) as the normal brain samples. The two pieces of porcine brain samples are put adjacent to each other to mimic the two cerebral hemispheres of a human brain. The dielectric constant ε_r and conductivity σ of the porcine brain tissues are measured to be 50.2 and 1.62 S/m, respectively, at 2.45 GHz using a dielectric probe (N1501, Keysight). We then insert a cylindrical-shaped phantom into the porcine brain sample to mimic a brain hemorrhage spot, which is made of 59.9% water, 29.9% glycerin and 10.2% gelatin by weight. We fabricate two samples to do the *ex-vivo* testing, with one having a 10-mm-diameter hemorrhage spot and the other having a 5-mm-diameter hemorrhage spot.

TABLE II
ACOUSTIC PARAMETERS OF BRAIN AND SKULL

Material	Density (kg/m^3)	Speed of sound (m/s)	Attenuation coefficient (dB/MHz/cm)
Brain tissue	1046	1540	0.59
3D-printed skull	1180	2500	0
Bovine skull	1920-2060	1800-3200	18-33
Oil	900	1470	0

TABLE III
DIELECTRIC PARAMETERS OF BRAIN TISSUE AND PHANTOM

Material	Simulated contrast	Permittivity	Conductivity (S/m)
Brain tissue	-	50.2	1.62
High-contrast hemorrhage phantom	+49%	55.2	2.42
Low-contrast hemorrhage phantom	+28%	59.3	2.01
Ischemia phantom	−37%	78.4	1.05

Permittivity and conductivity of the used hemorrhage phantom material are given in Table III. Numerical simulations using CST software package reveals that the absorbed microwave power (defined as $\sigma|\boldsymbol{E}|^2$ with $\boldsymbol{E}$ denoting the vector electric field) in the hemorrhage spot is about 1.49 times higher than that in the normal brain tissue, which means the contrast is 49%. The contrast of experimental samples used in this work specifically refers to the contrast in microwave power absorption. Although both permittivity and conductivity can affect the contrast in microwave absorption, conductivity plays a more important role. To be more specific, the contrast is roughly proportional to conductivity. As reported in [59], [60], the microwave contrast of brain hemorrhage against normal brain tissues can range from 20% to 50%. The higher the blood contained in the hemorrhage spot, the higher the contrast. Two contrast 49% and 28% (in Section IV-A) are used in this work, which fall in the reported range and can well test the robustness of the proposed technique in dealing with different contrast.

Schematic of the experimental system is given in Fig. 4(c). We put the brain sample in the printed skull and immerse both of them in a tank filled with coupling oil with an SOS of 1470 m/s. We use a waveguide (WR430) antenna resided beneath the tank to radiate microwave signals to the sample. The antenna is fed through a coaxial cable by a microwave source pumping 2.45-GHz 0.5-μs-wide pulses with a duty cycle of 0.05% and

peak power of 20 kW. We adopt a 2.25-MHz single-element flat ultrasound transducer configured in the way shown in Fig. 4(c) to detect the generated acoustic signals, which is controlled to do circular scanning around the skull at 180 discrete locations with a scanning radius of 63 mm. The acoustic signals are sequentially amplified by 59 dB, recorded by a data acquisition card with a sampling rate of 15 MHz (PXI-5922, National Instruments), and processed by a personal computer.

B. Construction of Training Sets

We perform two groups of simulations via *k-Wave* toolbox, each of which establishes 1500 training sets and 100 for validation. The two groups respectively correspond to the two fabricated samples with a 10-mm-diameter hemorrhage spot and a 5-mm-diameter hemorrhage spot. Dimensions of the entire *k-Wave* acoustic simulation region and imaging region are respectively set to be 300 × 300 pixels ($S_x = S_y = 300$) and 120 × 120 pixels ($I_x = I_y = 120$) with a pixel size of 0.5 mm × 0.5 mm. The initial pressure in the hemorrhage spot is set to be 1.5 times higher than that in the brain tissue, i.e., 50% contrast. The applied SOS and mass density in different materials are listed in Table II [61]. The entire region is set to be acoustically lossless. 180 transducers are located on a 63-mm-radius circle around the sample to record the simulated acoustic signals.

We vary some parameters to enrich the training sets. To be specific, the thickness of the skull in each training set is nonuniform to account for the nature of typical human skulls, which is changed from 2 to 10.5 mm in a random manner. Size of the brain sample is varied from 900 to 2025 mm^2 and rotate the entire sample, both in a random manner. Moreover, size of the hemorrhage spot is adjusted in the range of 7 to 11 mm for the first group of 1500 training sets and 3 to 7 mm for the second group of 1500 training sets, respectively. We use two irregular convex polygons, with one totally contained within the other, to represent the inner and outer boundaries of the skull. We randomly adjust the two polygons to vary the shape of the skull. These operations greatly enhance the richness of the datasets. We then add white Gaussian noises to the simulated raw signals with a of 20 dB SNR. We finally use the DAS algorithm to yield an input image.

In the ground truth image, pixels in the hemorrhage spot are set to 1 and the rest pixels in the surrounding area are set to 0. Although the SOS distribution is not included in the ground truth, it affects the input images to a great extent.

C. Ex-Vivo Testing Results

We perform two sets of *ex-vivo* experiments using the two fabricated samples to get the experimental testing data. For the first sample, the DAS image in Fig. 5(a) only partially reveals a blurred boundary of the 10-mm-diameter hemorrhage spot. The recovered image from the trained network is given in Fig. 5(b). The image contrast is noticeably enhanced and the artifacts are significantly suppressed, making the hemorrhage spot clearly distinguishable. Its boundary, size, and shape are all reconstructed with high fidelity. For the second sample, the 5-mm-diameter hemorrhage spot is barely seen in the DAS result in Fig. 5(c), but is reconstructed with high quality in the recovered image in Fig. 5(d). Sizes of the images in Sections III and IV are 60 mm × 60 mm unless otherwise specified.

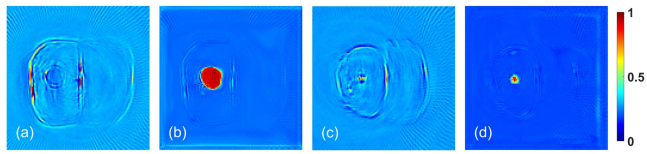

Fig. 5. Printed-skull-based *ex-vivo* experimental testing results. (a) and (b) are respectively the DAS and recovered images for the sample with a 10-mm-diameter hemorrhage spot. (c) and (d) are respectively the DAS and recovered images for the sample with a 5-mm-diameter hemorrhage spot.

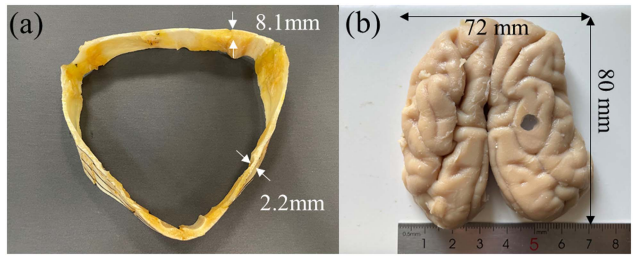

Fig. 6. (a) Photo of the homemade bovine skull ring. (b) Photo of the porcine brain sample with a 5-mm-diameter hemorrhage spot.

IV. Ex-Vivo Testing Using a Bovine Skull

A. Sample Preparation

Realistic human skulls bear high acoustic attenuation [62], [63], [64], which also complicates the acoustic signal propagation. In order to make the proposed ResAttU-Net-based DL-MITAT framework more convincing, we further perform *ex-vivo* experimental testing exercising a buffalo bovine skull. As displayed in Fig. 6(a), we cut three 30-mm-wide arc-shaped pieces from the top of the bovine skull and connect them end to end to form a ring using hot-melt adhesive. The thickness of the bovine skull ring is ranged from 2.2 to 8.1 mm and the largest outer diameter is about 155 mm.

The hemorrhage-mimicking phantom used here is made of 73% water, 14.6% glycerin and 12.4% gelatin [65]. This leads to a contrast of 28% that is around the lower limit of the reported contrast range of brain hemorrhage [59]. Because a target with a low contrast is generally hard to be differentiated by MITAT [14], [66] and ultrasonography [67], it is very meaningful to investigate such a challenging case. This is also helpful to evaluate the capability of the DL-MITAT technique in dealing with different practical scenarios. For the bovine skull based *ex-vivo* experiments, we prepare two samples, with one containing a 10-mm hemorrhage spot and the other only having normal brain tissues. The scanning radius is increased to 85 mm. Other experimental setups used in this section are the same as those described in Section III-A.

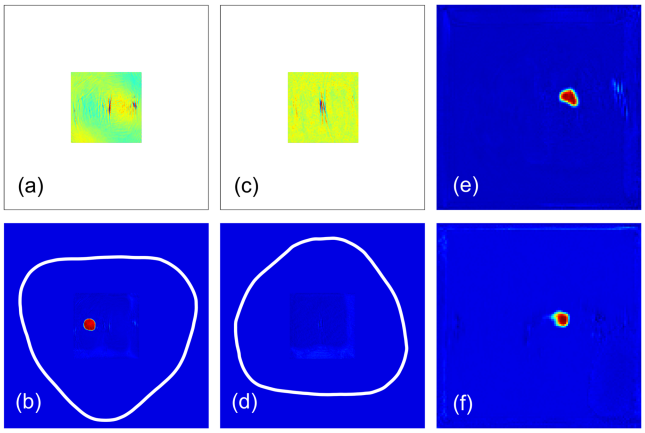

Fig. 7. Bovine-skull-based *ex-vivo* experimental testing results. (a) and (b) are respectively the DAS and recovered images for the 10-mm-hemorrhage sample. (c) and (d) are respectively the DAS and recovered images for a normal brain sample. (e) and (f) are respectively the recovered images for the 5-mm- hemorrhage and 3-mm-hemorrhage sample. Sizes of (a)–(d) are 175 mm × 175 mm. Sizes of (e) and (f) are 60 mm × 60 mm.

B. Construction of Training Sets

For this more complicated case, we establish in total 4000 training sets together with 200 sets for validation. The acoustic simulation region and imaging region are respectively defined as 350 × 350 pixels ($S_x = S_y = 350$) and 120 × 120 pixels ($I_x = I_y = 120$) with a pixel size of 0.5 mm × 0.5 mm. The initial pressure in the hemorrhage spot is set to be 1.2 times higher than that in the brain tissue, i.e., 20% contrast. Configurations of the acoustic properties are identical to those in Section III-B unless otherwise specified.

We take several measures to guarantee the data richness of the training sets. First, we vary the area of brain tissue from 1600 to 7225 mm^2, the thickness of the skull from 2 mm to 8.5 mm, and the location of the hemorrhage spot in the brain tissue (details can be found in the supplementary material), which are all done in a random manner. Second, we sweep the diameter of the hemorrhage spot from 7 to 11 mm. Third, in light of the fact that the acoustic properties of human skulls have big variation ranges, it is imperative to vary them in a reasonable range in the training sets, as summarized in Table II [62], [63], [64]. These data enriching measures can considerably increase the practicability of the DL-MITAT mechanism for diversified realistic scenarios.

After the simulations are done, we add Gaussian noises with a SNR of 15 dB to the simulated raw signals. We then acquire an input image via the conventional DAS method and initiate the training process of the network.

C. Detection of Samples With and Without a Hemorrhage Spot

We conduct two *ex-vivo* experiments respectively making use of the bovine skull with the two samples. For the first sample, its recovered image in Fig. 7(b) unambiguously reveals the 10-mm hemorrhage spot with very high fidelity in terms of shape, location and dimension. For the second sample, its recovered image in Fig. 7(d) does not show any white spot indicative of the occurrence of hemorrhage, which can be considered as a normal case, i.e., without a hemorrhage spot. It is worth mentioning that the training sets for this normal case are identical to those Section IV-B. In other words, even if no normal case is included in the training sets, the sample without a hemorrhage spot can still be faithfully reconstructed. Thus, it is proved that the proposed DL-MITAT methodology is able to reliably distinguish hemorrhage from normal cases. This is crucial for reducing false positive rate in potential clinical applications. On the contrary, the DAS images shown in Fig. 7(a) and (c) exhibit lots of streak-shaped artifacts and fail to discriminate between the hemorrhage and normal cases.

For each of these two cases, we implement one more ultrasound pulse echo experiment to recover the outer boundary of the bovine skull, which are displayed as the white irregular-shaped rings in Fig. 7(b) and (d). By this manner, we can accurately locate the detected hemorrhage spot inside the brain, which is of great practical value for the diagnosis and treatment of brain hemorrhage.

D. Detection of Smaller Hemorrhage Spots

To gain further insight of the proposed DL-MITAT approach, we perform systematic studies of the capability and limitations of the ResAttU-Net starting from this subsection.

We first fabricate and examine two new samples with smaller hemorrhage spots, i.e., 5-mm and 3-mm diameter. We rebuild 4000 training sets together with 200 sets for validation, in which the diameter of the hemorrhage spot is swept from 3 to 7 mm. The image given in Fig. 7(e) exhibits a well recovered hemorrhage spot with a size of 5 mm.

We find that the 3-mm hemorrhage spot with a 28% contrast cannot be handled by the trained network. This is probably because a small hemorrhage spot with low contrast generates low-SNR acoustic signals (about 4 dB lower than that of a 3-mm 49%-contrast hemorrhage spot). The network is trained using input images obtained by acoustic signals having good SNR. If the measured SNR is not high enough, the input image of the testing phase can be very noisy and the trained network may fail to generate a good recovered image. Therefore, we make another sample containing a 3-mm hemorrhage spot bearing an enhanced contrast of 49%, which is the same as that used in Section III-A. The imaging result showcased in Fig. 7(f) noticeably manifests the hemorrhage spot at correct location with an accurate size. So, diagnosing a hemorrhage spot of 3 mm with a relatively high contrast is within the capability of the proposed technique. This finding is very meaningful since the smallest detectable hemorrhage spot by the commonly exercised clinical approaches is about 3 mm [68].

E. Comparison of Different Networks

We then compare the performance of the ResAttU-Net with another three networks, including original U-Net, ResU-Net, and AttU-Net without residual skip connection. As suggested by Fig. 8, the designed ResAttU-Net outperforms the other networks in terms of image quality for all the four experimental cases. The quantitative comparison based on NRMSE, defined in (9), labeled in the images also supports

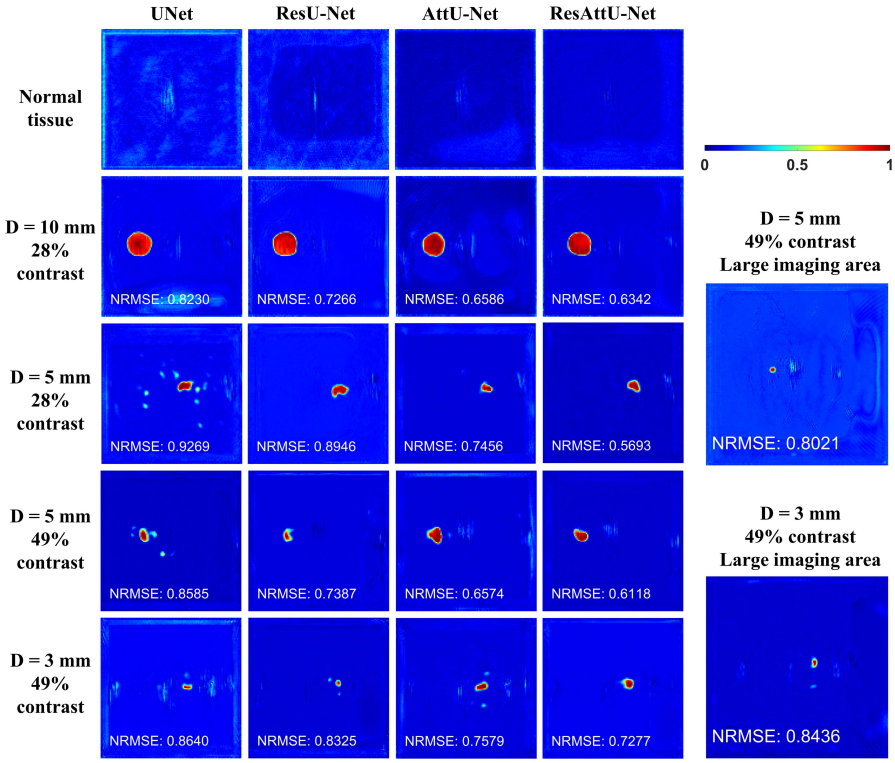

Fig. 8. Bovine-skull-based *ex-vivo* experimental testing results comparing the performance of different networks using different samples.

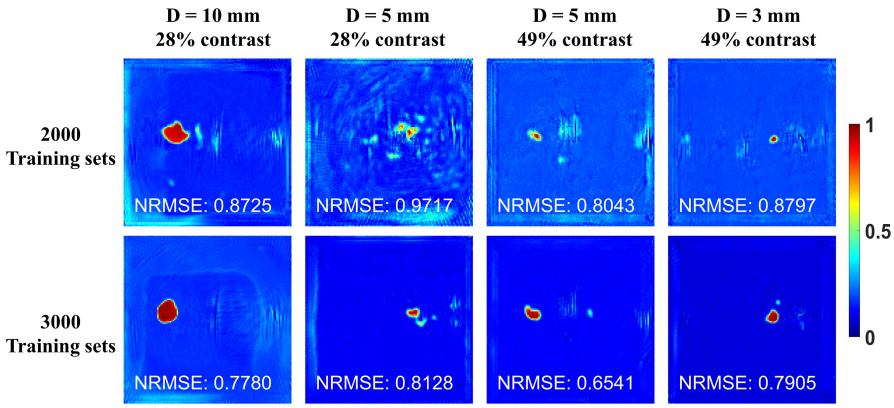

Fig. 9. Bovine-skull-based *ex-vivo* experimental testing results studying the effect of number of training sets using different samples.

this conclusion. In addition, the AttU-Net compares favorably to the ResU-Net as expected according to the description in Section II-B.

$$\mathrm{NRMSE} = \sqrt{\frac{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\left|Y\left[m,n\right]-X\left[m,n\right]\right|^2}{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\left|X\left[m,n\right]\right|^2}} \tag{9}$$

In (9), X and Y are respectively the ground truth image and recovered image.

F. Number of Training Sets

Next, we investigate the effect of number of training sets on the imaging quality. By maintaining the network structure, we reduce the number of training sets to 3000 and 2000. The recovered images of four tested cases are presented in Fig. 9 with the NRMSE results labeled. It is straightforward to see that the image quality becomes worse as the number of training sets and available information decreases. Moreover, the 5-mm 28%-contrast 3000-training-sets case results in almost the same NRMSE as the 5-mm 49%-contrast 2000-training-sets case. This fact implies that a higher contrast can efficiently reduce the demanded number of training sets to render a recovered image of the same quality. Furthermore, comparison with the results in Section III reveals that a more complicated scenario needs more training sets to train the network and achieve acceptable performance.

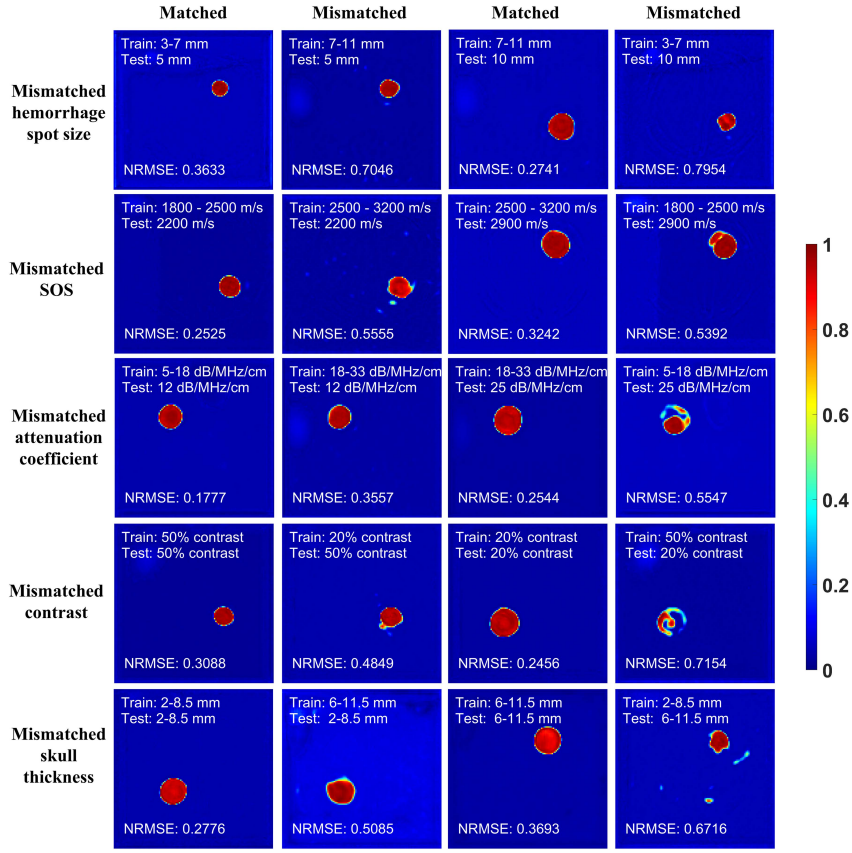

Fig. 10. Bovine-skull-based simulation testing results for mismatched conditions.

G. Mismatch Between Training Sets and Testing Data

Finally, we investigate some conditions using testing data that does not fall into the range of the training sets, which are named as mismatched cases [69]. We study four kinds of mismatches. Since some testing data involves samples that are difficult to realize in experiments, we apply the simulation method for this study. In total 200 testing sets are built for each case and the NRMSE is calculated for the corresponding 200 recovered images. For each case, one of the 200 recovered images is given in Fig. 10 as an example.

The first case has mismatch in the size of the hemorrhage spot. We try two kinds of mismatches for this case. The first kind utilizes 7-11 mm spots in the training sets and 3-7 mm spots in the testing sets. The second kind is featured by 3-7 mm spots in the training sets and 7-11 mm spots in the testing sets. Fig. 10 shows that the first kind recovers a hemorrhage spot bigger than its actual size, while the second kind recovers a hemorrhage spot smaller than its actual size.

The second and third cases respectively pertain to mismatched SOS and attenuation coefficient of the skull. Similarly, we test two kinds of mismatches. In the first kind, the acoustic parameter in the testing sets is lower than that in the training sets. In the second kind, the acoustic parameter in the testing sets is higher than that in the training sets. The quality of the recovered images for these mismatched cases in Fig. 10 is obviously degraded compared to the matched counterparts. The results of the first three kinds of mismatches undoubtedly indicate that performance of the applied network is unfavorable in the presence of mismatched parameters in the training and testing sets. This is a general limitation for supervised learning.

The fourth case evaluates the effect of mismatched contrast by two different settings. The first uses 20% contrast for training while 50% contrast for testing. The resulting recovered image is a little worse than the matched case that employs 50% contrast in both data sets. The second defines 50% contrast in training while 20% contrast in testing. The image is much worse than the matched case that utilizes 20% contrast in both data sets. This finding tells that if the contrast in the testing set is higher than that in the training sets, an acceptable recovered image can still be secured. However, if the contrast in the testing set is lower than that in the training sets, the image quality is severely reduced. This is probably because of the lower SNR in the testing set than that in the training sets. This finding is also the reason why we fabricate samples with a contrast a little higher than that in the training sets.

H. Detection of an Irregular-Shaped Hemorrhage Spot

Beside the circular hemorrhage spot, we also test an irregular-shaped hemorrhage spot (49% contrast) as shown in Fig. 11, which is about 15 mm long and 7 mm wide. The training sets are the same as the first mismatched contrast case in Fig. 10 using 50% contrast, which use rounded hemorrhage spots. The recovered image given in Fig. 11 show reliably reconstructed

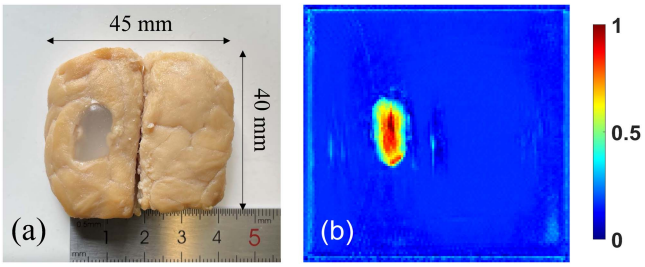

Fig. 11. Bovine-skull-based *ex-vivo* experimental testing result for an irregular-shaped hemorrhage phantom. (a) Photo of the sample. (b) Recovered image.

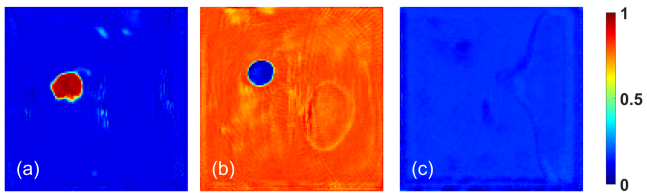

Fig. 12. Bovine-skull-based *ex-vivo* experimental testing results to distinguish among (a) Hemorrhagic stroke, (b) Ischemic stroke and (c) Normal case.

shape and dimension of the irregular-shaped hemorrhage spot, demonstrating the good generalization of the proposed DL-MITAT approach.

I. Distinguish Between Hemorrhagic and Ischemic Stroke

For potential clinical applications, it is meaningful to test if the proposed method can distinguish between hemorrhagic and ischemic stroke [70].We make a new sample by embedding a 10-mm ischemia-mimicking phantom into porcine brain tissue. The contrast of the ischemia spot is -37% (within the reported range of ischemic stroke [60]), which means the microwave absorption in the spot is 37% lower than that in the porcine brain tissue. We rebuild 4000 training sets for this subsection. The first 2000 sets are for hemorrhagic stroke that follow the same setup as those in Section IV-H. The second 2000 sets are for ischemic stroke and adopt the same parameter variation ranges as the first 2000 sets for hemorrhagic stroke except using a -50% contrast. In the ground truth for the ischemic stroke, the ischemia spot is set to 0 and the background region is set to 1. We then use three testing sets to test this trained network, including a hemorrhagic sample with 49% contrast, an ischemic sample with -37% contrast, and a normal sample. The recovered images presented in Fig. 12 exhibit very good quality and preliminarily prove that the proposed technique is able to classify hemorrhagic stroke, ischemic stroke and normal case. Many other aspects of the transcranial ischemia detection need to be further researched, which is our ongoing work.

J. Discussions

Although we only demonstrate 2-D conditions in this work, the proposed ResAttU-Net-based DL-MITAT methodology is applicable to 3-D scenarios, which is simply because the basic procedure is not limited by the dimensions of applications. The training sets can also be established by 3-D *k-Wave* simulations. It is worth mentioning that there are two possible challenges for the potential 3-D applications. The first one is the greatly increased computational burden for obtaining the training sets and training the network as compared to the 2-D case. The second one is that a deeper embedded hemorrhage spot demands higher input microwave energy to maintain the SNR of thermoacoustic signals and quality of the recovered image. This may increase the system cost and cause over heating in the scalp. A newly reported focused microwave mechanism may be used to reduce the demanded microwave power to get a desired SNR for a deeply embedded hemorrhage spot [71], [72]. The proposed method also holds the potential to be realized by a handheld system that is more suitable for clinical usage [73].

For the applied ResAttU-Net that is a kind of supervised learning method, plenty of training sets are demanded to train the network. Although experiments are usually conducted to get training sets for DL-based photoacoustic imaging [41], the adopted simulation method in this work bears two distinct advantages. First, in view of the large amount of required training sets, the simulation method is considerably more time- and cost-efficient than the experimental counterpart. Second, the experimental method is infeasible for some scenarios since desired samples exhibiting accurate microwave and acoustic properties are impossible to be realized. For example, each training set in Section IV-F uses a skull with specific SOS, density and attenuation coefficient. However, it is impossible to obtain so many different materials with exact predefined acoustic parameters to do experiments. The simulation approach used in this work is evidenced to be efficient and robust for DL-MITAT. This mechanism can potentially be used in a variety of DL-based applications.

Because the major purpose of this work is to detect brain hemorrhage, the only target of interest is the hemorrhage spot. This means recovering the background normal brain tissues in the image is of less importance for the current task. For locating the hemorrhage spot inside the brain, we can reconstruct the boundary of the skull using the ultrasound pulse echo technique, which can assist clinical diagnosis and treatment of brain hemorrhage. By ignoring the normal brain tissues in the ground truth image, the network is able to mainly extract features in the region with the hemorrhage spot while largely suppress the features in the background region. This can also facilitate the calculation the attention coefficient (in (5)) in the attention gate block, which is because the attention coefficient is used to enhance the weight of the target and decrease the weight of other irrelevant features in the background region.

To save the computational cost, we use a 60 mm $\times$ 60 mm limited reconstruction region. We also try a 120 mm $\times$ 120 mm reconstruction area covering the entire region in the bovine skull and show the imaging results in Fig. 8. It is seen that the recovered hemorrhage spots still bear high quality and the artifacts in the background region does not affect the conclusion of a successful detection.

In the simulation setup of the training sets, we use 180 transducers located on a circle enclosing the entire brain sample. This means we collect time-domain acoustic signal at 180

locations to do the image processing. In the experiments, we scan a single-element transducer along a circle with a step of 2° to obtain the required 180 measurements, which is essentially equivalent to using a circular transducer array.

In the *k-Wave* simulations, the 180 transducers are set to point transducers, which has an ideal uniform receiving pattern as used by previous work [57]. Thus, the point transducer does not induce any scattering or reflection of the acoustic waves. The applied single-element transducer in the experiments is designed to have very good impedance matching performance, which means the surface of the transducer almost does not reflect acoustic waves when it is listening. Even if some acoustic waves are reflected by the transducer, which are then reflected back by the skull and finally received by the transducer, such signals can be easily filtered and will not affect the performance of the proposed method. Accordingly, there is no need to compensate for the wave scattering. Instead, we did compensation for the simulated time-domain acoustic signals. To be specific, we add 15-dB Gaussian noise to the simulated signals and filter them using a filtering function obtained from experimentally measured signals to increase the resemblance to the measured ones.

This protocol can be directly applied in scenarios with larger and thicker skulls simply by reestablishing the training sets. We perform simulations to test a 6∼11.5 mm thick skull that can mimic an adult human skull [23]. Good imaging results in Fig. 10 can still be obtained so long as the skull thickness variation range in the testing data matches with that in the training sets.

We also investigate the microwave power penetration depth in the brain and safety issues of this technique in the supplementary material, along with other topics such as irregular-shaped hemorrhage and ischemic stroke.

V. Conclusion

In this work, we present the DL-MITAT modality to address the acoustic inhomogeneity issue and demonstrate its effectiveness in transcranial brain hemorrhage detection. We propose a new network structure ResAttU-Net to implement the DL-MITAT technique. We use the simulation method to build training sets, which is much more time-efficient and feasible than the experiment-based method. We demonstrate the validity of the technique by *ex-vivo* experiments using a lossless printed skull and an 8.1-mm thick bovine skull. The obtained results provide preliminary verification of the ability of the DL-MITAT method in eliminating the adverse effects of acoustic inhomogeneity and conducting transcranial hemorrhage detection. It is also able to discriminate the hemorrhage from ischemia and normal cases and detect a hemorrhage spot as small as 3-mm large. In addition, we show that the ResAttU-Net outperforms other networks for the current application, more training sets or a higher contrast can render a better image, and mismatched conditions cannot be handled by the DL-MITAT technique. This work not only opens a compelling avenue for transcranial brain hemorrhage detection in a cost-effective, time-efficient, accurate and noninvasive manner, but also establishes a novel modality for various promising applications of MITAT including other transcranial brain imaging applications. It is also helpful for dealing with the acoustic inhomogeneity problems in ultrasound imaging modality.

Acknowledgment

The authors thank the financial support from the ShanghaiTech-United Imaging Intelligence Joint Lab.

References

[1] T. Bowen, "Radiation-induced thermoacoustic soft-tissue imaging," in *Proc. Ultrason. Symp.*, 1981, pp. 817–822.
[2] F. Caspers and J. Conway, "Measurement of power density in a lossy material by means of electro-magnetically induced acoustic signals for non-invasive determination of spatial thermal absorption in connection with pulsed hyperthermia," in *Proc. IEEE 12th Eur. Microw. Conf.*, 1982, pp. 565–568.
[3] X. Wang et al., "Microwave-induced thermoacoustic imaging model for potential breast cancer detection," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 10, pp. 2782–2791, Oct. 2012.
[4] J. C. Lin, "The microwave auditory effect," *IEEE J. Electromagn. RF Microw. Med.*, vol. 6, no. 1, pp. 16–28, Mar. 2022.
[5] Y. Xu and L. V. Wang, "Rhesus monkey brain imaging through intact skull with thermoacoustic tomography," *IEEE Trans. Ultrason. Ferroelect. Freq. Control*, vol. 53, no. 3, pp. 542–548, Mar. 2006.
[6] X. Wang et al., "Computational feasibility study of contrast-enhanced thermoacoustic imaging for breast cancer detection using realistic numerical breast phantoms," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 5, pp. 1489–1501, May 2015.
[7] X. Jin, C. H. Li, and L. V. Wang, "Effects of acoustic heterogeneities on transcranial brain imaging with microwave-induced thermoacoustic tomography," *Med. Phys.*, vol. 35, no. 7, pp. 3205–3214, Jul. 2008.
[8] Z. Chi et al., "Detection and monitoring of osteoporosis in a rat model by thermoacoustic tomography," *IEEE J. Electromagn. RF Microw. Med. Biol.*, vol. 4, no. 4, pp. 234–239, Dec. 2020.
[9] L. Huang et al., "Quantitative thermoacoustic tomography: Recovery of conductivity maps of heterogeneous media," *Appl. Phys. Lett.*, vol. 101, no. 24, Dec. 2012, Art. no. 244106.
[10] Z. Chi et al., "Thermoacoustic tomography of in vivo human finger joints," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 6, pp. 1598–1608, Jun. 2019.
[11] S. Liu et al., "Microwave induced thermoacoustic tomography based on probabilistic reconstruction," *Appl. Phys. Lett.*, vol. 112, no. 1, Jun. 2018, Art. no. 263701.
[12] S. Zhao et al., "Ultrashort-pulse-microwave excited whole-breast thermoacoustic imaging with uniform field of large size aperture antenna for tumor screening," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 2, pp. 725–733, Feb. 2022.
[13] X. Liang et alet al.., "Thermoacoustic endoscopy," *Appl. Phys. Lett.*, vol. 116, no. 1, pp. 013702–1–013702–5, Jan. 2020, Art. no. 013702.
[14] D. Zhang, B. Wang, and X. Wang, "Enhanced and modulated microwave-induced thermoacoustic imaging by ferromagnetic resonance," *Appl. Phys. Exp.*, vol. 12, no. 7, Jun. 2019, Art. no. 077001.
[15] B. Wang et al., "Three-dimensional microwave-induced thermoacoustic imaging based on compressive sensing using an analytically constructed dictionary," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 1, pp. 377–386, Jan. 2020.
[16] P. M. Meyers et al., "Current status of endovascular stroke treatment," *Circulation*, vol. 123, no. 22, pp. 2591–2601, 2011.
[17] E. S. Donkor, "Stroke in the 21st century: A snapshot of the burden, epidemiology, and quality of life," *Stroke Res. Treat.*, vol. 2018, Art. no. 3238165, doi: 10.1155/2018/3238165.
[18] S. Chen, L. Zeng, and Z. Hu, "Progressing hemorrhagic stroke: Categories, causes, mechanisms and managements," *J. Neurol.*, vol. 261, no. 11, pp. 2061–2078, 2014.
[19] C. S. Kidwell et al., "Comparison of MRI and CT for detection of acute intracerebral hemorrhage," *J. Amer. Med. Assoc.*, vol. 292, pp. 1823–1830, Oct. 2004.
[20] Z. Liu et al., "Transcranial thermoacoustic tomography: A comparison of two imaging algorithms," *IEEE Trans. Med. Imag.*, vol. 32, no. 2, pp. 289–294, Feb. 2013.
[21] S. Liu et al., "Reducing acoustic inhomogeneity based on speed of sound autofocus in microwave induced thermoacoustic tomography," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 8, pp. 2206–2214, Aug. 2020.
[22] Y. Zhao et al., "Thermoacoustic tomography of in vivo rat brain," *J. Innov. Opt. Health Sci.*, vol. 10, no. 4, Jun. 2017, Art. no. 1740001.

[23] A. Yan et al., "Microwave-induced thermoacoustic tomography throughan adult human skull," *Med. Phys.*, vol. 46, pp. 1793–1797, 2019.

[24] L. Huang, T. Li, and H. Jiang, "Thermoacoustic imaging of hemorrhagic stroke: A feasibility study with a human skull," *Med. Phys.*, vol. 44, no. 4, pp. 1494–1499, Jan. 2017.

[25] J. C. Bamber, "Acoustical characteristics of biological media," in *Encyclopedia of Acoustics*, M. J. Crocker Ed. Hoboken, NJ, USA: Wiley, 1997, pp. 1703–1726, ch. 141.

[26] F. Fry and J. Barger, "Acoustical properties of the human skull," *J. Acoust. Soc. Amer*, vol. 63, no. 5, pp. 1576–1590, May 1978.

[27] K. Shapoori et al., "An ultrasonic-adaptive beamforming method and its application for trans-skull imaging of certain types of head injuries; Part I: Transmission mode," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 5, pp. 1253–1264, May 2015.

[28] M. Xu, Y. Xu, and L. V. Wang, "Time-domain reconstruction algorithms and numerical simulations for thermoacoustic tomography in various geometries," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 9, pp. 1086–1099, Sep. 2003.

[29] M. Xu and L. V. Wang, "Pulsed-microwave-induced thermoacoustic tomography: Filtered backprojection in a circular measurement configuration," *Med. Phys.*, vol. 29, no. 8, pp. 1661–1669, 2002.

[30] C. G. Hoelen and F. F. de Mul, "Image reconstruction for photoacoustic scanning of tissue structures," *Appl. Opt.*, vol. 39, no. 31, pp. 5872–5883, Dec. 2000.

[31] B. Wang et al., "Efficient dictionary construction method for microwave induced thermoacoustic compressive sensing imaging," *Appl. Phys. Lett.*, vol. 113, 2018, Art. no. 053701.

[32] Y. Sun et al., "A low-cost compressive thermoacoustic tomography system for hot and cold foreign bodies detection," *IEEE Sensors J.*, vol. 21, no. 20, pp. 23588–23596, Oct. 2021.

[33] E. M. Lillie et al., "Evaluation of skull cortical thickness changes with age and sex from computed tomography scans," *J. Bone Mineral Res.*, vol. 32, no. 2, pp. 299–307, Feb. 2016.

[34] T. R. Wang and Y. Jing, "Transcranial ultrasound imaging with speed of sound-based phase correction: A numerical study," *Phys. Med. Biol.*, vol. 58, no. 19, pp. 6663–6681, Oct. 2013.

[35] L. Deng, A. Hughes, and K. Hynynen, "A noninvasive ultrasound resonance method for detecting skull induced phase shifts may provide a signal for adaptive focusing," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 9, pp. 2628–2637, Sep. 2020.

[36] B. Zhu et al., "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487–492, Mar. 2018.

[37] D. Lee et al., "Deep residual learning for accelerated MRI using magnitude and phase networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1985–1995, Sep. 2018.

[38] Q. Dou et al., "Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1558–1567, Jul. 2017.

[39] D. Nie et al., "Medical image synthesis with deep convolutional adversarial networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 12, pp. 2720–2730, Dec. 2018.

[40] F. Liu et al., "Deep learning MR imaging-based attenuation correction for PET/MR imaging," *Radiology*, vol. 286, no. 2, pp. 676–684, Feb. 2018.

[41] N. Davoudi, X. L. Dean-Ben, and D. Razansky, "Deep learning optoacoustic tomography with sparse data," *Nature Mach. Intell.*, vol. 10, no. 1, pp. 453–460, Sep. 2019.

[42] T. Tong et al., "Domain transform network for photoacoustic tomography from limited-view and sparsely sampled data," *Photoacoustics*, vol. 19, 2020, Art. no. 100190.

[43] X. Wang, R. S. Witte, and H. Xin, "Thermoacoustic and photoacoustic characterization of few-layer graphene by pulsed excitations," *Appl. Phys. Lett.*, vol. 108, no. 14, 2016. Art. no. 143104.

[44] K. Xu et al., "Deep learning-based inversion methods for solving inverse scattering problems with phaseless data," *IEEE Trans. Antennas Propag.*, vol. 68, no. 11, pp. 7457–7470, Nov. 2020.

[45] X. Chen et al., "A review of deep learning approaches for inverse scattering problems," *Prog. Electromagn. Res.*, vol. 167, pp. 67–81, Jun. 2020.

[46] K. Xu et al., "Fast full-wave electromagnetic inverse scattering based on scalable cascaded convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 2001611.

[47] X. Liang et al., "Investigation of artifacts by mapping SAR in thermoacoustic imaging," *J. Innov. Opt. Health Sci.*, vol. 14, no. 5, Apr. 2021, Art. no. 2150011.

[48] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput. Assist. Interv.*, 2015, pp. 234–241.

[49] K. He et al., "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[50] S. Antholzer, M. Haltmeier, and J. Schwab, "Deep learning for photoacoustic tomography from sparse data," *Inverse Problems Sci. Eng.*, vol. 27, no. 7, pp. 987–1005, Sep. 2018.

[51] H. Lee et al., "Deep-neural-network based sinogram synthesis for sparse-view CT image reconstruction," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 2, pp. 109–119, Mar. 2019.

[52] X. Xiao et al., "Weighted Res-UNet for high-quality retina vessel segmentation," in *Proc. IEEE 9th Int. Conf. Inf. Technol. Med. Educ.*, 2018, pp. 327–331.

[53] O. Oktay et al., "Attention U-Net: Learning where to look for the pancreas," in *Proc. Int. Conf. Med. Imag. Deep Learn.*, 2018, pp. 1–10.

[54] N. Abraham and N. M. Khan, "A novel focal Tversky loss function with improved attention u-net for lesion segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag.*, 2019, pp. 683–687.

[55] E. Thomas et al., "Multi-res-attention UNet: A CNN model for the segmentation of focal cortical dysplasia lesions from magnetic resonance images," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 5, pp. 1724–1734, May 2021.

[56] B. E. Treeby and B. T. Cox, "k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields," *J. Biomed. Opt.*, vol. 15, no. 2, Mar. 2010, Art. no. 021314.

[57] J. Zhang et al., "Deep-learning-enabled microwave-induced thermoacoustic tomography based on sparse data for breast cancer detection," *IEEE Trans. Antennas Propag.*, vol. 70, no. 8, pp. 6336–6348, Aug. 2022.

[58] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent*, 2015, pp. 1–15.

[59] M. Hopfer et al., "Electromagnetic tomography for detection, differentiation, and monitoring of brain stroke: A virtual data and human head phantom study," *IEEE Antennas Propag. Mag.*, vol. 59, no. 5, pp. 86–97, Oct. 2017.

[60] K. Olympia et al., "Experimental validation of microwave tomography with the DBIM-TwIST algorithm for brain stroke detection and classification," *Sensors*, vol. 20, no. 3, Jan. 2020, Art. no. 840.

[61] P. A. Hasgall et al., "IT'IS database for thermal and electromagnetic parameters of biological tissues," Feb. 2022, doi: 10.13099/VIP21000-04.

[62] H. N. Minh, J. Du, and K. Raum, "Estimation of thickness and speed of sound in cortical bone using multifocus pulse-echo ultrasound," *IEEE Trans. Ultrason. Ferroelect. Freq. Control*, vol. 67, no. 3, pp. 568–579, Mar. 2020.

[63] C. Demené et al., "Transcranial ultrafast ultrasound localization microscopy of brain vasculature in patients," *Nature Biomed. Eng.*, vol. 5, no. 3, pp. 219–228, Mar. 2021.

[64] M. Mozaffarzadeh et al., "Refraction-corrected transcranial ultrasound imaging through the human temporal window using a single probe," *IEEE Trans. Ultrason. Ferroelect. Freq. Control*, vol. 69, no. 4, pp. 1191–1203, Apr. 2022.

[65] P. M. Meaney et al., "Electrical characterization of glycerin: Water mixtures: Implications for use as a coupling medium in microwave tomography," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 5, pp. 1471–1478, May 2017.

[66] D. Wu et al., "Contrast agents for photoacoustic and thermoacoustic imaging: A review," *Int. J. Mol. Sci.*, vol. 15, no. 12, pp. 23616–23639, Dec. 2014.

[67] A. D. Bar-Zion et al., "Denoising of contrast-enhanced ultrasound cine sequences based on a multiplicative model," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 8, pp. 1969–1980, Aug. 2015.

[68] V. C. Cappendijk et al., "In vivo detection of hemorrhage in human atherosclerotic plaques with magnetic resonance imaging," *J. Magn. Reson. Imag.*, vol. 20, no. 1, pp. 105–110, Jul. 2004.

[69] Z. Luo et al., "Quantitative reconstruction of dielectric properties based on deep-learning-enabled microwave-induced thermoacoustic tomography," *IEEE Trans. Microw. Theory Techn.*, vol. 71, no. 6, pp. 2652–2663, Jun. 2023.

[70] S. Chennareddy et al., "Portable stroke detection devices: A systematic scoping review of prehospital applications," *BMC Emerg. Med.*, vol. 22, no. 111, pp. 1–13, Jun. 2022.

[71] L. Xu and X. Wang, "Focused microwave breast hyperthermia monitored by thermoacoustic imaging: A computational feasibility study applying realistic breast phantoms," *IEEE J. Electromagn. RF Microw. Med.*, vol. 4, no. 2, pp. 81–88, Jun. 2020.

[72] J. Li et al., "A preclinical system prototype for focused microwave breast hyperthermia guided by compressive thermoacoustic tomography," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 7, pp. 2289–2300, Jul. 2021.

[73] L. Wu et al., "A handheld microwave thermoacoustic imaging system with an impedance matching microwave-sono probe for breast tumor screening," *IEEE Trans. Med. Imag.*, vol. 41, no. 5, pp. 1080–1086, May 2022.