Title

Radiation-induced Acoustic Signal Denoising using a Supervised Deep Learning Framework for Imaging and Therapy Monitoring.

Journal

Physics in medicine and biology, 4(05-09)

ISSN

2331-8422

Authors

Jiang, Zhuoran
Wang, Siqi
Xu, Yifei
et al.

Publication Date

2023-10-01

DOI

10.1088/1361-6560/ad0283

IOP Publishing Phys. Med. Biol. 68 (2023) 235010 https://doi.org/10.1088/1361-6560/ad0283

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

RECEIVED
31 May 2023

REVISED
29 September 2023

ACCEPTED FOR PUBLICATION
11 October 2023

PUBLISHED
29 November 2023

PAPER

Radiation-induced acoustic signal denoising using a supervised deep learning framework for imaging and therapy monitoring

Zhuoran Jiang[1,2,8], **Siqi Wang**[3,8], **Yifei Xu**[3], **Leshan Sun**[3], **Gilberto Gonzalez**[4], **Yong Chen**[4], **Q Jackie Wu**[1,2], **Liangzhong Xiang**[3,5,6,*] **and Lei Ren**[7,*]

[1] Medical Physics Graduate Program, Duke University, Durham, NC 27705, United States of America
[2] Department of Radiation Oncology, Duke University Medical Center, Durham, NC 27710, United States of America
[3] Department of Biomedical Engineering, University of California, Irvine, CA 92617, United States of America
[4] Department of Radiation Oncology, University of Oklahoma Health Sciences Center, Oklahoma City, OK, 73104, United States of America
[5] Department of Radiological Sciences, University of California, Irvine, CA 92697, United States of America
[6] Beckman Laser Institute & Medical Clinic, University of California, Irvine, CA 92612, United States of America
[7] Department of Radiation Oncology, University of Maryland, Baltimore, MD 21201, United States of America
[8] Contributed equally.
[*] Authors to whom any correspondence should be addressed.

E-mail: lren@som.umaryland.edu **and** liangzhx@hs.uci.edu

Keywords: radiation-induced acoustic signal denoising, x-ray-induced acoustic, protoacoustic, electroacoustic, deep learning

Abstract

Radiation-induced acoustic (RA) imaging is a promising technique for visualizing the invisible radiation energy deposition in tissues, enabling new imaging modalities and real-time therapy monitoring. However, RA imaging signal often suffers from poor signal-to-noise ratios (SNRs), thus requiring measuring hundreds or even thousands of frames for averaging to achieve satisfactory quality. This repetitive measurement increases ionizing radiation dose and degrades the temporal resolution of RA imaging, limiting its clinical utility. In this study, we developed a general deep inception convolutional neural network (GDI-CNN) to denoise RA signals to substantially reduce the number of frames needed for averaging. The network employs convolutions with multiple dilations in each inception block, allowing it to encode and decode signal features with varying temporal characteristics. This design generalizes GDI-CNN to denoise acoustic signals resulting from different radiation sources. The performance of the proposed method was evaluated using experimental data of x-ray-induced acoustic, protoacoustic, and electroacoustic signals both qualitatively and quantitatively. Results demonstrated the effectiveness of GDI-CNN: it achieved x-ray-induced acoustic image quality comparable to 750-frame-averaged results using only 10-frame-averaged measurements, reducing the imaging dose of x-ray-acoustic computed tomography (XACT) by 98.7%; it realized proton range accuracy parallel to 1500-frame-averaged results using only 20-frame-averaged measurements, improving the range verification frequency in proton therapy from 0.5 to 37.5 Hz; it reached electroacoustic image quality comparable to 750-frame-averaged results using only a single frame signal, increasing the electric field monitoring frequency from 1 fps to 1k fps. Compared to lowpass filter-based denoising, the proposed method demonstrated considerably lower mean-squared-errors, higher peak-SNR, and higher structural similarities with respect to the corresponding high-frame-averaged measurements. The proposed deep learning-based denoising framework is a generalized method for few-frame-averaged acoustic signal denoising, which significantly improves the RA imaging's clinical utilities for low-dose imaging and real-time therapy monitoring.

1. Introduction

Radiation deposits energy when it travels through human bodies and interacts with tissue atoms through various mechanisms, which enables numerous medical applications. Radiation has been widely used for

diagnostic imaging. For example, low-energy (kV) x-ray has been widely used in mammography or computed tomography (CT) for tissue characterization and disease diagnosis. Radiation has also been widely used for cancer treatment. For example, ionizing radiation has been extensively developed in radiation therapy using photon, electron, or proton beams to kill cancer cells. Such radiation particles damage the DNA structures in living cells directly or indirectly with free radicals and ions, preventing them from dividing and growing. Non-ionizing radiation has also been developed for cancer treatments, such as radiofrequency ablation and electroporation.

Despite the success of employing radiation in medical applications, the efficacy of using radiation for imaging and treatment remains to be improved. One critical issue is the lack of visibility of radiation energy deposition in tissues. For imaging, x-ray-based imaging modalities have historically relied on detecting penetrating radiation to visualize the internal tissues. This mechanism necessitates higher dose for imaging and full-view acquisition for three-dimensional (3D) reconstruction. For treatment, the efficacy of radiation depends on the energy deposition accuracy, which can be affected by many factors such as patient positioning, anatomy changes, and dose calculation uncertainties. Currently, there is a lack of effective tools to verify the treatment accuracy in real time to ensure the efficacy for tumor control and normal tissue sparing. For example, in radiation therapy, portal imaging can provide some 2D *in vivo* dose deposition, but it cannot provide 3D dose verification in real time.

Various techniques are being investigated to address the challenges above for radiation-based imaging and treatment. One promising direction is the radiation-induced acoustic (RA) imaging (Hickling *et al* 2014, 2018), which involves detecting acoustic signals generated from pulsed radiation beams and can be detected by ultrasound transducers. The pressure distribution reconstructed from these acoustic signals linearly correlates to the radiation energy deposition, making RA imaging a valuable tool for diagnostic imaging and therapy monitoring. For example, x-ray-induced acoustic computed tomography (XACT) is a novel imaging technique that excites the tissues with pulsed kV x-ray beams to induce acoustic signals (Xiang *et al* 2016, Li *et al* 2020, Robertson *et al* 2020, Samant *et al* 2020, Pandey *et al* 2021, Pogue *et al* 2021, Wang *et al* 2021a). It leverages the high contrast of tissue's radiation absorption properties to reveal fine structures deep within the body. Compared to x-ray imaging and CT, XACT features much lower imaging dose and can reconstruct 3D images from a single-view acquisition. Similar to XACT, proton pencil beams in proton therapy can also induce detectable acoustic signals, giving rise to the protoacoustic imaging technique for real-time proton dose verification (Hayakawa *et al* 1995, Jones *et al* 2014, Ahmad *et al* 2015, Kipergil *et al* 2017, Jones *et al* 2018, Nie *et al* 2018, Van Dongen *et al* 2019, Deurvorst *et al* 2022). Mast *et al* (2023) demonstrated the feasibility of measuring and reconstructing the proton irradiation-induced thermoacoustic source distribution using ultrasound arrays. Kalunga *et al* (2023) found that the ultrasound sensor arrangement can affect the localization accuracy of proton Bragg peak, and reported the in-silico localization errors within 2% when the sensors are optimally positioned and the experimental localization errors between 0.4 and 1.0 mm depending on the sensor arrangement and beam energy. Samant *et al* (2022) performed a simulation study to demonstrate the feasibility of the 3D protoacoustic imaging using a planar ultrasound array for proton dose verification.

In addition, studies have found that electric fields used in the electroporation for tumor ablation can also induce acoustic signals due to the electrical energy deposition (Neal *et al* 2014, Zarafshani *et al* 2019a, 2019b, Wang *et al* 2019, 2021a). This phenomenon led to the development of the electroacoustic tomography (EAT) technique, allowing the electroporation process to be monitored in real time.

Despite the promise of these novel RA techniques, a significant challenge with RA imaging is its low signal-to-noise ratio (SNR) in the acoustic signal. The detected acoustic signals are usually contaminated with background noises, such as electronic and system thermal noises, resulting in poor SNR and limiting the techniques' utility. Averaging is the most common way to improve the signal SNR by eliminating uncorrelated random noise (You and Choi 2020). Typically, hundreds to thousands of frames are required to be averaged to achieve satisfactory SNR (Jones *et al* 2015, Tang *et al* 2017), which significantly prolongs the imaging time and degrades the temporal resolution for therapy monitoring and moving-target imaging. In the case of ionizing radiation-based techniques such as XACT, the use of averaging requires the acquisition of a large number of frames, leading to a considerable increase in the imaging dose. Filtering is another widely used technique to denoise acoustic signals. It involves decomposing signals in different domains (e.g. frequencies and wavelet coefficients) and ruling out noises with thresholds (Kruger *et al* 1995, Wang *et al* 2004, Patil and Chavan 2012, Ngui *et al* 2013, Najafzadeh *et al* 2020). The main limitation of this method lies in the tradeoff between reducing the residual noises and preserving the true acoustic signals.

In the recent decade, deep learning has demonstrated excellent performance in various medical imaging tasks (Balakrishnan *et al* 2018, Würfl *et al* 2018, Zhu *et al* 2018, Hesamian *et al* 2019, Jiang *et al* 2020, Chang *et al* 2021, Jiang *et al* 2022a). Its applications have also been explored in acoustic imaging. Antholzer *et al* (2019) developed a deep convolutional neural network (CNN) to reconstruct accurate photoacoustic images from sparsely sampled data. Hariri *et al* (2020) used a multi-level wavelet CNN to improve the low-fluence

photoacoustic image quality. Jiang *et al* (2022b) developed a cascaded CNN to correct the limited-view artifacts in the matrix array-based proton acoustic imaging. In these studies, deep learning models were used as a post-processing technique, which reduced the artifacts and enhanced the quality of the images reconstructed by conventional algorithms. Earlier research (Jiang *et al* 2019a, 2021, 2022a) has indicated that the models can achieve better performance by utilizing advanced signal preprocessing methods to improve the input quality. Several studies have explored this preprocessing area for acoustic imaging. Gutta *et al* (2017) used a fully-connected network to map the limited bandwidth to the full bandwidth photoacoustic signals. Awasthi *et al* (2020) employed a CNN to remove the gaussian noise in photoacoustic signals in a simulation study. Yao *et al* (2020) proposed a recurrent neural network to verify proton range in one-dimension (1D), and later, proposed a different deep learning model with wavelet decomposition (Yao *et al* 2021) to reconstruct 3D proton range with high accuracy. Both studies used the acoustic signals measured by sparsely-distributed ultrasound sensors in a simulation study. And for the few-average signal denoising, Wang *et al* (2021b) proposed a dictionary learning-based workflow for both signal denoising and image enhancement in the photoacoustic microscopy (PAM), which reduced the laser fluence by 5 times without compromising the image quality. However, the proposed method used the K-SVD algorithm for sparse coding, assuming that the structures in the underlying signals have sparse representations. This assumption may not be true for images with complex structures and thus limits its applications. Wang *et al* (2023) performed a preliminary study to demonstrate the feasibility of deep learning to verify proton Bragg peak using few-averaged protoacoustic signals. However, the experimental data involved in this study is limited and they are measured using accelerometers at individual points in 1D. No images are available.

In this study, we developed a general deep-learning model to generate accurate and high-SNR acoustic signals from few-frame-averaged measurements. Compared to the previous dictionary learning-based method (Wang *et al* 2021b), the proposed deep learning-based method does not assume the properties of underlying signals and images, making it more generalizable for different tasks. To handle signals with diverse time windows, convolutions with multiple dilations are used in each inception block in the proposed model. This design allows for the encoding and decoding of signal features at different temporal scales. The model's performance was evaluated using the experimental data of kV x-ray-induced acoustic, protoacoustic, and electroacoustic signals, which have potential applications in (1) low-dose imaging with XACT, (2) real-time range verification in proton therapy using protoacoustic imaging, and (3) real-time therapy monitoring for electroporation using EAT. Results showed that the proposed model realized high-SNR signals from extremely few-frame acquisitions for all modalities, demonstrating the model's excellent generalizing ability across various RA techniques using different radiation sources. This major advance can greatly improve the clinical utilities of these novel radiation-induced acoustic techniques for low-dose diagnostic imaging and real-time treatment verification.

2. Methods

2.1. Problem formulation

Let $x \in R^{C\times T}$ be the real-valued RF signals acquired with a transducer having C channels in T timesteps, and $y \in R^{C\times T}$ be the corresponding ground truth signals. Then the problem can be formulated as finding a denoising pattern between x and y so that

$$\arg \min_{f} \quad \|f(x) - y\|_2^2$$

where f denotes the denoising function estimated by a deep learning model in this study.

2.2. Deep learning-based RF acoustic signal denoising

In this study, a general deep inception CNN (GDI-CNN) was used for the RF acoustic signal denoising. The few-averaged noisy signals are used as the input to the model, which generates the high-SNR denoised signals.

The proposed GDI-CNN utilizes the encoder-decoder architecture. In the encoder branch, seven inception blocks are stacked to extract high-dimensional hidden features from the noisy input signals. In the decoder branch, seven inception blocks are stacked to decode the features to generate the final denoised signals. In each inception block, seven convolutional layers with increasing dilation rates are used to capture features at different temporal resolutions. More details of the inception block can be found in the following figure A1 in the appendix A. Following the last inception block, a convolutional layer (filter size: 1×1, filter number: 1) with an activation function ('relu' for x-ray-induced and electroacoustic, and 'leaky relu' with 0.3 leaky rate for protoacoustic) is used to generate the final output denoised signals.

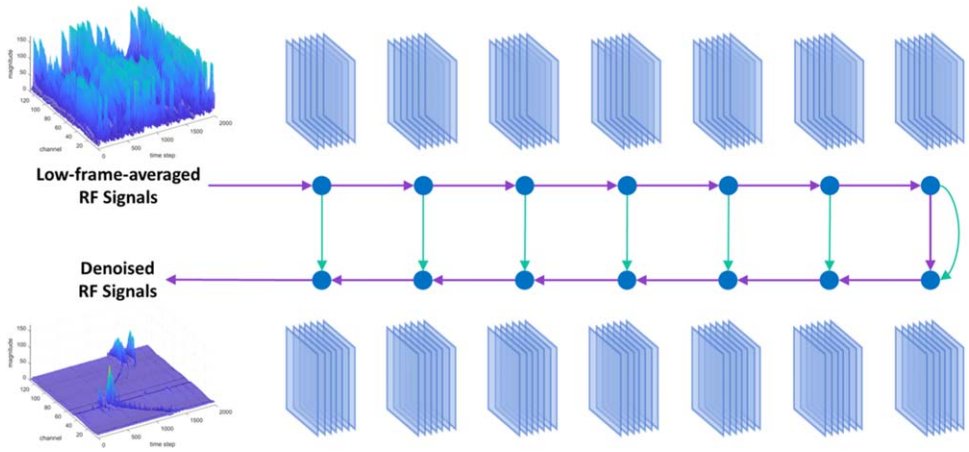


Figure 1. Structure of the proposed general deep inception convolutional neural network (GDI-CNN). It takes few-averaged signal as input, extracts and decodes hidden features with stacked inception convolutions (purple arrows →) and residual connections (green arrows →), and generates denoised signal in the end.

Compared to the original inception network (Szegedy *et al* 2015), this study uses convolutions with different dilations, rather than different filter sizes, to further enlarge the inception fields while significantly reducing the computational assumptions. Besides, residual connections bridge the inception blocks in the decoder with the mirrored ones in the encoder to facilitate the training of deep networks. Figure 1 shows the structure of the proposed general deep inception convolutional neural network (GDI-CNN). Source codes of the model implementation will be available publicly upon the acceptance of the manuscript.

3. Experiment design

3.1. Data acquisition

In this study, radiation-induced acoustic data of various modalities were collected to verify the proposed method's performance in few-frame-averaged signal denoising. For each modality, data were collected with different target geometries; and for each datum, RF signals from N repeated pulses ($N = 750$ for x-ray-induced acoustic and electroacoustic, and $N = 1500$ for protoacoustic) were collected, and were averaged to yield the N-frame-averaged signal, or fully-averaged signal, as the reference signal for model training and performance evaluation. The experiment setup of the x-ray-induced acoustic, protoacoustic, and electroacoustic data collection is illustrated in figure 2. Details are described in sections 3.1.1–3.1.3.

3.1.1. X-ray-induced acoustic data

X-ray acoustic signals were generated by irradiating agar phantoms using a portable x-ray source as shown in figure 2(a). Specifically, a battery-powered pulsed x-ray source (XR200, Golden Engineering, IN, USA) operated at 150 kVp was used to generate 50-nanosecond x-ray pulses (2.6 mR, cone angle: 40°) at a repetition rate of 20 Hz. X-ray beams were projected from the bottom of a water tank at a distance of 30 cm, in which a target-holding phantom was placed along the beam trail. The phantoms were made of 3% agar-water solution (BactoTM, Becton, Dickinson and Company, NJ, USA), containing various targets made of 1/16 inch and 1/8 inch thick lead wires.

The acoustic signals were collected by a ring-shaped 128-channel ultrasound transducer array (PA probe, Doppler Co. Limited, Guangzhou, China; central frequency: 5 MHz, bandwidth: ⩾60%). The detector was placed in the water tank, surrounding the phantom. To trigger the data acquisition, a Ce:Lu_2SiO_5 crystal (MTI Corporation, CA, USA) was used to scintillate at a wavelength of 418 nm upon incident with stray x-rays. The scintillator was connected to a photodetector (PDA100A, Thorlabs, NJ, USA) for light-to-volt conversion and followed with a waveform generator (33600A, Keysight, CA, USA) to relay the trigger signal to the 128-channel data acquisition system (SonixDAQ, DK Medical, Canada). Raw acoustic signals were conditioned by a low-noise preamplifier (TomoWave Laboratories, Inc., TX,USA). For each trigger, acoustic signals were collected for 67.5 μs (sampling rate: 40 MHz), forming a frame. For each x-ray acoustic data acquisition, 750 frames were collected to ensure the SNR.

In this experiment, 58 x-ray-induced acoustic data with different target sizes and locations were collected for model training and testing.

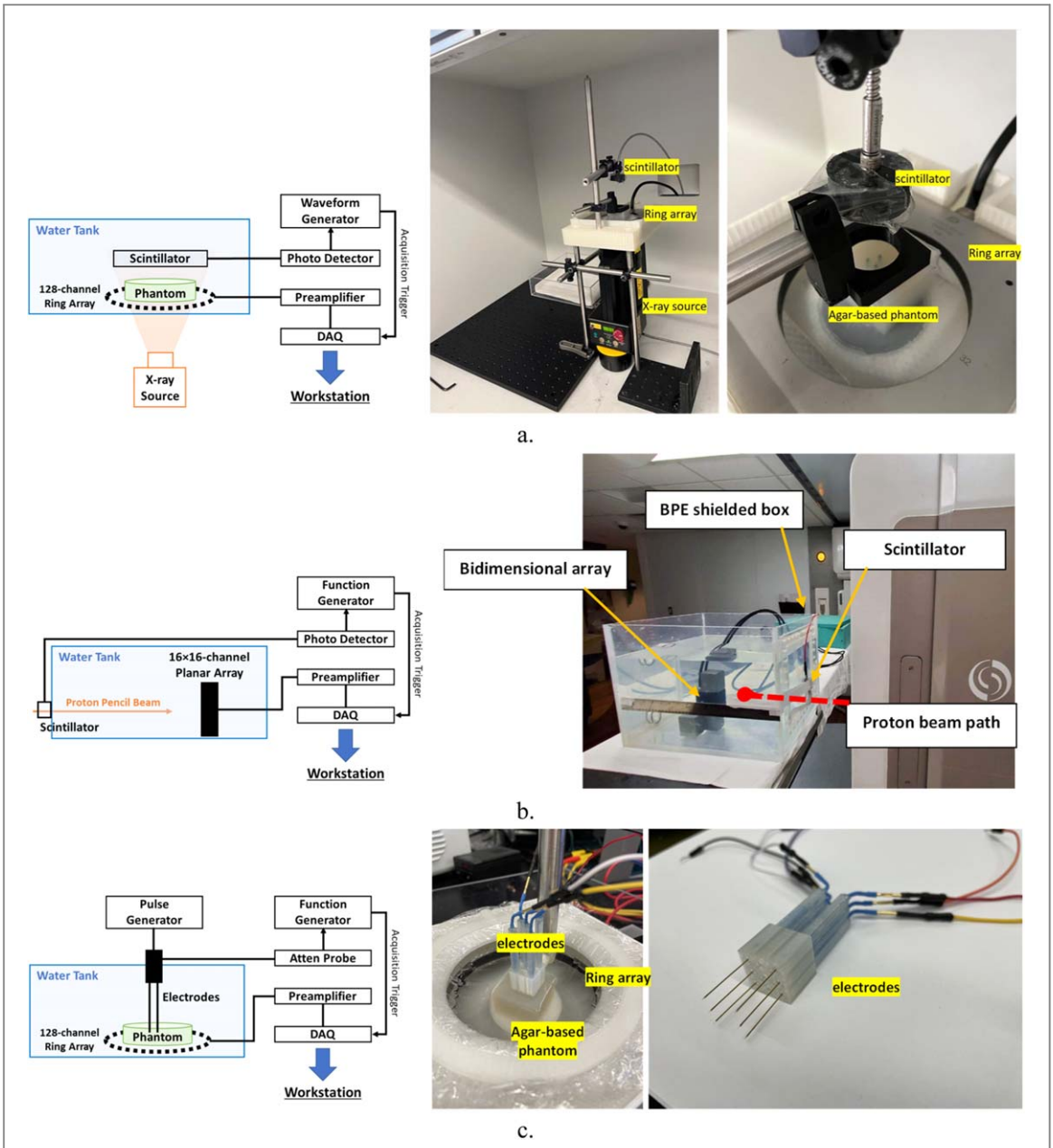


Figure 2. Illustration (left) and photos (right) of the experiment setup of (a) x-ray-induced acoustic data collection, (b) protoacoustic data collection, and (c) electroacoustic data collection.

3.1.2. Protoacoustic data

Protoacoustic signals were generated by irradiating a water tank using a clinical synchrocyclotron as shown in figure 2(b). Specifically, a clinical synchrocyclotron Mevion s250i (Mevion Medical Systems, MA, USA) was used to deliver proton beams (energy: 107.6 MeV, proton range: 8.69 cm, beam pulse width: 4 μs, uncollimated Bragg Peak) at a repetition rate of 750 Hz. The charge of each proton beam was set to 8 pC, resulting in a dose of about 1 cGy.

The water tank has a dimension of 32 cm $\times$ 32 cm $\times$ 40 cm and was placed on the treatment couch.

The acoustic signals were collected by a custom planar 16 $\times$ 16-channel ultrasound transducer array (Matrix probe, Doppler Co. Limited, Guangzhou, China; central frequency: 1 MHz, bandwidth: $\geqslant$60%, element size: 3 $\times$ 3 mm^2). The detector was placed in the water tank at a depth of 12.7 cm with respect to the proton beam entrance. To trigger the data acquisition, a scintillation crystal (BC408, Epic-Crystal, Guangzhou, China) was attached to the beam entrance window to detect the proton beam passage. The scintillator output was connected to a function generator (SDG-2122X, SIGLENT, OH, USA) to relay the trigger signal to the data acquisition system (Legion ADC, Photosound Technologies, USA), which was housed in a Borated Polyethylene (BPE) shielded box for neutron shielding. For each trigger, acoustic signals were collected for 67.5 μs (sampling rate:

40 MHz), forming a frame. For each protoacoustic data acquisition, 1500 frames were collected to ensure the SNR.

In this experiment, 5 protoacoustic data with various proton beam coordinates were collected for the model training and testing.

3.1.3. Electroacoustic data

Electroacoustic signals were generated by irradiating agar phantoms using an electrical pulse generator as shown in figure 2(c). Specifically, a custom nanosecond electrical pulse generator (VilniusTECH, Lithuania) was used to deliver 100-nanosecond electrical pulses (repetition rate: 1000 Hz) via several tungsten electrodes (573400, A&M Systems, Carlsborg, USA) to an agar phantom in a water tank. The voltage on the electrodes was set to 100 Volt, resulting in an electric field of hundreds of volts per centimeter. A multi-electrode holder was 3D-printed to enable different electrode arrangements. Distances between electrodes varied between 5 and 15 mm.

The acoustic signals were collected using the same 128-channel ring array mentioned in section 3.1.1. The detector was placed in the water tank, surrounding the phantom. A polyethylene film was used to isolate the electrodes from the ultrasound transducer to reduce the direct impact of the high-voltage pulsed electric field on the piezoelectric ultrasound probe surface. To trigger the data acquisition, a parallelly connected high-voltage probe (P4250, Keysight Technologies, USA) was used to detect the pulse generation directly from the output. The other ends of the probes were connected to a function generator to relay the trigger signal to the 128-channel data acquisition system (SonixDAQ, DK Medical, Canada). Raw electroacoustic signals were boosted by a custom low-noise preamplifier. For each trigger, acoustic signals were collected for 67.5 μs (sampling rate: 40 MHz), forming a frame. For each electroacoustic data acquisition, 750 frames were collected to ensure the SNR.

In this experiment, 52 electroacoustic data from various electrode arrangements and locations were collected for model training and testing.

3.2. Model training

The proposed model was trained in a modality-specific method. For each imaging technique, a model was trained and tested using the corresponding dataset. Specifically, for the XACT, 58 data were acquired, of which 47, 5, and 6 data were used for model training, validation, and testing, respectively. For each datum, 10 frames were averaged to generate a sample, which was used as the input to the model; and the 750-frame-averaged signals were used as the 'ground truth' signals. As a result, 3525 samples were used for model training, 375 samples were used for validation, and 450 samples were used for testing. For the protoacoustic, five data were acquired. Due to the limited dataset, four of the five data were used for model training, and the other one data was used for model validation. For each datum, 20 frames were averaged to generate a sample, which was used as the input to the model; and the 1500-frame-averaged signals were used as ground truth. As a result, 300 samples were used for model training and 75 samples were used for validation. For the EAT, 52 data were acquired, of which 43 data were used for model training, 3 for validation, and 6 for testing. For each datum, a single frame was a sample, which was used as the model input; and the 750-frame-averaged signals were used as the ground truth. To balance the computing time and model performance, one frame out of every ten was used for denoising. As a result, 3225 samples were used for model training, 225 samples were used for validation, and 450 samples were used for testing.

To address the negativity caused by the limited bandwidth (Shen *et al* 2020), for the x-ray-induced acoustic and electroacoustic signals, envelope was performed on the 750-frame-averaged signals; for the protoacoustic signals, we keep the original signals as the negativity is still under investigation and there is no consensus in addressing this issue.

To improve the learning target ('ground truth' signal) quality for the protoacoustic data, the discrete wavelet transform (DWT) was used to further improve the SNR of the 1500-frame-averaged signals, which uses a scaled and predefined wavelet and scaling functions to convolute with signals. For DWT, we implemented the 'Coiflets' wavelet filter (order = 5) (Ngui *et al* 2013) and the 'sqtwolog' threshold selection (Donoho and Johnstone 1994) using the MATLAB wavelet toolbox. The SNRs of the fully-averaged (750-frame-averaged) signals were already good and thus no wavelet denoising was performed for these two modalities.

During the model training, the few-frame-averaged signal was used as input to the model, whose weights were optimized by minimizing the mean squared errors (MSE) between the denoised signal and the corresponding ground truth signal. The optimizer was set to 'Adam' (Kingma and Ba 2014) with a learning rate of 0.001. The batch size was set to 1 to account for the memory.

An overall workflow for the model training is shown in figure 3.

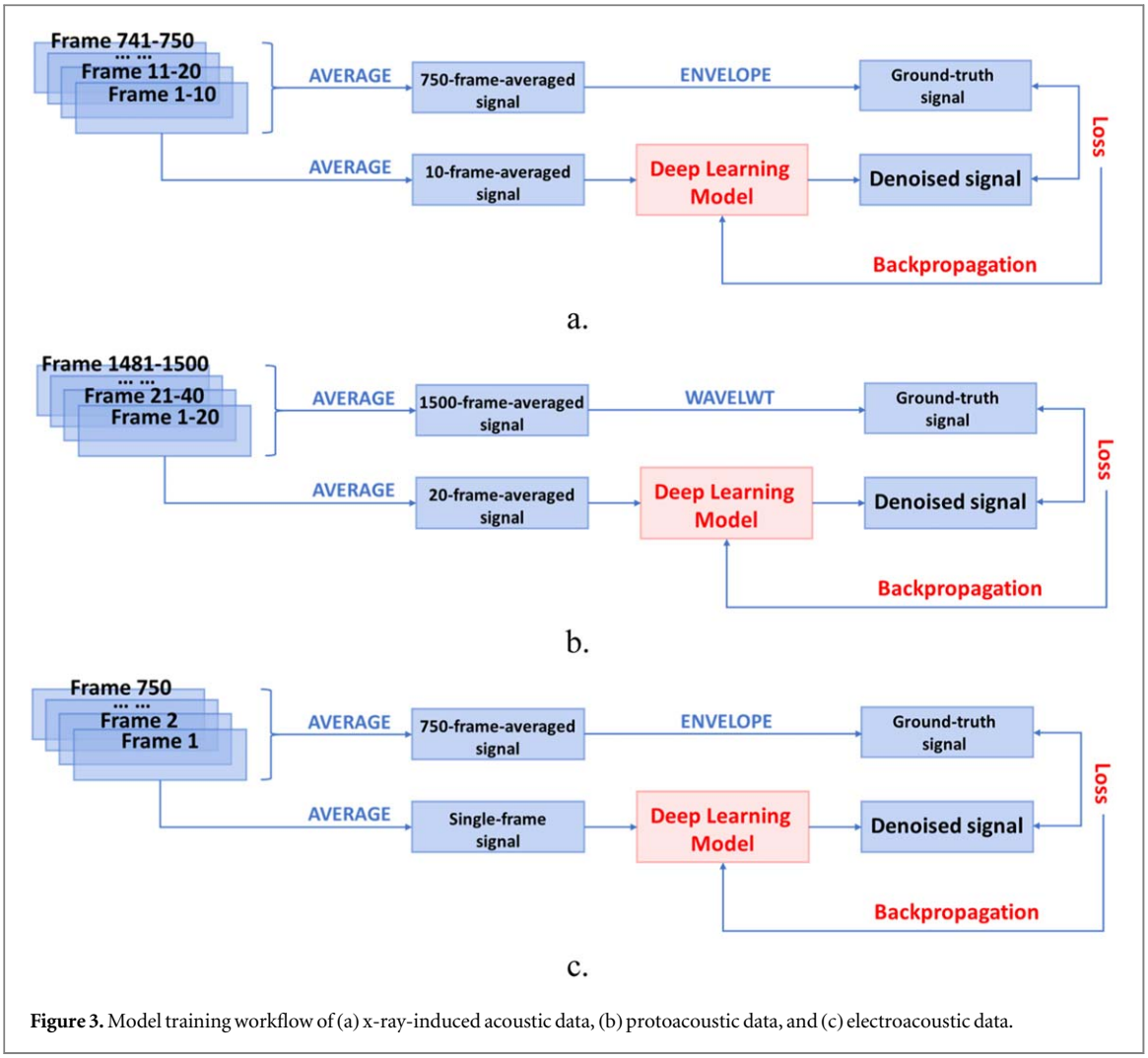


Figure 3. Model training workflow of (a) x-ray-induced acoustic data, (b) protoacoustic data, and (c) electroacoustic data.

3.3. Model evaluation

The signals denoised by the proposed model were compared to the 'ground truth' signals. Meanwhile, the widely used lowpass filtering was also included in this study as a comparison baseline. The cutoff frequency for the lowpass filter was set to 6 MHz, 200 kHz, and 6 MHz for x-ray-induced acoustic, protoacoustic, and electroacoustic signals, respectively. The lowpass filter was implemented using the MATLAB signal processing toolbox.

To further demonstrate the values of the denoising, acoustic images were reconstructed using the back-projection algorithm from the denoised signals, and were then compared to those reconstructed from the ground truth signals. An overall workflow for the model evaluation is shown in figure 4.

There are some slight differences in the preprocessing of RA signals for different modalities before reconstruction. The differences are due to the different characteristics of the acoustic data and correspondingly the variations in the routine preprocessing methods used for different data. For the x-ray-induced acoustic and electroacoustic data, envelope was performed to address the negative values caused by the limited bandwidth; the negativity issue was not addressed for the protoacoustic data in this study as this issue is still under investigation and no consensus is established. For the protoacoustic data, the fully-averaged signals still have poor SNR, and thus wavelet was used for denoising before reconstruction; the fully-averaged signals for the x-ray-induced and electroacoustic signals have good SNRs, and thus no wavelet was performed.

The results were evaluated both qualitatively and quantitatively using metrics of MSE, peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM).

The performance of the proposed GDI-CNN was compared to the U-Net for all three involved modalities. For U-Net, we implemented two versions: the original U-Net (Ronneberger *et al* 2015) (hereafter referred to as U-Net), and the lite U-Net (hereafter referred to as U-Net-Lite). U-Net-Lite has the same architecture (e.g. scale levels and skip connections) as U-Net but has fewer convolutional layers and fewer filter numbers in each scale level, which makes it have a similar order of trainable parameters to the proposed GDI-CNN. The number of the model's parameters is shown in tables 1 and 2. U-Net and U-Net-Lite were trained and tested following the same process as the proposed GDI-CNN.

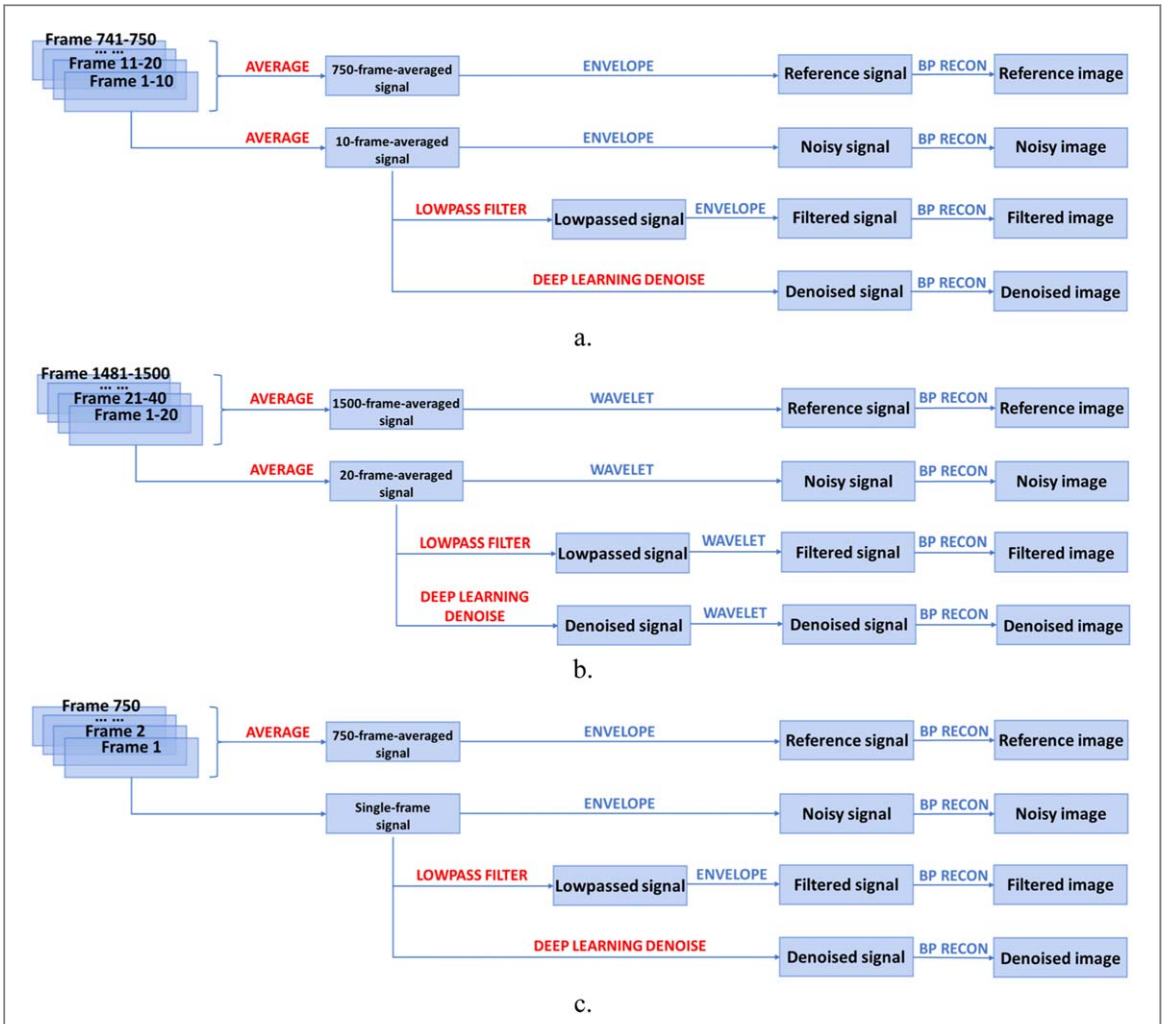


Figure 4 Evaluation workflow of (a) x-ray-induced acoustic data, (b) protoacoustic data, and (c) electroacoustic data. Reference signals/images are used as the ground truth for result evaluation.

Table 1. Quantitative analysis of radio-frequency acoustic signals.

Imaging techniques	Metrics	Values[a]			
		Filtered	U-Net (34 539 585)[b]	U-Net-Lite (1 180 081)[b]	**Proposed** (1 476 609)[b]
X-ray acoustic	MSE	0.0099 ± 0.0053	$\mathbf{0.0021 \pm 0.0033}$	0.0033 ± 0.0076	0.0031 ± 0.0068
	PSNR	20.934 ± 3.2003	$\mathbf{29.394 \pm 4.6138}$	28.645 ± 4.9885	28.930 ± 5.1021
	SSIM	0.5851 ± 0.0961	$\mathbf{0.8193 \pm 0.1094}$	0.7993 ± 0.1399	0.8063 ± 0.1324
Protoacoustic	MSE	0.0346 ± 0.0202	0.0056 ± 0.0017	0.0074 ± 0.0020	$\mathbf{0.0041 \pm 0.0024}$
	PSNR	15.269 ± 2.3976	22.678 ± 1.3303	21.477 ± 1.1673	$\mathbf{24.305 \pm 1.9456}$
	SSIM	0.3690 ± 0.0368	0.6306 ± 0.0519	0.5957 ± 0.0478	$\mathbf{0.7521 \pm 0.0443}$
Electroacoustic	MSE	0.0068 ± 0.0014	0.0004 ± 0.0001	0.0004 ± 0.0001	$\mathbf{0.0003 \pm 0.0001}$
	PSNR	21.754 ± 0.9117	34.382 ± 1.8914	34.490 ± 1.9424	$\mathbf{35.243 \pm 1.8811}$
	SSIM	0.5934 ± 0.0491	0.8816 ± 0.0178	0.8926 ± 0.0134	$\mathbf{0.9018 \pm 0.0122}$

[a] Values are calculated with data normalized to [0, 1], and expressed as mean ± std.

[b] Number in brackets indicate the total parameters of the model.

4. Results

4.1. Radiation-induced acoustic signal denoising

Figure 5 shows a representative 10-frame-averaged x-ray-induced acoustic data in the testing dataset. It is the image in the ring array plane shown in figure 2(a). The images (h1–m1) can be understood as a combination of dose distribution and target shapes. The three high-intensity dots (orange to red) indicate the lead wires. Outside the lead wires is the uniform agar phantom. The intensity falloff indicates the deposited dose falloff. It is caused

Table 2. Quantitative analysis of reconstructed images.

Imaging techniques	Metrics	Values[a]			
		Filtered	U-Net (34 539 585)[b]	U-Net-Lite (1 180 081)[b]	**Proposed** (1 476 609)[b]
X-ray acoustic	MSE	0.0114 ± 0.0133	$\mathbf{0.0069 \pm 0.0158}$	0.0075 ± 0.0173	0.0093 ± 0.0204
	PSNR	21.822 ± 4.7896	$\mathbf{28.773 \pm 7.5435}$	28.458 ± 7.4704	26.430 ± 6.7062
	SSIM	0.7769 ± 0.1046	$\mathbf{0.9317 \pm 0.0797}$	0.9310 ± 0.0798	0.9204 ± 0.0861
Protoacoustic	MSE	0.0178 ± 0.0148	0.0036 ± 0.0014	0.0024 ± 0.0014	$\mathbf{0.0011 \pm 0.0003}$
	PSNR	18.655 ± 3.0915	24.809 ± 2.0970	27.072 ± 2.900	$\mathbf{29.604 \pm 1.2060}$
	SSIM	0.8313 ± 0.0602	0.9843 ± 0.0031	0.9821 ± 0.0034	$\mathbf{0.9905 \pm 0.0014}$
Electroacoustic	MSE	0.0321 ± 0.0218	0.0002 ± 0.0002	0.0003 ± 0.0004	$\mathbf{0.0001 \pm 0.0001}$
	PSNR	16.385 ± 4.0832	39.965 ± 4.0093	37.645 ± 5.5664	$\mathbf{40.804 \pm 3.5220}$
	SSIM	0.6307 ± 0.1661	0.9859 ± 0.0057	0.9804 ± 0.0107	$\mathbf{0.9878 \pm 0.0045}$

[a] Values are calculated with data normalized to [0, 1], and expressed as mean ± std.

[b] Number in brackets indicate the total parameters of the model.

by the uneven exposure of x-ray radiation, which is inherited from the x-ray source used in this study. Artifacts indicated by the red arrows are caused by the channel cross talk in the DAQ device used in this study. Due to the few-frame average, the RF signals are severely contaminated by background noises. The lowpass filter partially removes the noises but still yields poor SNR. Overall, the deep learning models, including U-Net, U-Net-Lite, and the proposed GDI-CNN, all effectively remove the noises while accurately preserving the signals, showing very good agreement with the reference 750-frame-averaged signals. The acoustic image reconstruction further demonstrates the effectiveness of the deep learning methods. The targets in the agar phantom are completely diminished in the few-frame-averaged reconstruction, and can be hardly distinguished in the lowpassed reconstruction. In contrast, the image reconstructed from the deep learning denoised signals has clear and accurate structures (as shown in figures 5(j)–(l)), which has high similarity to the reference image reconstructed from the 750-frame-averaged signals.

Figure 6 shows a representative 20-frame-averaged protoacoustic data in the validation dataset (the one data left out from the model training dataset). Figures 6(h)–(m) shows the 3D pressure in three orthogonal planes. Specifically, referred to the experimental setup shown in figure 2(b), the first row is the plane orthogonal to the proton beam, and the second and the third rows are the planes along the proton beam. Since the proton beams are delivered to a homogenous acoustic medium, water, the results can be understood as pure dose distribution. The high-intensity region (red) corresponds to the proton Bragg peak. The low-intensity region (blue) can be ignored when interpreting the image. It corresponds to the negative acoustic signals in figures 6(A), (d). Noted that the protoacoustic imaging technique is still under development, and the image interpretation is an ongoing research subject. Artifacts indicated by the red arrows are caused by the channel cross talk in the DAQ device used in this study. Signals cannot be distinguished from the few-frame-averaged signals. The lowpass filter partially removes the noises but the signals are still highly contaminated. The deep learning models, including U-Net, U-Net-Lite, and the proposed GDI-CNN, all effectively remove the noises while preserving the signals. The acoustic image reconstructed from the few-frame-averaged signals has poor image quality with severe artifact contamination, and the one from lowpass-filtered signals shows obvious artifacts and errors. In contrast, the image reconstructed from the deep learning denoised signals shows a clear and accurate proton range, agreeing well with the 1500-frame-averaged results. Furthermore, despite the lower intensity, the proposed GDI-CNN demonstrates a more accurate shape than the U-Net and the U-Net-Lite, as indicated by figures 6(j)–(m). In addition, GDI-CNN better removes the artifacts caused by the DAQ channel cross talk, as indicated by the red arrows.

Figure 7 shows representative single-frame electroacoustic data in the testing dataset. Figures 7(h1)–(m1) shows the image in the ring array plane shown in figure 2(c). The three high-intensity dots indicate the areas of strong electrical energy deposition around the electrodes in the agar phantom. The intensity falloff indicates the deposited energy falloff, which is caused by the electric field falloff. Artifacts indicated by the red arrows are caused by the channel cross talk in the DAQ device used in this study. Similar to the x-ray-induced acoustic results, the deep learning-based denoising method considerably improved the SNR of the electroacoustic signals. The reconstructed acoustic images further confirm the effectiveness of the deep learning models. However, as indicated by the black arrows in figures 7(c)–(f), the U-Net and the U-Net-Lite tend to smooth out the low-intensity signals, which leads to a more severe limited-angle response in their denoised signals. In contrast, the proposed GDI-CNN well preserve the low-intensity signals. Therefore, the cylinder

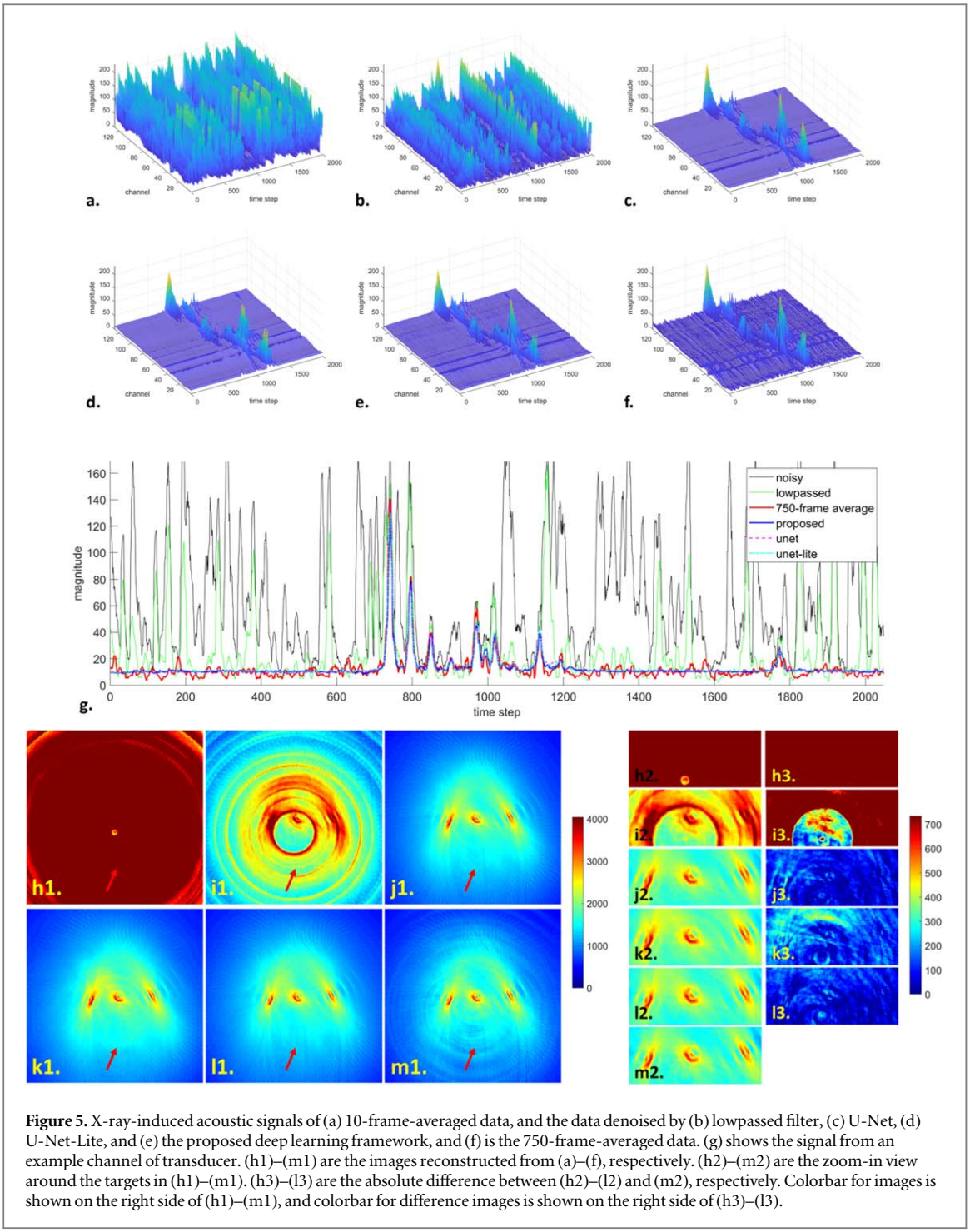


Figure 5. X-ray-induced acoustic signals of (a) 10-frame-averaged data, and the data denoised by (b) lowpassed filter, (c) U-Net, (d) U-Net-Lite, and (e) the proposed deep learning framework, and (f) is the 750-frame-averaged data. (g) shows the signal from an example channel of transducer. (h1)–(m1) are the images reconstructed from (a)–(f), respectively. (h2)–(m2) are the zoom-in view around the targets in (h1)–(m1). (h3)–(l3) are the absolute difference between (h2)–(l2) and (m2), respectively. Colorbar for images is shown on the right side of (h1)–(m1), and colorbar for difference images is shown on the right side of (h3)–(l3).

electrode distorts less in the proposed image than in the U-Net and U-Net-Lite images, as shown in figures 7(j)–(m).

Tables 1 and 2 show the quantitative results in both acoustic signal and image domains. Compared to the lowpass-filtered few-frame-averaged results, the signals/images denoised by the deep learning models showed substantially improved quality. The proposed GDI-CNN shows the best performance for the protoacoustic and electroacoustic denoising. For the x-ray-induced acoustic denoising, the U-Net demonstrates better quantitative metrics on average among all samples than GDI-CNN. Note that the reference x-ray-induced acoustic images reconstructed using 750-frame-averaged signals still showed considerable noises with limited SNR (shown in figure 5(m)), possibly affecting the metric calculation based on the reference images. Quantitative evaluation within regions of interest (ROIs) (targets in the x-ray-induced acoustic images, and electrodes in the

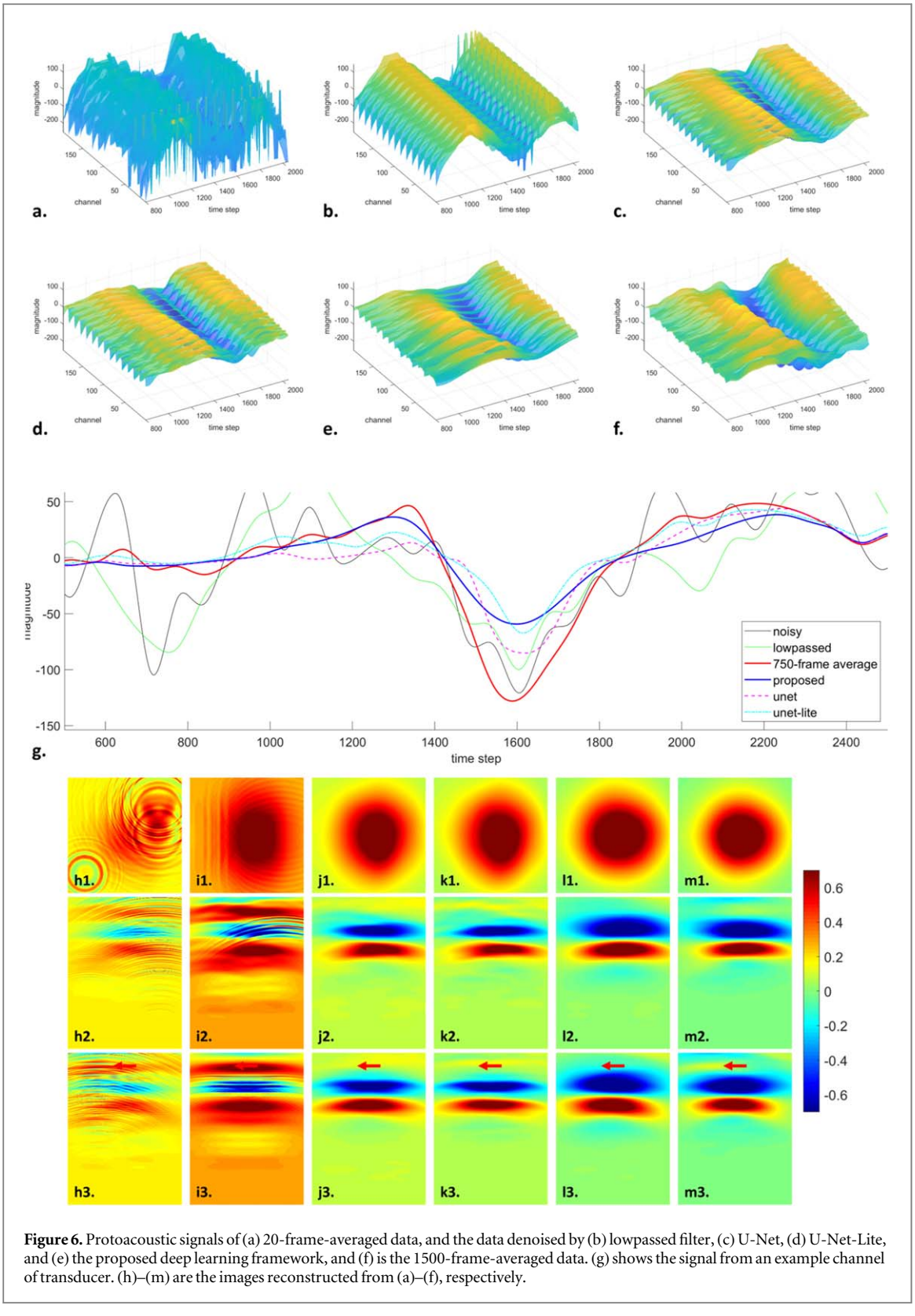


Figure 6. Protoacoustic signals of (a) 20-frame-averaged data, and the data denoised by (b) lowpassed filter, (c) U-Net, (d) U-Net-Lite, and (e) the proposed deep learning framework, and (f) is the 1500-frame-averaged data. (g) shows the signal from an example channel of transducer. (h)–(m) are the images reconstructed from (a)–(f), respectively.

electroacoustic images) is consistent with the whole-image evaluation results. More details of the ROI evaluation can be found in the appendix C.

4.2. Runtime

The proposed deep learning framework was implemented using the Keras framework with the TensorFlow backend. The model training and testing were performed on a computer equipped with a CPU of Intel Xeon

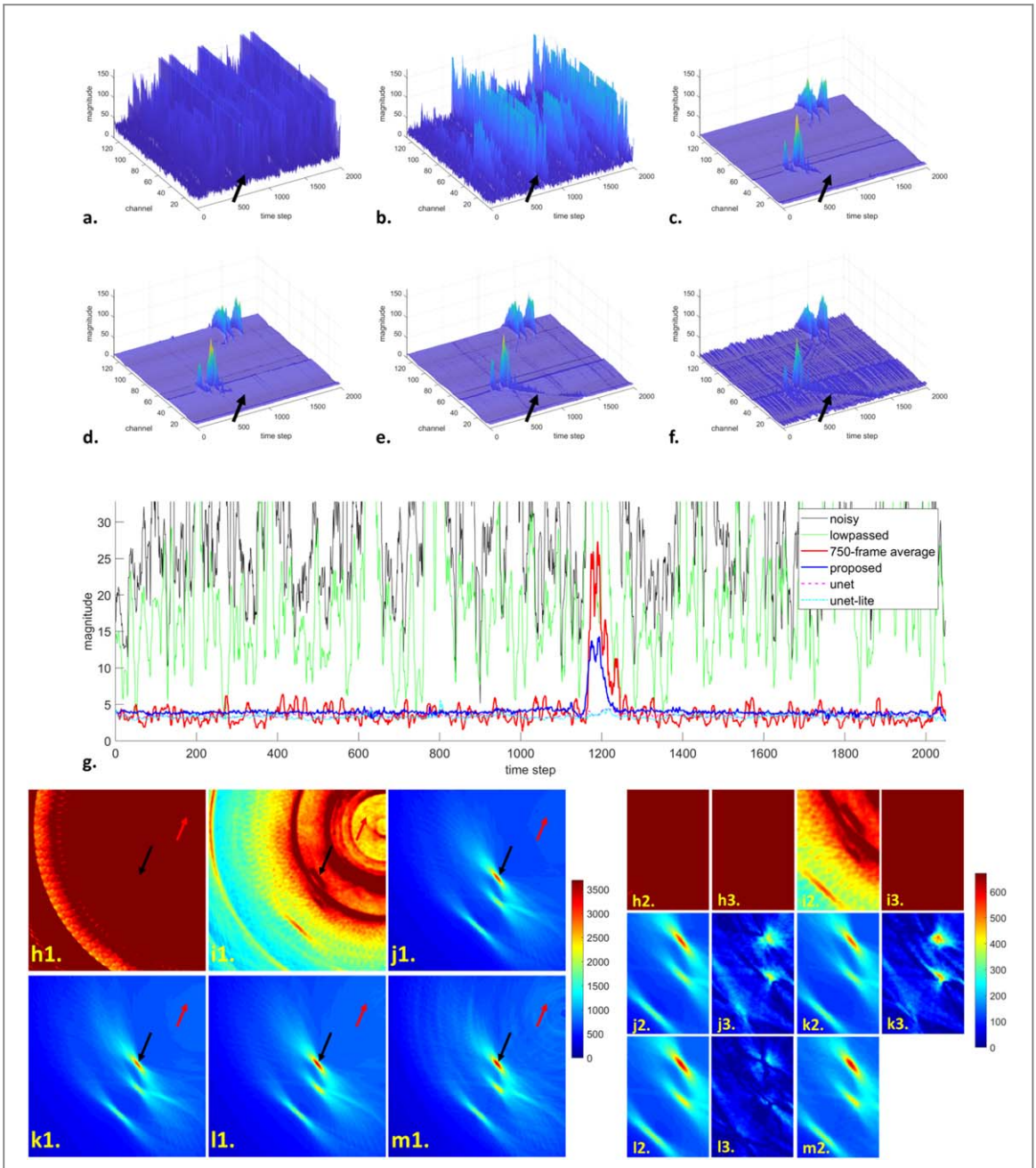


Figure 7. Electroacoustic signals of (a) single frame data, and the data denoised by (b) lowpassed filter, (c) U-Net, (d) U-Net-Lite, and (e) the proposed deep learning framework, and (f) is the 750-frame-averaged data. (g) shows the signal from an example channel of transducer. (h1)–(m1) are the images reconstructed from (a)–(f), respectively. (h2)–(m2) are the zoom-in view around the electrodes in (h1)–(m1). (h3)–(l3) are the absolute difference between (h2)–(l2) and (m2), respectively. Colorbar for images is shown on the right side of (h1)–(m1), and colorbar for difference images is shown on the right side of (h3)–(l3).

and 32GB memory and a GPU of NVIDIA Titan RTX (24GB memory). The denoising process is fully automated, which takes about 0.25 s for 128-channel x-ray-induced acoustic and electroacoustic data, and about 1.4 s for 16 × 16-channel protoacoustic data. Parallel computing can be used to achieve real-time denoising.

5. Discussion

Radiation-induced acoustic imaging is a novel modality to reveal the *in vivo* radiation energy deposition. It gives rise to various techniques, such as the XACT for diagnostic imaging, the protoacoustic imaging for real-time proton range verification in proton therapy, and the EAT for electroporation monitoring. To improve their clinical utilities, it is highly desirable to reconstruct acoustic images from few-frame-averaged signals for (1)

better temporal resolution with less dynamic blurriness, and (2) lower imaging dose. However, conventional averaging method or filter-based algorithms can hardly achieve satisfactory SNR from few-frame-averaged signals. In this study, we presented a deep learning-based framework to denoise few-frame-averaged acoustic signals. Results demonstrated the effectiveness and efficiency of the proposed method in achieving accurate and high-SNR signals from few-frame-averaged measurements. The acoustic image quality is considerably improved using the signals denoised by the proposed method. Both the denoised few-frame-averaged signals and the corresponding reconstructed images show very high agreement with the reference high-frame-averaged ones.

The proposed deep learning-based framework can be generalized to denoise various kinds of acoustic signals. The radiation pulse width can vary among different applications, resulting in acoustic signals of different frequencies. As a result, signal features can be presented at various temporal scales, for which inception blocks with multiple dilation rates are used in our proposed GDI-CNN. This inception-style architecture enables multi-scale feature extraction, allowing the model to capture and process temporal dependencies at diverse scales concurrently. This multi-scale representation improves the model's sensitivity to both rapid and slow time-varying signals, and thus generalizes the model for denoising acoustic signals from various sources. In addition, these inception blocks offer an exponential expansion of the receptive field while maintaining parameter efficiency compared to larger kernels. And without pooling operation, the temporal and spatial resolutions are potentially better preserved by the model. In this study, x-ray acoustic and electroacoustic signals are of mega-hertz, while protoacoustic signals are of tens of kilo-hertz due to the different beam pulse durations. As shown in the results, the proposed model achieved excellent denoising performance for all these modalities, indicating its generalizing ability for acoustic signals with large temporal characteristic variations.

Results demonstrated GDI-CNN's capabilities of reducing crosstalk artifacts in the involved acoustic imaging. Crosstalk is primarily caused by inadequate insulation between channels. It can lead to unwanted signals appearing in multiple channels simultaneously when background noises or interferences occur, which is random. The proposed model can perceive the random noise as inconsistencies across the dataset. Consequently, during the training phase, the model learns to ignore this noise to better approximate the underlying denoising patterns. In addition, L2 regularization was used in the model training, which discouraged the model from overfitting to the training data, particularly the random noise present in the reference fully-averaged signals. This strategy can make the model generalize better and produce cleaner outputs during testing. We also observe this 'student beats the teacher' phenomenon in previous studies (Ghafoorian *et al* 2018, Jiang *et al* 2019b), wherein the deep learning models can predict results of better quality than the training ground truth.

The proposed method makes the radiation-induced acoustic imaging a valuable tool for both imaging and therapy monitoring. In the context of ionizing imaging, 'as low as reasonably achievable (ALARA)' principle suggests that reasonable efforts should be made to minimize the radiation exposure to patients to reduce the risk of adverse effects. As shown in results, the proposed method achieves comparable XACT image quality to 750-frame-averaged results using only 10-frame-averaged signals, considerably reducing the imaging dose by 98.7%. For the therapy monitoring, temporal resolution is an important performance metric besides accuracy. High temporal resolution not only ensures the real-time efficiency but also reduces dynamic blurriness induced by temporal averaging. Results showed that the proposed method realized proton range accuracy comparable to 1500-frame-averaged results using only 20-frame-averaged signals, which greatly improves the proton range verification frequency from 0.5 to 37.5 Hz. And the method reached EAT image quality comparable to 750-frame-averaged results using only a single frame signal, substantially increasing the electric field monitoring frequency from 1 Hz to 1 kHz. Although the temporal frequency can be higher than the clinical requirement, it pushes the temporal averaging to the minimum and provides a possibility to visualize the energy deposition process with ultra-high temporal speed. These significant improvements can substantially expand the clinical utilities of these novel imaging techniques for various applications in diagnosis and treatment.

There are some limitations of this study. First, the datasets used in this study are relatively small due to the laborious data collection. Additionally, experimental data were collected from agar phantoms (containing targets) or water tanks for feasibility demonstration. In the future, more acoustic data from more complex tissues and patients are warranted to further evaluate the clinical utility of the proposed denoising method. Second, we focused on three kinds of RA signal denoising in this study due to the data availability of our labs. In the future, more imaging modalities, such as photoacoustic and MV x-ray-induced acoustic data, can be collected to further test the performance and generalizing ability of the proposed method. Third, we evaluated the results using general metrics such as element-wise errors (MSE), peak SNR (PSNR), and structural similarities (SSIM) in this study. In the future, task-specific metrics can be developed to make the evaluation

more practical, which provides a more targeted assessment of the denoising performance. Furthermore, it is noted that the proposed deep learning model can experience degraded performance when further reducing the averaging frame number. Task-specific metrics can be used to explore the minimum averaging number required by the model. Fourth, it is note that, in this study, we focused on the signal denoising which is a preprocessing step for the RA imaging. The final RA image quality can be additionally improved by the advances in reconstruction and postprocessing techniques. In the future, the efficacy of the proposed denoising technique can be further evaluated in the entire chain of RA imaging. Fifth, we did not perform thorough ablation tests to clarify the contribution of each component in the proposed GDI-CNN in this study. More careful examination of the network structure can be conducted in future studies, which could further optimize the model's performance.

6. Conclusion

The proposed deep learning-based denoising framework is a generalized method for few-frame-averaged acoustic signal denoising, which significantly improves the radiation-induced acoustic imaging's clinical utilities including low-dose XACT, real-time protoacoustic-based proton range verification, and real-time electroacoustic-based electroporation monitoring.

Acknowledgments

This work was supported by the National Institutes of Health under Grant No. R01-EB028324, R01-EB032680, R37CA240806, and R01-CA279013. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This work was also supported by the American Cancer Society (133697-RSG-19-110-01-CCE), and the UCI Chao Family Comprehensive Cancer Center (P30CA062203).

Data availability statement

The data cannot be made publicly available upon publication because no suitable repository exists for hosting data in this field of study. The data that support the findings of this study are available upon reasonable request from the authors. The experimental data and the source codes of the proposed denoising method will be available publicly upon acceptance of this paper.

Author Contributions

Dr. Lei Ren and Dr. Liangzhong Xiang are the corresponding authors who supervised and instructed the entire project. Zhuoran Jiang led the deep learning-based method development, experiment design, and results analysis. Siqi Wang led the experiment conduction and troubleshooting of acoustic data collection. Specifically, Siqi Wang was responsible for the X-ray acoustic data collection; Yifei Xu was responsible for the electroacoustic data collection; and Gilberto Gonzalez was responsible for the protoacoustic data collection. Leshan Sun developed the codes for acoustic image reconstruction, lowpass filtering, and wavelet-based denoising. Dr. Yong Chen and Dr. Q. Jackie Wu provided support for the experiment and study conduction. All the authors contributed to the manuscript and approved the publication of this study.

Appendix A. Detailed structure of the inception block in the proposed GDI-CNN

Figure A1 Shows the detailed structure of the inception block in the proposed GDI-CNN.

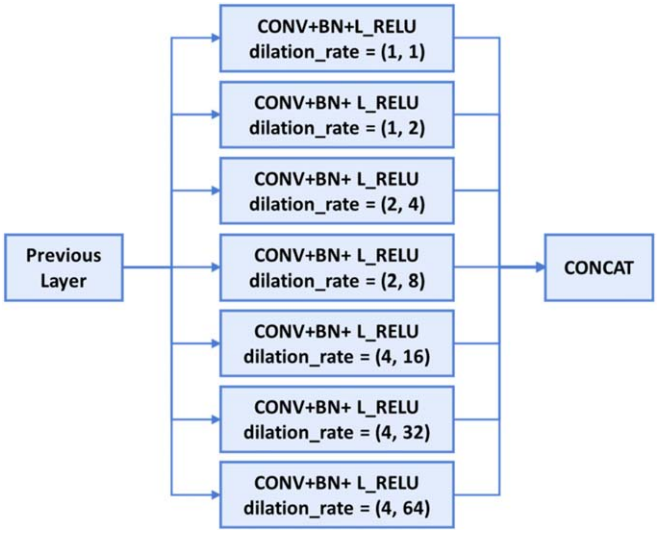


Figure A1. Structure of the inception block. CONV is a dilation convolutional layer (filter size: 3 × 3, filter number: 16, dilation rate is indicated in the box); BN is a batch normalization layer; L_RELU is the activation function ('relu' for x-ray-induced and electroacoustic, and 'leaky relu' with 0.3 leaky rate for protoacoustic); CONCAT is a concatenation layer in the channel dimension.

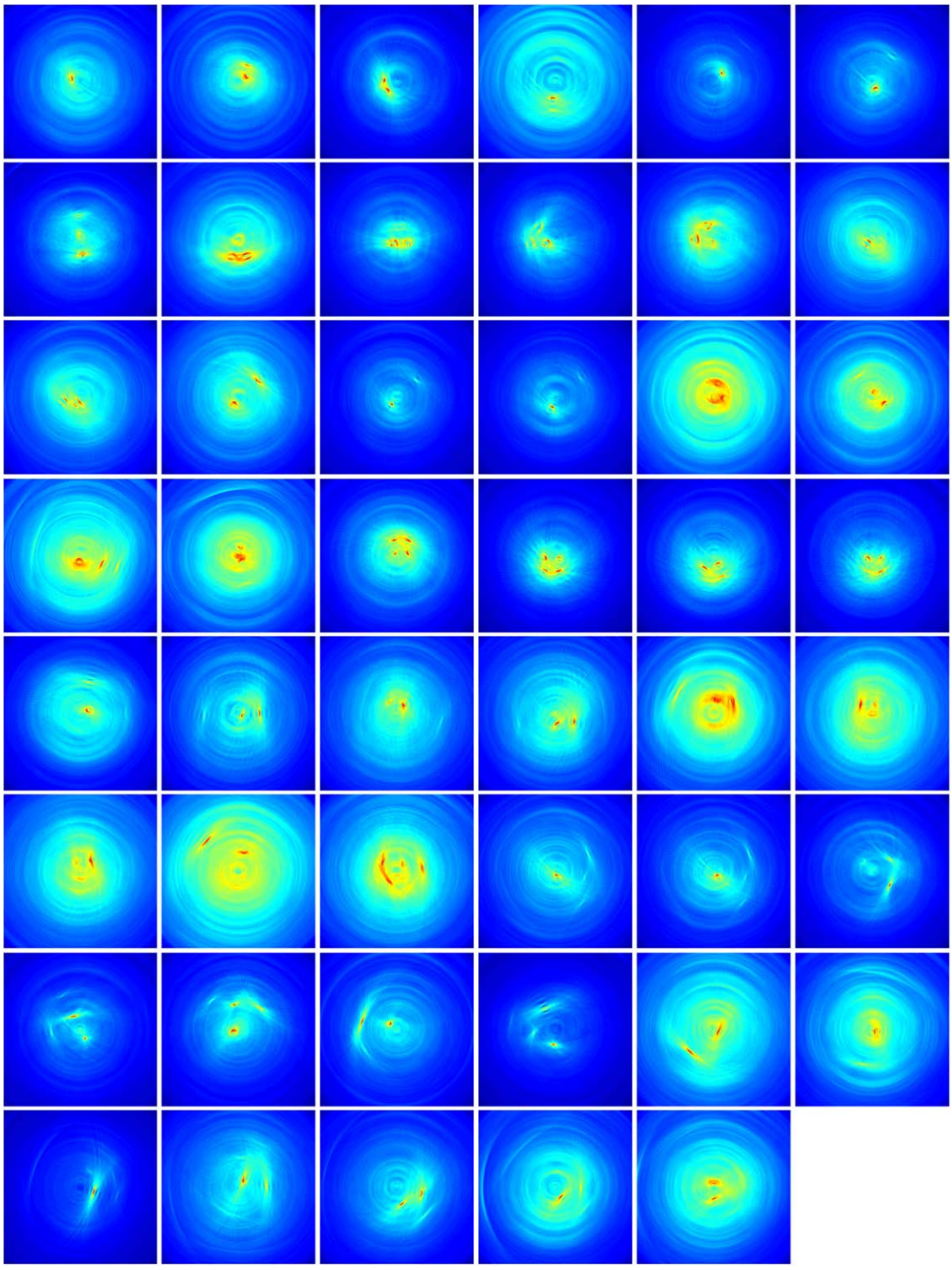
Appendix B. Datasets involved in this study

For x-ray-induced acoustic, 58 data were acquired with different target sizes and locations (as shown in figure B1). The dataset was randomly split into 47, 5, and 6 data for training, validation, and testing, respectively. As a result, the testing data are not seen during the model training nor validation.

For protoacoustic, five data were acquired with different proton beam coordinates (as shown in figure B2). Due to the limited dataset, one datum was randomly selected for validation while the other four data were used for training. As a result, the validation data are not seen during the model training. Figure B2 shows the reconstruction of the protoacoustic dataset involved in this study.

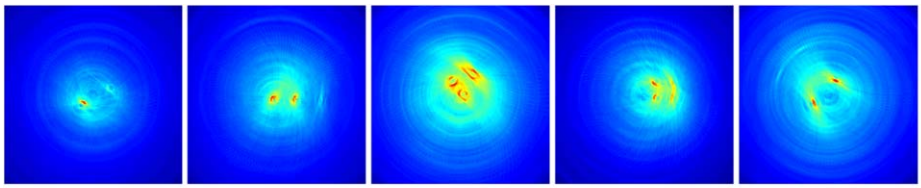
For electroacoustic, 52 data were acquired with different electrode arrangements and locations (as shown in figure B3). The dataset was randomly split into 43, 3, and 6 data for training, validation, and testing, respectively. As a result, the testing data are not seen during the model training nor validation. Figure B3 shows the reconstruction of the XACT dataset involved in this study.

Due to the limited experimental resources in our lab, imaging targets have simple structures. To increase the diversity among data, targets were designed with different sizes/shapes and at different positions relative to the transducer, resulting in markedly different response in the acquired signals. In addition, noises in the acquired acoustic signals are random. Therefore, the testing data can be considerably distinct from the training and validation data. In future studies, evaluation of more complex structures and materials is warranted.

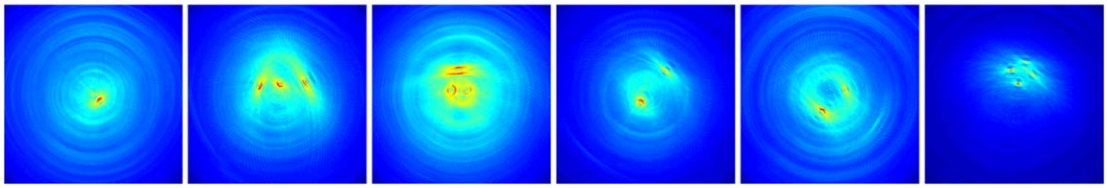


(a) training data

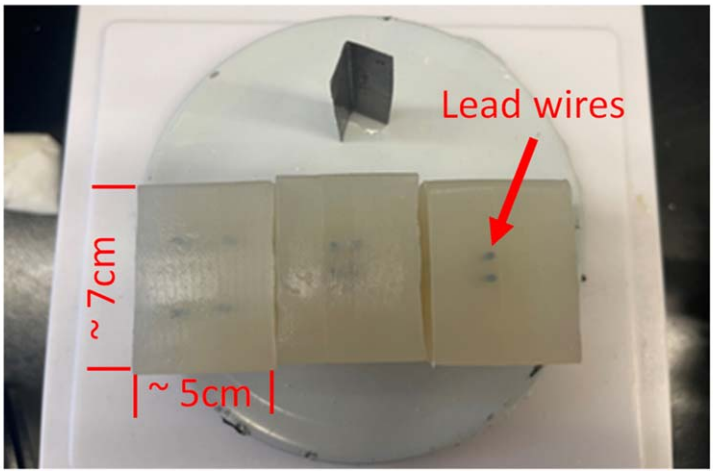
Figure B1. The reconstruction of the x-ray-induced acoustic data in the (a) training, (b) validation, and (c) testing dataset. Each subfigure in (a)–(c) shows the 2D image reconstructed from the fully-averaged signals of one datum. Different data were acquired with different target sizes and locations. Hotspots indicate the targets (lead wires) in the phantom. Images are shown in the 'jet' colormap (dynamic range is set to [0, max(image)*1.1]). (d) shows some example phantoms used in this study. The phantom size is about 5 cm wide, 7 cm long, and 4.5 cm high.



(b) validation data



(c) testing data



(d) example phantoms

Figure B1. (Continued.)

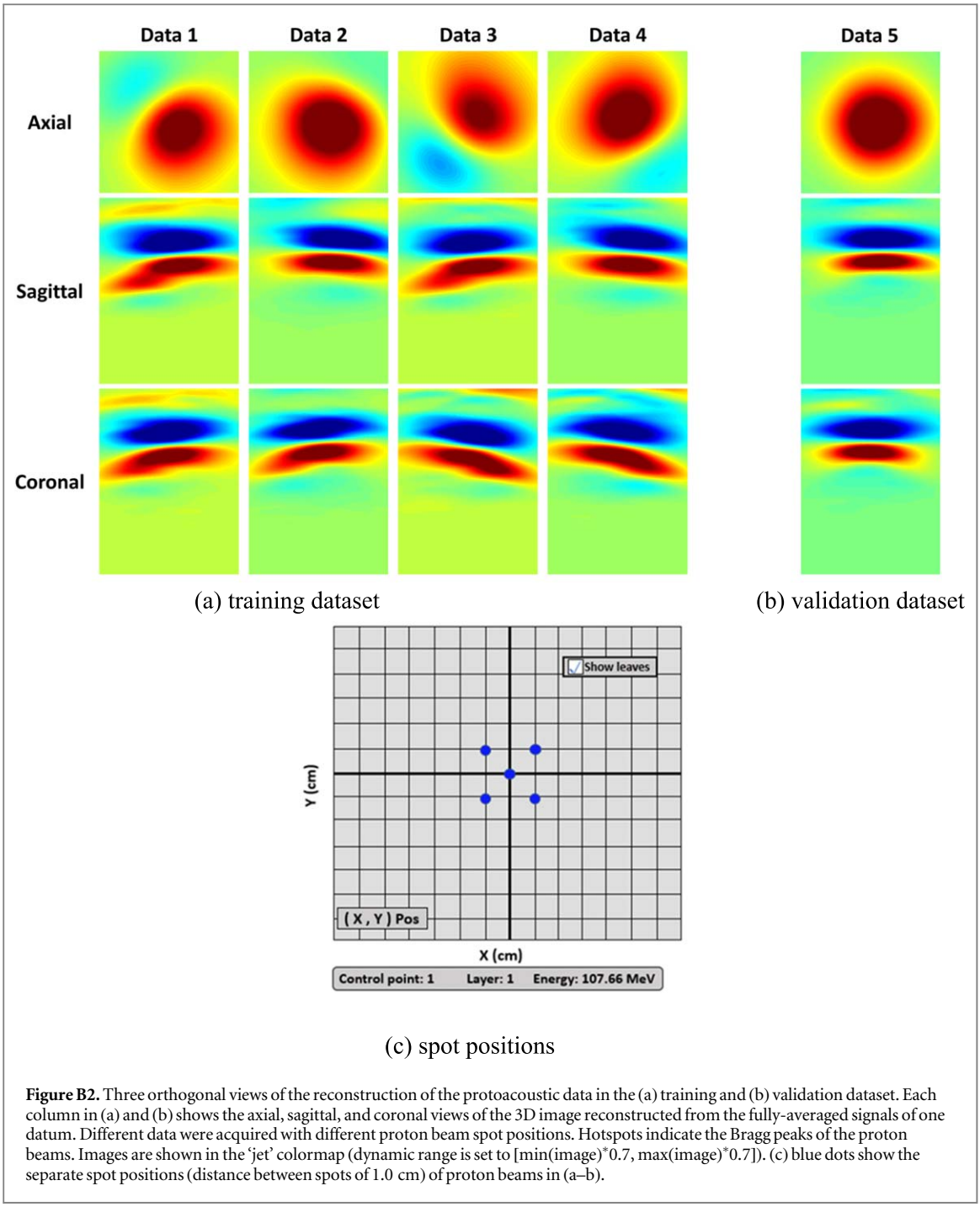
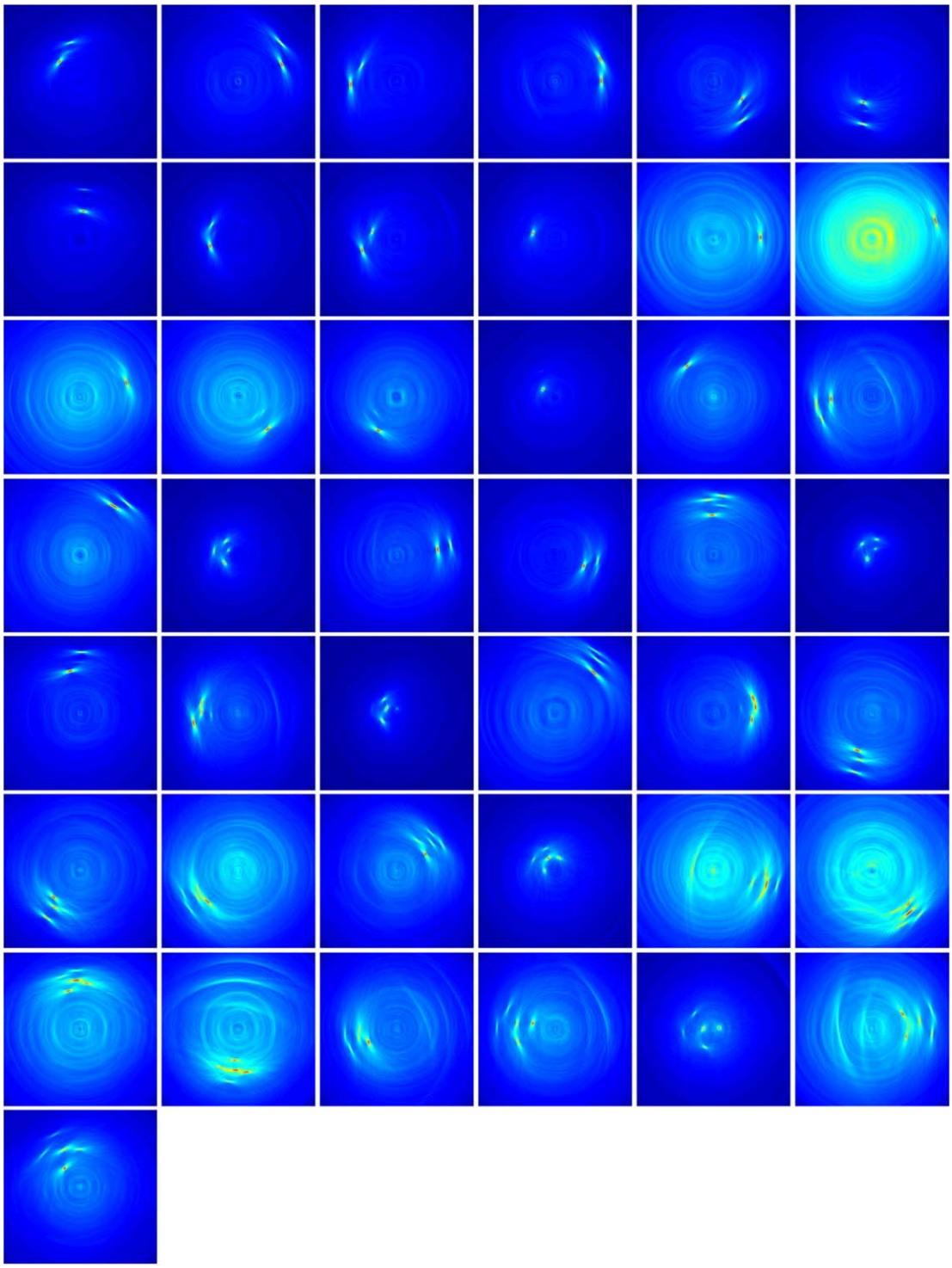


Figure B2. Three orthogonal views of the reconstruction of the protoacoustic data in the (a) training and (b) validation dataset. Each column in (a) and (b) shows the axial, sagittal, and coronal views of the 3D image reconstructed from the fully-averaged signals of one datum. Different data were acquired with different proton beam spot positions. Hotspots indicate the Bragg peaks of the proton beams. Images are shown in the 'jet' colormap (dynamic range is set to [min(image)*0.7, max(image)*0.7]). (c) blue dots show the separate spot positions (distance between spots of 1.0 cm) of proton beams in (a–b).



(a) training data

Figure B3. The reconstruction of the electroacoustic data in the (a) training, (b) validation, and (c) testing dataset. Each subfigure in (a)–(c) shows the 2D image reconstructed from the fully-averaged signals of one datum. Different data were acquired with different electrode arrangements and locations. Hotspots indicate the targets (lead wires) in the phantom. Images are shown in the 'jet' colormap (dynamic range is set to [0, max(image)*1.1]). (d) shows some example electrode arrangements. Electrodes were held by a 3D-printed multi-electrode holder (holes are arranged in a 3 × 3 layout and distance between adjacent holes is 5 mm). Experiment device for (d) can be referred to the figure 2(c) in the manuscript.

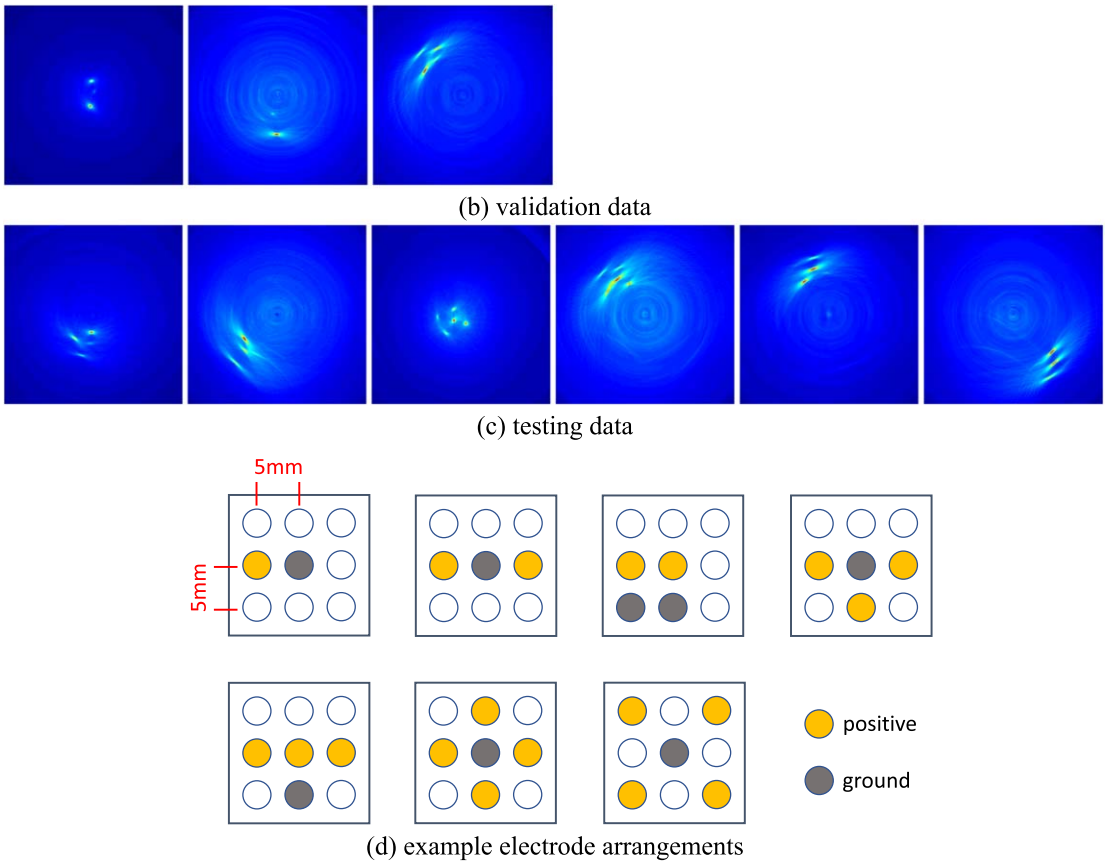


Figure B3. (Continued.)

Appendix C. Quantitative evaluation for ROIs in x-ray-induced and electroacoustic imaging

For x-ray-induced acoustic images, ROIs were defined using a threshold of 55% of the maximum values of the ground truth images. An example of the ROIs is shown in the following figure C1 in black solid lines. Quantitative results are shown in the following table C1, which are consistent with the whole-image evaluation.

For electroacoustic images, ROIs were defined using a threshold of 45% of the maximum values of the ground truth images. An example of the ROIs is shown in the following figure C2 in black solid lines.

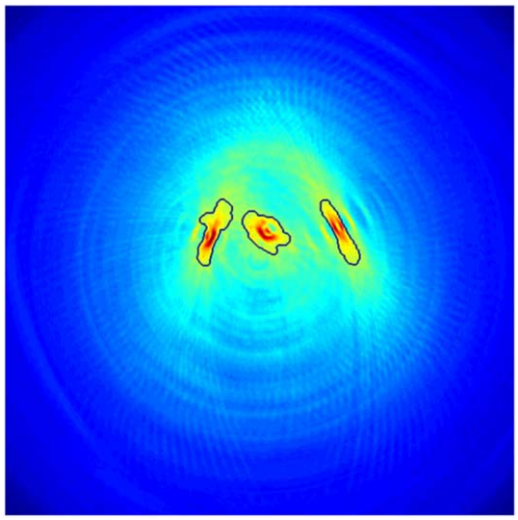


Figure C1. Regions of interest (ROIs) defined in the x-ray-induced acoustic images. ROIs are indicated by the black solid lines.

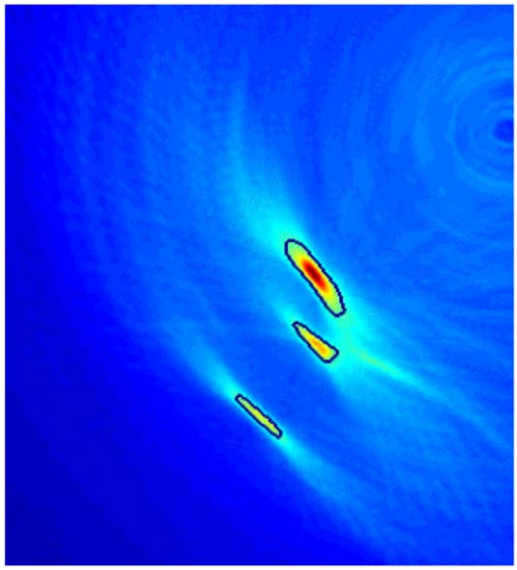


Figure C2. Regions of interest (ROIs) defined in the Electroacoustic images. ROIs are indicated by the black solid lines.

Table C1. Quantitative analysis of regions of interest (ROIs) in the x-ray-induced acoustic images.

Imaging techniques	Metrics	Values[a]			
		Filtered	U-Net (34 539 585)[b]	U-Net-Lite (1 180 081)[b]	**Proposed** (1 476 609)[b]
X-ray acoustic	MSE	0.0182 ± 0.0244	$\mathbf{0.0126 \pm 0.0283}$	0.0139 ± 0.0305	0.0165 ± 0.0355
	PSNR	20.498 ± 5.5069	$\mathbf{25.308 \pm 6.4666}$	25.104 ± 6.7423	23.733 ± 6.4593
	SSIM	0.8210 ± 0.1512	$\mathbf{0.9360 \pm 0.0775}$	0.9322 ± 0.0791	0.9310 ± 0.0845

[a] Values are calculated with data normalized to [0, 1], and expressed as mean ± std.
[b] Numbers in brackets indicate the total parameters of the model.

Table C2. Quantitative analysis of regions of interest (ROIs) in the Electroacoustic images.

Imaging Techniques	Metrics	Values[a]			
		Filtered	U-Net (34 539 585)[b]	U-Net-Lite (1 180 081)[b]	**Proposed** (1 476 609)[b]
Electroacoustic	MSE	0.0300 ± 0.0337	0.0014 ± 0.0012	0.0018 ± 0.0017	$\mathbf{0.0011 \pm 0.0008}$
	PSNR	17.185 ± 3.9856	30.196 ± 3.6893	29.022 ± 3.6921	$\mathbf{30.341 \pm 2.8499}$
	SSIM	0.8555 ± 0.1468	0.9897 ± 0.0095	0.9871 ± 0.0125	$\mathbf{0.9920 \pm 0.0061}$

[a] Values are calculated with data normalized to [0, 1], and expressed as mean ± std.
[b] Numbers in brackets indicate the total parameters of the model.

Quantitative results are shown in the following table C2, which are consistent with the whole-image evaluation. These results confirm GDI-CNN's superior performance to U-Net in terms of electrode shape accuracy.

References

Ahmad M *et al* 2015 Theoretical detection threshold of the proton-acoustic range verification technique *Med. Phys.* **42** 5735–44

Antholzer S, Haltmeier M and Schwab J 2019 Deep learning for photoacoustic tomography from sparse data *Inverse Prob. Sci. Eng.* **27** 987–1005

Awasthi N *et al* 2020 Sinogram super-resolution and denoising convolutional neural network (SRCN) for limited data photoacoustic tomography arXiv:2001.06434

Balakrishnan G *et al* 2018 An unsupervised learning model for deformable medical image registration[C] *Proc. of the IEEE Conf. on Computer Vision And Pattern Recognition* pp 9252–60

Chang Y *et al* 2021 A generative adversarial network (GAN)-based technique for synthesizing realistic respiratory motion in the extended cardiac-torso (XCAT) phantoms *Phys. Med. Biol.* **66** 115018

Deurvorst F R *et al* 2022 A spatial and temporal characterisation of single proton acoustic waves in proton beam cancer therapy *J. Acoust. Soc. Am.* **151** 1200–10

Donoho D L and Johnstone I M 1994 Ideal spatial adaptation by wavelet shrinkage *Biometrika* **81** 425–55
Ghafoorian M *et al* 2018 Student beats the teacher: deep neural networks for lateral ventricles segmentation in brain MR[C] *Medical Imaging 2018: Image Processing* (SPIE) 10574, pp 744–9
Gutta S *et al* 2017 Deep neural network-based bandwidth enhancement of photoacoustic data *J. Biomed. Opt.* **22** 116001
Hariri A *et al* 2020 Deep learning improves contrast in low-fluence photoacoustic imaging *Biomed. Opt. Express* **11** 3360–73
Hayakawa Y *et al* 1995 Acoustic pulse generated in a patient during treatment by pulsed proton radiation beam *Radiat. Oncol. Invest.* **3** 42–5
Hesamian M H *et al* 2019 Deep learning techniques for medical image segmentation: achievements and challenges *J. digit. Imaging* **32** 582–96
Hickling S *et al* 2018 Ionizing radiation-induced acoustics for radiotherapy and diagnostic radiology applications *Med. Phys.* **45** e707–21
Hickling S, Léger P, El and Naqa I 2014 Simulation and experimental detection of radiation-induced acoustic waves from a radiotherapy linear accelerator[C] *2014 IEEE Int. Ultrasonics Symp.* (IEEE) pp 29–32
Jiang Z *et al* 2019a Augmentation of CBCT reconstructed from under-sampled projections using deep learning *IEEE Trans. Med. Imaging* **38** 2705–15
Jiang Z *et al* 2020 A multi-scale framework with unsupervised joint training of convolutional neural networks for pulmonary deformable image registration *Phys. Med. Biol.* **65** 015011
Jiang Z *et al* 2021 Enhancing digital tomosynthesis (DTS) for lung radiotherapy guidance using patient-specific deep learning model *Phys. Med. Biol.* **66** 035009
Jiang Z *et al* 2022a Fast four-dimensional cone-beam computed tomography reconstruction using deformable convolutional networks *Med. Phys.* **49** 6461–76
Jiang Z *et al* 2022b 3D *in vivo* dose verification in prostate proton therapy with deep learning-based proton-acoustic imaging *Phys. Med. Biol.* **67** 215012
Jiang Z, Yin F and Ren L 2019b Student beats the teacher: deep learning using a 3d convolutional neural network (CNN) for augmentation of CBCT reconstructed from under-sampled projections[C] *MEDICAL PHYSICS. 111 RIVER ST, HOBOKEN 07030-5774, NJ USA* (WILEY) 46, E222–222
Jones K C *et al* 2014 Proton beam characterization by proton-induced acoustic emission: simulation studies *Phys. Med. Biol.* **59** 6549
Jones K C *et al* 2015 Experimental observation of acoustic emissions generated by a pulsed proton beam from a hospital-based clinical cyclotron *Med. Phys.* **42** 7090–7
Jones K C *et al* 2018 Acoustic-based proton range verification in heterogeneous tissue: simulation studies *Phys. Med. Biol.* **63** 025018
Kalunga R *et al* 2023 On the robustness of multilateration of ionoacoustic signals for localization of the Bragg peak at pre-clinical proton beam energies in water *Phys. Med. Biol.* **68** 105010
Kingma D P and Ba J 2014 Adam: a method for stochastic optimization arXiv:1412.6980
Kipergil E A *et al* 2017 An analysis of beam parameters on proton-acoustic waves through an analytic approach *Phys. Med. Biol.* **62** 4694
Kruger R A *et al* 1995 Photoacoustic ultrasound (PAUS)—reconstruction tomography *Med. Phys.* **22** 1605–9
Li Y *et al* 2020 3D x-ray-induced acoustic computed tomography with a spherical array: A simulation study on bone imaging *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **67** 1613–9
Mast T D *et al* 2023 Reconstruction of thermoacoustic emission sources induced by proton irradiation using numerical time reversal *Phys. Med. Biol.* **68** 025003
Najafzadeh E *et al* 2020 Photoacoustic image improvement based on a combination of sparse coding and filtering *J. Biomed. Opt.* **25** 106001
Neal R E *et al* 2014 *In vivo* irreversible electroporation kidney ablation: experimentally correlated numerical models *IEEE Trans. Biomed. Eng.* **62** 561–9
Ngui W K *et al* 2013 Wavelet analysis: mother wavelet selection methods *Appl. Mech. Mater.* **393** 953–8
Nie W *et al* 2018 Proton range verification in homogeneous materials through acoustic measurements *Phys. Med. Biol.* **63** 025036
Pandey P K *et al* 2021 Model-based x-ray-induced acoustic computed tomography *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **68** 3560–9
Patil P B and Chavan M S 2012 A wavelet based method for denoising of biomedical signal[C] *Int. Conf. on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012)* (IEEE) pp 278–83
Pogue B W *et al* 2021 Review of *in vivo* optical molecular imaging and sensing from x-ray excitation *J. Biomed. Opt.* **26** 010902
Robertson E *et al* 2020 X-ray-induced acoustic computed tomography (XACT): initial experiment on bone sample *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **68** 1073–80
Ronneberger O, Fischer P and Brox T 2015 U-net: Convolutional networks for biomedical image segmentation[C] *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th Int. Conf., Munich, Germany, October 5–9, 2015, Proc., Part III 18* (Springer Int. Publishing) 234–41
Samant P *et al* 2020 X-ray induced acoustic computed tomography *Photoacoustics* **19** 100177
Samant P *et al* 2022 3D protoacoustic imaging through a planar ultrasound array: a simulation workflow *IEEE Trans. Radiat. Plasma Med. Sci.* **7** 83–95
Shen K *et al* 2020 Negativity artifacts in back-projection based photoacoustic tomography *J. Phys. D: Appl. Phys.* **54** 074001
Szegedy C *et al* 2015 Going deeper with convolutions[C] *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 1–9
Tang S *et al* 2017 X-ray-induced acoustic computed tomography with an ultrasound transducer ring-array *Appl. Phys. Lett.* **110** 10
Van Dongen K W A *et al* 2019 Reconstructing 3D proton dose distribution using ionoacoustics *Phys. Med. Biol.* **64** 225005
Wang J *et al* 2023 Deep learning-based protoacoustic signal denoising for proton range verification *Biomed. Phys. Eng. Express* **9** 045006
Wang M *et al* 2019 Feasibility of electroacoustic tomography: a simulation study *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **67** 889–97
Wang S *et al* 2021a X-ray-induced acoustic computed tomography (XACT) imaging with single-shot nanosecond x-ray *Appl. Phys. Lett.* **119** 18
Wang S, Zarafshani A and Xiang L 2021a Electroacoustic tomography (EAT): 2D electric field reconstruction for electroporation treatment monitoring *Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling* (SPIE) 11598, 545–50
Wang Y *et al* 2004 Photoacoustic imaging with deconvolution algorithm *Phys. Med. Biol.* **49** 3117
Wang Z, Zhou Y and Hu S 2021b Sparse coding-enabled low-fluence multi-parametric photoacoustic microscopy *IEEE Trans. Med. Imaging* **41** 805–14
Würfl T *et al* 2018 Deep learning computed tomography: Learning projection-domain weights from image domain in limited angle problems *IEEE Trans. Med. Imaging* **37** 1454–63
Xiang L *et al* 2016 High resolution x-ray-induced acoustic tomography *Sci. Rep.* **6** 26118
Yao S *et al* 2020 Feasibility study of range verification based on proton-induced acoustic signals and recurrent neural network *Phys. Med. Biol.* **65** 215017

Yao S *et al* 2021 Further investigation of 3D dose verification in proton therapy utilizing acoustic signal, wavelet decomposition and machine learning *Biomed. Phys. Eng. Express* **8** 015008

You K and Choi H 2020 Inter-stage output voltage amplitude improvement circuit integrated with class-b transmit voltage amplifier for mobile ultrasound machines *Sensors* **20** 6244

Zarafshani A *et al* 2019a Electroacoustic tomography (EAT): linear scanning with a single element transducer[C] *Medical Imaging 2019: Ultrasonic Imaging and Tomography* (SPIE) 10955, 237–42

Zarafshani A, Zheng B and Xiang L 2019b Electroacoustic tomography system using ultra-short electric filed excitation source induced acoustic signals *J. Acoust. Soc. Am.* **145** 1920–1

Zhu B *et al* 2018 Image reconstruction by domain-transform manifold learning *Nature* **555** 487–92