Deep-Learning-Based Electrical Noise Removal Enables High Spectral Optoacoustic Contrast in Deep Tissue

Christoph Dehner, Ivan Olefir, Kaushik Basak Chowdhury, Dominik Jüstel, and Vasilis Ntziachristos

***Abstract*— Image contrast in multispectral optoacoustic tomography (MSOT) can be severely reduced by electrical noise and interference in the acquired optoacoustic signals. Previously employed signal processing techniques have proven insufficient to remove the effects of electrical noise because they typically rely on simplified models and fail to capture complex characteristics of signal and noise. Moreover, they often involve time-consuming processing steps that are unsuited for real-time imaging applications. In this work, we develop and demonstrate a discriminative deep learning approach to separate electrical noise from optoacoustic signals prior to image reconstruction. The proposed deep learning algorithm is based on two key features. First, it learns spatiotemporal correlations in both noise and signal by using the entire optoacoustic sinogram as input. Second, it employs training on a large dataset of experimentally acquired pure noise and synthetic optoacoustic signals. We validated the ability of the trained model to accurately remove electrical noise on synthetic data and on optoacoustic images of a phantom and the human breast. We demonstrate significant enhancements of morphological and spectral optoacoustic images reaching 19% higher blood vessel contrast and localized spectral contrast at depths of more than 2 cm for images acquired in vivo. We discuss how the proposed denoising framework is applicable to clinical multispectral optoacoustic tomography and suitable for real-time operation.**

***Index Terms*— Breast cancer, denoising, dynamic MSOT, photoacoustic tomography, signal decomposition, sinogram.**

I. Introduction

ELECTRICAL noise is a key source of signal corruption in optoacoustic imaging and arises due to thermal effects (thermal noise) and from electromagnetic interference (parasitic noise); the latter possibly generated by the optoacoustic system itself or the environment [1]. While thermal noise can be modeled as white Gaussian noise [2], parasitic noise entails complex spatiotemporal correlations, and thus cannot be efficiently captured by an analytical model [1]. Both thermal and parasitic noise cause artifacts in reconstructed optoacoustic images that severely decrease morphological and spectral contrast. Whereas shielding hardware can suppress some parasitic noise, this solution is device-specific and often incomplete [1]. Therefore, additional computational denoising techniques, which are applicable across platforms, are needed to remove both parasitic and thermal noise.

Noise in optoacoustic images hinders the detection and identification of fine structures in tissue, particularly as the signal-to-noise ratio (SNR) in the data decreases with increasing depth [3]. Besides the reduction of image contrast, noise also challenges the quantification and spectral un-mixing of optoacoustic images acquired at multiple wavelengths [4], [5]. In particular, corrupted spectral information decreases the spatial resolution of multispectral optoacoustic tomography (MSOT) because it necessitates averaging the spectra obtained from large tissue regions for reliable quantification [6]–[8]. Efficient noise reduction algorithms are therefore critical for improving the performance of spectral optoacoustics.

Frequency filtering using band-pass filters cannot adequately separate thermal and parasitic noise from optoacoustic signals because the frequency content of optoacoustic signals and noise overlap significantly. For this reason, data averaging and regularization methods have been more commonly considered to minimize the effects of electrical noise from optoacoustic tomographic images [3], [7]–[12]. While data averaging effectively reduces zero-mean electrical noise, combining multiple

Manuscript received 29 April 2022; accepted 30 May 2022. Date of publication 3 June 2022; date of current version 27 October 2022. This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 862811 (RSENSE), from the European Research Council (ERC) under Grant Agreement No. 694968 (PREMSOT), and by the Deutsche Forschungsgemeinschaft (DFG), Sonderforschungsbereich-824 (SFB-824), subproject A1. *(Corresponding authors: Vasilis Ntziachristos; Dominik Jüstel.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Local Ethics Committee of the Technical University of Munich under Application No. 27/18 S.

Christoph Dehner, Ivan Olefir, Kaushik Basak Chowdhury, and Vasilis Ntziachristos are with the Institute of Biological and Medical Imaging, Helmholtz Zentrum Munchen, 85764 Neuherberg, Germany, and also with the School of Medicine, Chair of Biological Imaging, Technical University of Munich, 80333 München, Germany (e-mail: christoph.dehner@helmholtz-muenchen.de; ivan.olefir@tum.de; kaushikbchowdhury@gmail.com; bioimaging.translatum@tum.de).

Dominik Jüstel is with the Institute of Biological and Medical Imaging and the Institute of Computational Biology, Helmholtz Zentrum Munchen, 85764 Neuherberg, Germany, and also with the School of Medicine, Chair of Biological Imaging, Technical University of Munich, 80333 München, Germany (e-mail: dominik.juestel@helmholtz-muenchen.de).

Digital Object Identifier 10.1109/TMI.2022.3180115

acquisitions reduces imaging rates and increases vulnerability to motion artifacts, particularly in clinical applications or when using a handheld system. Regularization of model-based reconstructions may decrease the effects of electrical noise, but this reduction is either limited to the noise characteristics captured by the regularization functional or realized at the cost of data fidelity, thereby corrupting the meaningfulness of the reconstructed image. Furthermore, iterative model-based reconstruction is computationally intensive, and therefore unsuitable for applications that require real-time feedback [4], [13], [14]. Another approach to reduce noise is based on sparse (typically Wavelet-based) representations for optoacoustic signals [1], [15], [16]. Noise is assumed to distort the sparsity properties of optoacoustic signals, which enables its removal, e.g., via thresholding techniques. However, the denoising performance of such methods is limited by their reliance on oversimplified models of noise and optoacoustic signals.

Recently, discriminative deep neural network models have achieved state-of-the-art performance on general image denoising tasks, like Gaussian denoising, deblocking, super-resolution, inpainting, and dehazing [17]–[20]. These deep neural network models capture the required data transformations for denoising in a data-driven way by utilizing large ground truth training datasets. In this way, discriminative deep neural networks are capable of more accurate, robust, and fast denoising than traditional methods that rely on rigid analytical models because they can adjust to complex data characteristics during training and are evaluated in real-time on modern GPUs. Similar deep-learning-based approaches have also been applied to remove reconstruction [21], [22] and reflection artifacts [23], [24] from optoacoustic images, and to enhance contrast of images acquired with low energy illumination elements such as LED-based systems [25]–[27].

In this work, we examine whether discriminative deep learning can separate thermal and parasitic noise from optoacoustic signals. We show that the modeling capabilities and the computational efficiency of a deep neural network facilitates denoising in optoacoustic tomography that is both precise enough to remove noise with complex characteristics and fast enough for real-time imaging applications. We design a deep neural network model to simultaneously denoise the entire sinogram of an optoacoustic scan, i.e. the complete dataset acquired from all transducers. Entire sinogram denoising enables the network to capture spatiotemporal correlations within both parasitic noise and true optoacoustic responses, and thus more efficiently separate the two. Exploiting the independence of electrical noise and optoacoustic signals, we train the network on a large ground truth dataset of experimentally acquired pure noise and synthetic optoacoustic sinograms. We validate that the model removes thermal and parasitic noise from both synthetic sinograms and optoacoustic images of a phantom. We lastly apply the trained model to clinical MSOT images of breast tissue and show significant enhancements in morphological and spectral contrast. The improved contrast allows for tissue components to be more accurately localized and quantified and yields more meaningful correlations with the spectra of known absorbers in tissue, thereby increasing access to endogenous biomarkers in deep tissue, such as breast vasculature or hemoglobin contrast inside a cancerous tumor.

II. Methods

In the following, we formalize our methodology for removing electrical noise from optoacoustic sinograms. First, we reformulate the denoising problem as a decomposition task. Based on this formulation, we derive a discriminative deep learning framework for denoising optoacoustic sinograms. At the end of the chapter, we explain the experiments that we use to validate this approach.

A. Denoising via Decomposition

In this section, we formalize the rationale for reformulating denoising of optoacoustic sinograms as a decomposition task and conclude that 1) electrical noise in optoacoustic tomography is an additive component that is independent from the optoacoustic signals, and 2) for denoising, an acquired optoacoustic sinogram s can be decomposed into a component s_{OA} containing the true optoacoustic sinogram and an electrical noise component s_{noise}: $s = s_{\mathrm{OA}} + s_{\mathrm{noise}}$.

An optoacoustic scan at a fixed excitation wavelength consists of measured optoacoustic pressure signals $s_d[t]$, indexed by time samples $t \in [1, 2, \ldots, N_{\mathrm{samples}}]$ and transducer locations $d \in [1, 2, \ldots, N_{\mathrm{transducers}}]$. The collection of signals recorded at all the transducers compose the sinogram $s[d, t] := s_d[t]$ of the scan. We model the measured optoacoustic sinograms probabilistically as samples s of a random field S, i.e. a collection of random variables that model all recorded signals of a sinogram $s_d[t]$, $t \in [1, 2, \ldots, N_{\mathrm{samples}}]$, $d \in [1, 2, \ldots, N_{\mathrm{transducers}}]$. In the remainder of this paper, we will denote random fields with capital letters, and specific samples of random fields with lower case letters. The main assumption that leads to the formulation of denoising as a decomposition problem is that S is the sum of two independent random fields S_{OA} and S_{noise}, which describe the signal content due to optoacoustic responses and electrical noise, respectively. This assumption is justified by the fact that electrical noise in optoacoustic tomography typically originates from common system thermal noise and electromagnetic interference that is not influenced by the optoacoustic signal [1]. The probability distributions underlying S_{OA} and S_{noise} are denoted by P_{OA} and P_{noise}:

$$\begin{aligned} &S = S_{\mathrm{OA}} + S_{\mathrm{noise}} \\ &\text{with } S_{\mathrm{OA}} \sim P_{\mathrm{OA}} \text{ and } S_{\mathrm{noise}} \sim P_{\mathrm{noise}} \text{ independent.} \end{aligned} \tag{1}$$

Optoacoustic scans at different wavelengths are modeled as independent realizations of S because the noise is caused by the electronics of the imaging system that are not affected by the wavelength switching of the laser. In summary, isolating the noise-free optoacoustic sinogram s_{OA} given s is equivalent to decomposing $s = s_{\mathrm{OA}} + s_{\mathrm{noise}}$ into its two components s_{OA} and s_{noise}.

B. Deep-Learning-Based Denoising Framework

Next, we step-by-step derive a deep-learning-based denoising framework that can decompose optoacoustic sinograms

into their signal and noise components. In short, we train a deep neural network to infer the electrical noise component from a noisy input sinogram using experimentally acquired pure noise sinograms and synthetic optoacoustic sinograms.

To solve the decomposition problem introduced in subsection A, we need access to the distributions P_{OA} and P_{noise}. However, both random fields S_{OA} and S_{noise} are non-homogeneous and anisotropic with intricate spatial and temporal correlations due to the physics underlying the signals and the fact that electrical noise in optoacoustic systems often contains complex parasitic noise (Fig. 2c) [1]. As a result, accurate explicit models for the complex distributions of optoacoustic signals and electrical noise P_{OA} and P_{noise} are difficult to obtain. We therefore present a data-driven approach that allows us to rely on the empirical distributions of P_{OA} and P_{noise} via sampling of S_{OA} and S_{noise}. We first explain sample acquisition and then elaborate on our methodology for solving the decomposition task.

Because of the independence of S_{OA} and S_{noise} and the wavelength independence of S_{noise}, electrical noise can be directly measured in the absence of any absorbers that would emit optoacoustic responses. We thus obtained samples of the electrical noise distribution S_{noise} of the test system by immersing the scanner in a water tank and measuring from 700 to 790 nm, where light absorption in water is negligible (i.e., the optical absorption coefficients of water at 700 – 790 nm are between 0.006 cm^{-1} and 0.026 cm^{-1}, whereas for example the absorption coefficients of oxygenated hemoglobin and fat at 800 nm and 930 nm are 4.4 cm^{-1} and 13 cm^{-1}, respectively [28]–[30]). Note that in theory the effects of water absorption on the acquired noise samples may be further decreased by interrupting the optical path between the laser and the imaging probe during the noise acquisitions, or by acquiring noise samples only at the wavelength with the lowest water absorption coefficient from the range 700 – 790 nm. However, in practice such adjustments were not essential, as the obtained electrical noise samples did not correlate with the respectively used wavelength from the range 700 – 790 nm (which confirms that water absorption is indeed negligible in the whole range), and as the presented denoising framework facilitated accurate denoising with the current noise acquisition setup.

Acquiring samples of S_{OA} in an experimental setup is a laborious and time-consuming process that requires averaging multiple scans of the same location to remove electrical noise. Additionally, patient or operator movement impedes accurate averaging. Therefore, instead of experimentally acquiring noise-free optoacoustic sinograms, we generated samples of P_{OA} via simulation by applying an accurate acoustic forward model of the scanner [31], [32] to publicly available images from the PASCAL VOC2012 dataset [33], a diverse collection of over 17 000 images covering a large range of features. Utilizing these images as underlying initial pressure distributions in the simulations enables us to account for a broad range of potential features in optoacoustic sinograms and should yield a good approximation of the empirical distribution of P_{OA}. In addition, the general scope of the training data ensures that the denoising model is universally and with uniform performance applicable to arbitrary scans acquired by the considered system.

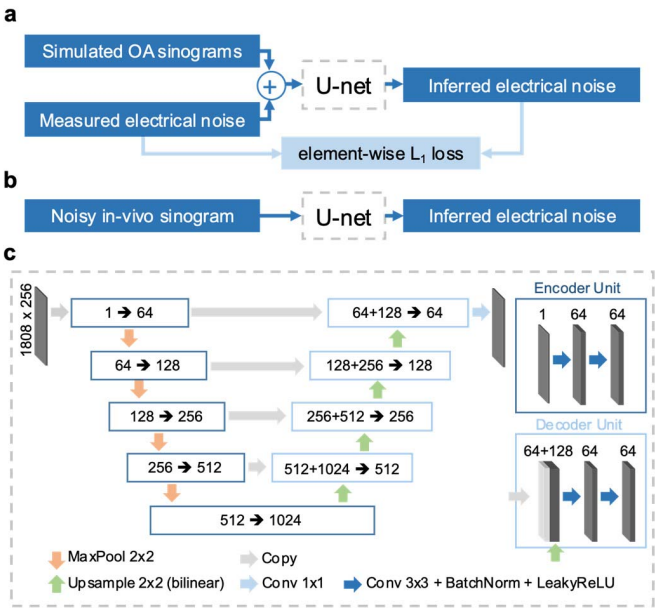

Fig. 1. Discriminative deep learning framework for denoising optoacoustic sinograms. a) Training setup of the method. The network was trained iteratively using simulated samples from the optoacoustic sinogram distribution of the test system and experimentally acquired samples of the electrical noise distribution of the system. b) Evaluation setup of the method. The trained neural network can infer the electrical noise from a noisy input sinogram. Subtracting the inferred noise from the input sinogram yields the denoised sinogram. c) U-net architecture of the deep neural network.

Using the samples from S_{OA} and S_{noise}, we utilized a U-net-like deep neural network [34] to solve the decomposition task at hand. Fig. 1 depicts the deep-learning-based approach. Fig. 1a shows the training setup. We iteratively trained the network on randomly selected pairs of samples s_{OA} and s_{noise} from S_{OA} and S_{noise} to infer the noise component (s_{noise}) from a noisy input sinogram $s = s_{\mathrm{OA}} + s_{\mathrm{noise}}$. In this way, the network is optimized to adopt the complex characteristics of P_{OA} and P_{noise}. Fig. 1b depicts the evaluation setup. Once trained, we can use the neural network to infer electrical noise from noisy input scans.

C. Experiments

As a test system for the denoising algorithm, we used a custom prototype of an MSOT Acuity Echo handheld scanner (iThera Medical GmbH, Munich, Germany). The system is equipped with a tunable laser that illuminates tissue with laser pulses of ~8 ns duration with an energy of 16 mJ and a repetition rate of 25 Hz. The ultrasound detector of the system consists of 256 piezoelectric transducers with a central frequency of 4 MHz, which are equidistantly placed on a circular arc with a radius of 6 cm and an angular coverage of 145°. Ultrasound signals are recorded with a sampling frequency of 40 MHz. Inside the imaging probe, heavy water with a speed of sound of approximately 1397 m/s is used as coupling medium. Fig. 2 provides an overview of the imaging system and its output. Fig. 2a shows the

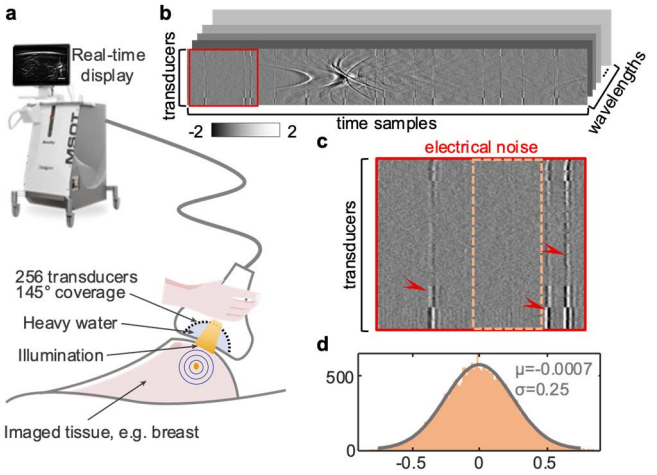

Fig. 2. Overview of the handheld MSOT system and its output used to evaluate the proposed denoising method. a) Illustration of the scanning procedure using the handheld imaging probe of the test system. b) Data layout of a measured multispectral stack of sinograms. The depicted sinogram shows the recorded signals during a representative scan of a human breast lesion at 960 nm. c) Magnification of the marked signals in b, which were recorded prior to responses from tissue and thus are predominately comprised of electrical noise. d) Histogram and fitted Gaussian distribution ($R^2 = 99.5\%$) for parts of the electrical noise with visually low amounts of parasitic noise (signals marked in c with the dashed rectangle) illustrating the characterization of the thermal noise of the system.

scanning procedure using the handheld imaging probe of the system. Fig. 2b illustrates the data layout of a multispectral stack acquired by the imaging system. A multispectral stack consists of 28 sinograms recorded at wavelengths from 700 – 970 nm in steps of 10 nm. Fig. 2c shows electrical noise from a representative optoacoustic in vivo scan. We observed that electrical noise in the system consists of two additive components: normally distributed thermal noise with a mean of zero and a standard deviation in the range of 0.2 – 0.3 and complex parasitic noise that is presumably caused by the switching-mode power supply of the system (examples marked with red arrows in Fig. 2c). Fig. 2d illustrates the characterization of thermal noise of the system based on parts of experimentally acquired electrical noise samples with visually low amounts of electrical noise and confirms that the thermal noise can be modeled as white Gaussian noise. The presented characterization of the thermal noise of the system nevertheless remains an approximation as the parasitic and thermal components of electrical noise cannot be completely separated.

We first evaluated the ability of the proposed deep learning framework to remove the combination of Gaussian thermal and complex parasitic electrical noise observed in the test system. We trained and evaluated a deep neural network on experimentally acquired samples of the electrical noise distribution S_{noise} of the system and simulated samples of the optoacoustic signal distribution S_{OA} (denoted as Dataset-EN). Next, we applied the trained denoising network to measurements of a physical phantom (denoted as Dataset-Ph). The arrangement of the phantom is shown in the inlay in Fig. 4a. Two plastic tubes with inner diameters of 3.0 mm and 0.86 mm and outer diameters of 3.2 mm and 1.52 mm were filled with ink and imaged cross-sectionally to simulate absorbers of different sizes and at different depths. These tubes were immersed into two layers of agar of slightly different densities mixed with Intralipid (6 ml 20% emulsion / 100 ml water) to mimic light scattering and small variations of the speed of sound distribution in biological tissue. Additionally, a copper plate was integrated into the arrangement as a reference that can be seen in both optoacoustic and ultrasound images.

To evaluate the denoising performance of the framework on in vivo scans, we subsequently applied the trained deep neural network to 81 multispectral optoacoustic scans of human breast cancer lesions (denoted as Dataset-BC). These scans were acquired in a study that was approved by the local ethics committee of the Technical University of Munich (27/18 S). All participants gave written informed consent upon recruitment.

Lastly, we evaluated the ability of the trained network to adapt to changes of the hardware configuration and of environmental conditions such as humidity and temperature of the used system, which might alter the amounts of thermal or parasitic electrical noise. For that, we applied the trained model to optoacoustic signals that were corrupted by a combination of measured electrical noise sinograms scaled with a $\text{factor}_{\text{EN}} \in \{0, 0.5, 1, \ldots, 3\}$, and white Gaussian noise with standard deviation $\sigma_{\text{GN}} \in \{0, 0.2, 0.4, \ldots, 2\}$ (denoted as Dataset-EN+GN). A summary of the four datasets is given in Table I.

D. Data Pre-Processing and Network Training

We band-pass filtered all recorded signals from 500 kHz – 10 MHz to remove signals outside the bandwidth of the transducers and reduce low frequency responses that would otherwise dominate the contrast in reconstructed optoacoustic images. Additionally, all signals were slightly cropped in the time domain to remove filtering artifacts at the signal boundaries and to make the number of signal samples divisible by 16, as required by the chosen neural network architecture, leading to 1808 time samples for each of the 256 detectors per scan.

A detailed illustration of the proposed neural network is given in Fig. 1c. We adopted the U-Net neural network architecture [34] with a depth of 5 layers and a width of 64 channels, and designed the network to infer only the electrical noise s_{noise} from a noisy input sinogram s to minimize the necessary expressiveness of the network [18]. The network was trained for 300 epochs using the L1 norm of the difference of inferred and ground truth noise as loss functional, and the ADAM optimizer [35] with batch size $= 1$ and momentum parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The learning rate was set to 0.0001 and was linearly decreased to zero in the last 50 epochs. To accelerate the learning process, we scaled all input data of the neural network by a constant factor of 0.004 to achieve a signal range of $[-1; 1]$. After having passed the neural network, all signals were rescaled to the original range. After training, the final model was selected based on the minimal loss on a validation split of the dataset. One training

TABLE I
STRUCTURE AND SIZE OF THE FOUR DATASETS USED IN THIS PAPER

Name	Description	Train split	Val. split	Test split
Dataset-EN	Simulated noise-free optoacoustic sinograms	3000 sinograms	590 sinograms	629 sinograms
	Measurements of pure electrical noise sinograms	2110 sinograms	590 sinograms	629 sinograms
Dataset-Ph	Scans of a phantom	-	-	28 sinograms in 1 multispectral stack
Dataset-BC	In vivo scans of human breast lesions	-	-	2268 sinograms in 81 multispectral stacks
Dataset-EN+GN	Simulated noise-free optoacoustic sinograms	-	-	629 sinograms
	Measurements of pure electrical noise sinograms + simulated white Gaussian noise	-	-	(7 x 11) x 629 sinograms of measured electrical noise sinograms that were scaled with factor$_{EN} \in \{0, 0.5, 1, \ldots, 3\}$ and additionally augmented by additive white Gaussian noise with a standard deviation $\sigma_{GN} \in \{0, 0.2, 0.4, \ldots, 2\}$

process took approximately three days on an NVIDIA GeForce RTX2070 GPU.

E. Signal-to-Noise Ratio

For Dataset-EN and Dataset-GN, we used the signal-to-noise ratio (SNR), i.e. the ratio of signal power and noise power, to quantify the noise levels in the signals before and after denoising. We calculated the power $P(s)$ of a whole sinogram $s[d,t], d \in [1,2,\ldots,N_{\text{transducers}}], t \in [1,2,\ldots,N_{\text{samples}}]$ as

$$P(s) := \frac{1}{N_{\text{transducers}} N_{\text{samples}}} \sum_{d=1}^{N_{\text{transdicers}}} \sum_{t=1}^{N_{\text{samples}}} s[d,t]^2. \tag{2}$$

Based on equation 2, the SNR of a sinogram s with ground truth noise s_{noise} and inferred noise s'_{noise} is defined as

$$\text{SNR} := 10 \log_{10} \left(\frac{P\left(s - s_{\text{noise}}\right)}{P\left(s_{\text{noise}} - s'_{\text{noise}}\right)} \right) dB. \tag{3}$$

Setting $s'_{\text{noise}} = 0$ yields the SNR before denoising; setting s'_{noise} to the output of the trained network yields the SNR after denoising.

Since computing the SNR requires direct access to the ground truth noise s_{noise} of a scan, the metric cannot be directly transferred to the in vivo scans of Dataset-BC. We therefore defined an alternative metric, SNR_{mean}, that enabled us to approximate the ground truth electrical noise for the in vivo scans from Dataset-BC by considering the per-pixel mean sinogram amplitudes across N_{scans} different scans of the dataset, $s^{(1)}, s^{(2)}, \ldots, s^{(N_{\text{scans}})}$, $\langle |s| \rangle := \frac{1}{N_{\text{scans}}} \sum_{n=1}^{N_{\text{scans}}} |s^{(n)}|$.

$$\text{SNR}_{\text{mean}} := 10 \log_{10} \left(\frac{P_{\langle |s| \rangle - \langle |s_{\text{noise}}| \rangle}}{P_{\langle |s_{\text{noise}}| \rangle - \langle |s'_{\text{noise}}| \rangle}} \right) dB. \tag{4}$$

We approximated the ground truth electrical noise $\langle |s_{\text{noise}}| \rangle$ from the first 100 averaged time samples of all scans in Dataset-BC $\langle |s[\cdot, t']| \rangle, t' \in [1,2,\ldots,100]$:

$$\langle |s_{\text{noise}}[d,t]| \rangle \approx \frac{1}{100} \sum_{t'=1}^{100} \langle |s[d,t']| \rangle \quad \text{for } t \in [1,2,\ldots,N_{\text{samples}}]. \tag{5}$$

Note that equation 5 yields a meaningful approximation of $\langle |s_{\text{noise}}[d,t]| \rangle$ for two reasons: First, the 100 signals recorded at the beginning of a scan do not contain optoacoustic responses but mostly electrical noise because they originate from the coupling medium inside the imaging probe. Second, we observed from the electrical noise sinograms in Dataset-EN that for the used test system, $\langle |s_{\text{noise}}[d,t]| \rangle$ is constant over time. Thus, an estimation of electrical noise based on a subset of time samples of the sinogram (i.e. the first 100 time samples in equation 4) is applicable to all time steps $t \in [1,2,\ldots,N_{\text{samples}}]$. Furthermore, sinograms from Dataset-BC were cropped to the first 1732 recorded signal samples before evaluating the SNR_{mean}, as subsequent signals originate from outside the designated field of view of the scans and contain strong reflections and filtering artifacts.

SNR_{mean} can approximate the SNR of in vivo images, for which the true amount of electrical noise is unknown. However, the metric may overestimate the SNR after denoising because the per-pixel mean amplitudes of the predicted noise $\langle |s'_{\text{noise}}| \rangle$ is subtracted in the denominator of equation 4, disregarding the possibility that the predicted and the ground truth noise have different signs: $\langle |s_{\text{noise}}| \rangle - \langle |s'_{\text{noise}}| \rangle = \langle |s_{\text{noise}} - s'_{\text{noise}}| \rangle$ only if $\forall 1 \ldots n \ldots N_{\text{scans}}: \text{sgn}\left(s_{\text{noise}}^{(n)'}\right) = \text{sgn}\left(s_n^{(n)}\right)$. An empirical comparison of SNR_{mean} and SNR on Dataset-EN (for which ground truth noise samples are also available) showed that SNR_{mean} correctly estimated the average true SNR before denoising (SNR_{mean} = 9.6 dB, avg. SNR = 9.3 dB) and overestimated the SNR after denoising by approximately 6 dB (SNR_{mean} = 26.5 dB, avg. SNR = 20.3 dB). The offset of the SNR_{mean} after denoising is however smaller than the reported SNR_{mean} improvements of

20 – 22 dB of the presented denoising method for Dataset-BC (see section III).

F. Evaluation on Reconstructed Images

To evaluate the effects of the denoising method visually and quantitatively on optoacoustic images, we reconstructed the initial pressure p_0 of all breast scans in Dataset-BC, both with and without denoising the recorded sinograms with the trained neural network, using a model-based reconstruction algorithm [31], [32]. We added two regularization terms to address the two main causes of the ill-posedness of the inverse problem: simple Tikhonov regularization to mitigate limited view noise and Laplacian-based regularization to mitigate sub-resolution noise.

$$p_0 := \arg\min_{p\geq 0} \|Mp - s\|_2^2 + \lambda_1 \|p\|_2^2 + \lambda_2 \|\Delta p\|_2^2. \quad (6)$$

The reconstructed optoacoustic images are of the size 400 × 400 pixels and correspond to FOVs of 3.99 cm × 3.99 cm. We denote the obtained datasets of reconstructed MSOT breast images as D_{original} and D_{denoised}. Additional reconstructions were also obtained using backprojection reconstruction [36], [37].

We quantified the effects of the denoising method in the reconstructed images by calculating the contrast resolution and the contrast-to-noise ratio of blood vessels. For that, we manually segmented blood vessels in the images, as well as background ROIs from the surroundings of all segmented vessels. The segmentations were based on scans at 870 nm, where blood contrast is at a maximum. The background areas were chosen so as not to overlap with regions below and above strong absorbers, which are affected by limited view artifacts. We chose vessels of different sizes and at different depths to obtain general estimates for the blood contrast in the dataset. Fig. 3f shows the segmented regions for a representative scan. Based on these segmentations, the contrast resolution of a scan with mean intensities I_{vessels} and $I_{\text{background}}$ in its respective vessel and background ROIs is defined as

$$\text{CR} := \frac{I_{\text{vessels}} - I_{\text{background}}}{I_{\text{vessels}} + I_{\text{background}}}, \quad (7)$$

and the contrast-to-noise ratio of a scan with standard deviation $\sigma_{\text{background}}$ in its background ROIs is defined as

$$\text{CNR} := 10 \log_{10} \left(\frac{I_{\text{vessels}} - I_{\text{background}}}{\sigma_{\text{background}}} \right) dB. \quad (8)$$

Negative values in backprojection images were set to zero before calculating the contrast resolution.

To evaluate the effects of denoising on the spectral contrast of MSOT, we applied blind spectral unmixing via non-negative matrix factorization (NMF) [38] to the multispectral optoacoustic images from D_{original} and D_{denoised} and compared the variety and biological interpretability of obtained spectral decompositions. For each of the two datasets (consisting each of 400 × 400 × 81 = 12 960 000 acquired spectra), we obtained a spectral decomposition into 10 non-negative spectral components H (size 10 × 28) and corresponding non-negative unmixing coefficients W (size 12 960 000 × 10).

$$(W, H) := \arg\min_{(W,H)\geq 0} \frac{1}{2} \|S - WH\|_F^2 + \lambda_1 \left(\|W\|_1 + \|H\|_1 \right) + \frac{1}{2} \lambda_F \left(\|W\|_F^2 + \|H\|_F^2 \right), \quad (9)$$

where S (size 12 960 000 x 28) denotes all spectra of the dataset, $\|M\|_F := \left(\sum_{i,j} m_{i,j}^2 \right)^{\frac{1}{2}}$ denotes the Frobenius norm, $\|M\|_1 := \sum_{i,j} |m_{i,j}|$ denotes the entrywise L^1-norm of a matrix $M = \left(m_{i,j}\right)_{i,j}$, and $M \geq 0$ refers to an entrywise inequality. The entrywise L^1-regularization was chosen to promote a maximally sparse decomposition of the spectra, guided by the fact that the spectral contrast of biological tissue is dominated by a small number of abundant chromophores. The number of spectral components chosen was purposefully greater than the number of different chromophores in tissue to also extract variants of the chromophores' absorption spectra that are perceived because of spectral coloring (i.e., distortions of the perceived absorption spectra of chromophores in deeper tissue by light absorption of chromophores in superficial tissue layers). The specific number of components and regularization parameters $\lambda_1 = 50.1$ and $\lambda_F = 50.1$ were selected via parameter space exploration and meaningfulness of the resulting spectral components. Furthermore, the residual norm

$$r_{\text{NMF}} := \|S - WH\|_F^2 / \|S\|_F^2 \quad (10)$$

was evaluated for each of the two NMF runs to quantify the accuracy of the obtained unmixing solution.

III. Results

The proposed deep learning framework for denoising optoacoustic sinograms significantly improves the SNR of both simulated and in vivo data in real-time. The average inference time of the employed deep neural network was 9 milliseconds. Based on the resulting increased quality of the optoacoustic signal data, the denoising method enables improved optoacoustic image contrast and spectral unmixing performance. In the following, we report the detailed findings in the signal, image, and spectral domains.

A. Denoising Performance in the Signal Domain

Optoacoustic signals and electrical noise are both complex broadband signals whose characteristics overlap significantly. The presented data-driven approach can disentangle these overlaps by accessing and separating the data manifolds of signal and noise in sinograms. We observed significant reductions in noise levels, both visually and quantitatively, upon application of the denoising method to sinograms that were corrupted by a combination of Gaussian thermal and complex parasitic electrical noise (Dataset-EN, Dataset-BC, and Dataset-EN+GN with $\text{factor}_{\text{EN}} > 0$), as well as to sinograms that were corrupted only by Gaussian noise (Dataset-EN+GN with $\text{factor}_{\text{EN}} = 0$). Fig. 3 summarizes these results. We begin by visually inspecting the effects of the denoising method for a representative scan of a breast lesion in Fig. 3a-f. Fig. 3a

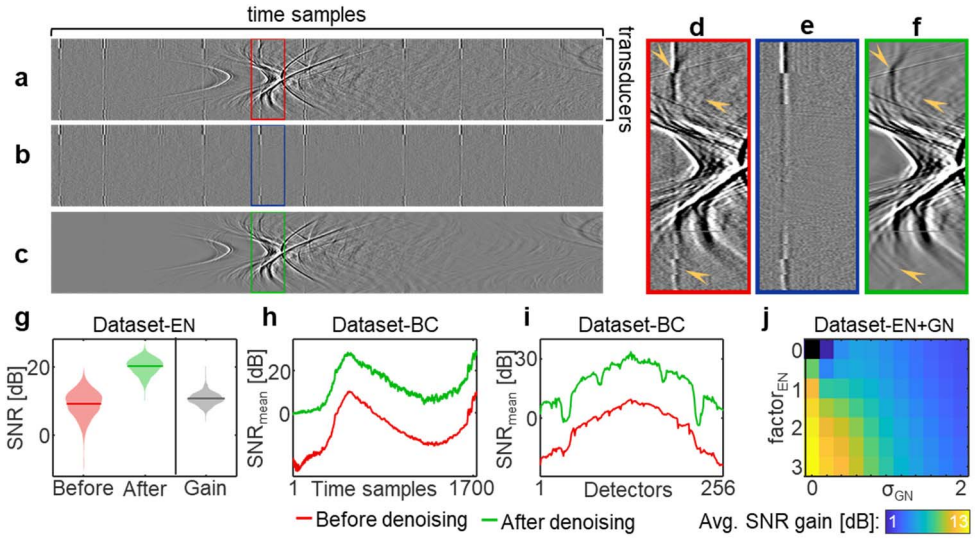

Fig. 3. Evaluation of the proposed denoising approach in the signal domain. a) Noisy sinogram from a representative scan of a human breast lesion. b) Electrical noise component inferred by the neural network. c) Denoised sinogram obtained by subtracting b from a. d-f) Magnifications of the marked areas in a-c. g-j) Quantitative evaluation of the denoising performance. g) Comparison of the SNR distributions in simulated optoacoustic sinograms that are distorted by electrical noise before and after denoising. The mean gain is 10.9 *dB*. h-i) Evaluation of in vivo scans of human breast lesions. h) Mean SNR (SNR_{mean}) of individual time samples. The average increase is 20.8 *dB*. i) Individual SNR_{mean} of all detectors. The average increase is 22.4 *dB*. j) Average SNR gains ("SNR after denoising – SNR before denoising") of the trained model for optoacoustic signals that were corrupted by a combination of measured electrical noise sinograms scaled with $factor_{EN} \in \{0, 0.5, 1, \ldots, 3\}$, and white Gaussian noise with standard deviations $\sigma_{GN} \in \{0, 0.2, 0.4, \ldots, 2\}$. One of the detectors (no. 61) was defective and excluded from the plots in g-j.

shows the noisy sinogram before denoising. Because of the radial nature of wave propagation and the circular shape of the used imaging probe, optoacoustic responses appear as bow-shaped structures in the sinogram. Also note that optoacoustic signals from a spherical target at the center of the circular transducer array appear as bow-shaped structures because of refraction at the interface between the coupling medium inside the imaging probe (heavy water with a speed of sound of 1397 m/s) and the imaged tissue (speed of sound typically in the range 1450 – 1550 m/s). The sinogram is distorted by additive electrical noise that is composed of zero-mean Gaussian noise and complex noise artefacts with strong spatiotemporal correlations (also shown in Fig. 2c). Fig. 3b depicts the electrical noise component inferred by the trained neural network, which demonstrates the network's ability to extract electrical noise. Finally, Fig. 3c shows the denoised sinogram, which was obtained by subtracting the inferred noise from the recorded sinogram. Fig. 3d-f depict enlargements of identical temporal sections of the images in Fig. 3a-c, highlighting the fine features of the optoacoustic signals that are exposed upon removal of electrical noise (yellow arrows).

Fig. 3g-j provide an in-depth quantitative analysis of the network's denoising performance. These results confirmed the ability of the network to consistently remove electrical noise with high accuracy from both synthetic and in vivo optoacoustic sinograms. Fig. 3g compares the distributions of SNRs within the test split of Dataset-EN before and after denoising. Application of the denoising method to the sinograms in Dataset-EN resulted in an average improvement in SNR of 10.9 dB, with improvements for individual sinograms ranging from 4.6 dB to 20.0 dB. After the neural network was trained and tested on Dataset-EN, we applied it to denoise scans of breast lesions (Dataset-BC) to demonstrate its applicability to in vivo data. Fig. 3h shows a plot of the mean SNR_{mean} of these individual time samples from Dataset-BC before and after application of the network, which indicates a time-independent increase in SNR_{mean} of approximately 20.8 dB after denoising. The uniformity of the increase in SNR_{mean} demonstrates that the trained neural network can extract electrical noise both from strong optoacoustic responses in superficial tissue (time samples 400 – 700), as well as from signals deeper in tissue, which have lower amplitudes due to light fluence attenuation (time samples 1100 – 1400).

Furthermore, we demonstrated the ability of the method to compensate for the variations in parasitic electrical noise within the transducer array of the test system (see Fig. 2c for details) to confirm the applicability of the trained network to in vivo scans. For that, we calculated the SNR_{mean} for Dataset-BC individually for all transducer elements, rather than for the whole sinograms, before and after denoising. As shown in Fig. 3i, applying the trained network to the breast scans from Dataset-BC improves the SNR_{mean} at all transducers by an average of 22.4 dB in a near uniform manner. Note that the transducers at the boundaries of the detector probe have lower SNRs than the central transducers due to the probe layout partially shielding the outermost transducers from arriving acoustic waves. The lower SNR_{mean} values at transducers 30-43, 79-87, 167-175, 213-227 result from acoustic noise waves propagating along the transducer array, which corrupts the ground truth noise estimation used to calculate the SNR_{mean} (see Equation 5). These noise waves depend on the imaged tissue, and therefore cannot be removed by the neural network.

Thus far, we have demonstrated the ability of the denoising method to accurately isolate the electrical noise of the test

system. Next, we investigated whether the trained model can also adopt electrical noise with altered levels of parasitic and thermal noise components, e.g. caused by changes in hardware configurations or environmental conditions such as humidity and temperature. We applied the trained model to optoacoustic signals that were corrupted with augmented electrical noise sinograms containing, in comparison to the measured electrical noise used during training, up to three times the amount of parasitic noise and approximately up to four times the amount of thermal noise (see Dataset-EN+GN in Table I). Fig. 3j summarizes the average denoising performance of the trained network, which afforded improvements in SNR for all tested combinations of parasitic and thermal noise components. These results demonstrate that the trained neural network can also generalize to previously unseen amounts of electrical noise and thus facilitate robust denoising in real-world imaging applications, where the amounts of electrical noise may change over time.

B. Denoising Enables High Contrast in Optoacoustic Images

Thus far, we have demonstrated the ability of the denoising network to isolate and remove electrical noise in optoacoustic sinograms. In this section, we analyze the effects of denoising on reconstructed optoacoustic images. First, we ensure that the denoising network successfully removes noise artifacts without distorting any true optoacoustic image structures using optoacoustic images of a phantom. Subsequently, we evaluate the improved image contrast due to denoising in a clinical dataset of scans of human breast lesions, to show the potential for improved diagnostic capability of optoacoustic tomography.

We utilized a model-based inversion algorithm to reconstruct optoacoustic images (i.e. initial spatial pressure distributions) from all scans in the datasets Dataset-Ph and Dataset-BC, both with and without denoising the recorded sinograms with the trained neural network. Fig. 4a-e illustrate the qualitative improvements to selected images upon application of the neural network. Fig. 4a shows an optoacoustic image of a phantom at 700 nm, reconstructed from a noisy sinogram. Zero-mean Gaussian noise in the recorded sinogram reduces the overall contrast in the image, whereas parasitic noise leads to ring artifacts that obscure potentially relevant image features. The arrangement of the phantom is shown in the inlay of Fig. 4a. Fig. 4b depicts the same optoacoustic image reconstructed from a denoised sinogram, demonstrating that the neural network can significantly reduce both the background noise and the ring artifacts. Fig. 4c plots the difference between Figs. 4a and b, which yields artifacts and background noise but no real structures, emphasizing the ability of the network to accurately identify and isolate noise in optoacoustic images. Fig. 4d and e show the optoacoustic images of a malignant breast tumor at 870 nm. The denoised image in Fig. 4e appears significantly richer in contrast than the original image in Fig. 4d and contains structures that are not visible prior to the denoising. To highlight the clinical relevance of the improved contrast, note that in Fig. 4e, the optoacoustic contrast inside the tumor core regions (outlined in blue) is separated from the noise that dominates these regions in Fig. 4d. Furthermore, we reconstructed the scans from Dataset-BC with the backprojection algorithm and confirmed that the presented denoising method also achieves visible reductions in the background noise and the ring artifacts for these images.

Next, we evaluated the contrast resolution (CR) of blood vessels in optoacoustic images reconstructed via model-based inversion from Dataset-BC to quantify the enhancement capabilities of the trained neural network in the image domain. Blood vessels and background ROIs were first manually segmented, as depicted in Fig. 4f, and used to calculate the blood contrast resolution. The distributions of contrast resolution in Dataset-BC before and after denoising are compared in Fig. 4g, which shows an average improvement of 0.083 with a range of 0.003 to 0.55 for individual images. As shown in Fig. 4h, the average improvement in blood contrast resolution is consistent across all wavelengths, demonstrating the network's ability to remove noise, independent of the varying strength of individual absorbers across the accessible spectrum. The denoising capabilities of the presented denoising method were also confirmed when evaluating the contrast-to-noise ratio of blood vessels in the images from model-based inversion, and the contrast resolution and contrast-to-noise ratio of blood vessels in images from backprojection reconstruction, with average improvements of 1.7 dB, 0.043, and 2.6 dB, respectively.

C. Deep-Learning-Based Denoising Enables High Spectral Contrast

A further central finding of this work is the ability of the presented denoising approach to significantly improve spectral contrast in MSOT, i.e. the differentiation of chromophores based on their absorption spectra. We found that upon application of the denoising method to the MSOT scans from Dataset-BC, the dominant absorbers in breast tissue – hemoglobin, lipids, and water – are more accurately identified and localized. To analyze the spectral contrast, we applied blind spectral unmixing via non-negative matrix factorization (NMF) to the original and denoised breast images and decomposed each of the two datasets into 10 spectra and corresponding unmixing coefficients. Note that unlike linear unmixing based on the reference absorption spectra of chromophores in tissue, NMF finds both the spectra and unmixing coefficients in a data driven way and thus extracts variants of the reference spectra that consider effects from spectral coloring.

Fig. 5 compares the spectral contrast of the original and denoised MSOT breast images from Dataset-BC. In Fig. 5a-c, we show the NMF spectra obtained from the original (Fig. 5a) and from the denoised (Fig. 5b) data next to the reference absorption spectra of the most prominent chromophores in tissue (oxygenated and deoxygenated hemoglobin, water, and lipids, see Fig. 5c). The spectra derived from the original data show a significant number of sharp peaks (spectra no. 3, 5, 7, 8, 9, 10) attributable to ring artifacts from parasitic noise, rather than specific absorbers in tissue. In contrast, the spectra

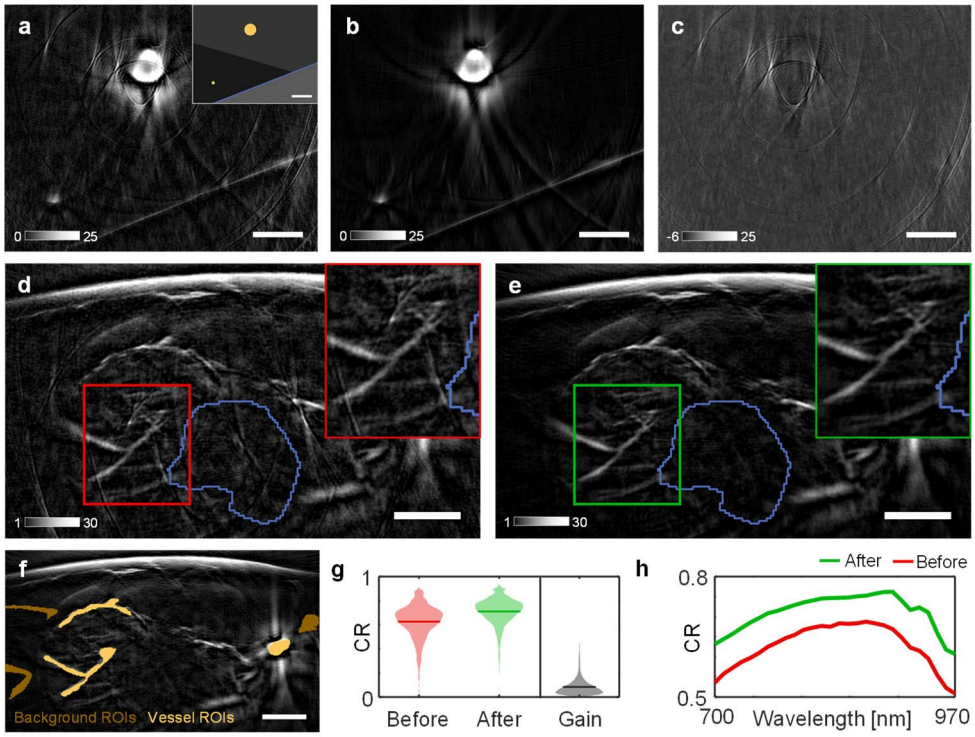

Fig. 4. Demonstration of improved image quality in denoised scans of a phantom and of human breast lesions. a) Optoacoustic image of a phantom before denoising. The overlayed image shows the arrangements of the phantom: tubes filled with ink (yellow), copper sheet (blue), and agar layers with slightly different speed of sound distributions (grey). b) Corresponding optoacoustic image after denoising. c) Difference between a and b. d-e) Optoacoustic image of a malignant breast tumor (d) before and (e) after denoising. The location of the hypoechoic tumor core, obtained from ultrasound images, is outlined in blue. f) Examples for the vessel and background ROIs that are used to compute the contrast resolution. g-h) Quantification of the contrast resolution (CR) of blood vessels in scans of breast lesions before and after denoising. The average increase is 0.083. The minimal gain is 0.003. The depicted optoacoustic images of the phantom and the breast lesion are obtained at 700 nm and 870 nm, respectively. All scale bars correspond to 5 mm.

obtained from the denoised images are broader, smoother, and are more easily related to the reference absorption spectra of hemoglobin (spectra no. 1, 2, 6, 7, 8, 10), fat (spectra no. 4, 9), and water (spectra no. 3, 5). The increased number of meaningful spectra found by NMF demonstrates superior spectral contrast of the denoised images compared to the original images. The improved unmixing accuracy upon application of the denoising method was also confirmed quantitatively by evaluating the residual norms of the NMF runs (see equation 10). The 10 NMF components obtained from the original MSOT breast scans could only represent 83.2% of the data ($r_{NMF} = 16.8\%$), whereas the 10 NMF components obtained from the denoised MSOT breast scans could represent 91.8% of the data ($r_{NMF} = 8.2\%$).

In Fig. 5d-g, we visualize the obtained spectral decompositions before and after denoising of a representative multispectral stack to visually confirm the ability of the network to enable better spectral contrast. We color-encode and blend the unmixing coefficients of three NMF spectra at a time, which correlate with the reference absorption spectra of hemoglobin (Fig. 5d,e), lipids, and water (Fig. 5f,g), covering approximately the same spectral regions for the original and the denoised data. To improve the dynamic range of the rendered images, we display the square roots of all coefficients in the visualizations. Whereas the visualizations derived from the original data are dominated by overlapping coefficients of different spectra (appearing as white in the color-encoding) and by ring noise artifacts (example marked with white arrows in Fig. 5f), the visualizations derived from denoised data show a reduction of noise artifacts and express significantly richer spectral contrast. In addition, while the tumor core (outlined in white) in Fig. 5d,f contains a lot of noise, this noise is removed by the denoising method in Fig. 5e,g, revealing hemoglobin contrast inside the tumor (white arrows in Fig. 5e). In summary, improved spectral contrast is observable in two ways upon application of the denoising method to the scans from Dataset-BC: First, blind spectral unmixing retrieves a more versatile set of spectral components and second, the denoising method enables a more meaningful decomposition of the acquired images into the found spectra.

IV. Conclusion

Optoacoustic signals are relatively weak and thus susceptible to corruption by electrical noise during the imaging process, which impedes morphological and spectral contrast. In this work, we presented a discriminative deep-learning-based denoising method for optoacoustic sinograms, which employs a deep neural network trained on samples of experimentally acquired electrical noise and simulated ground truth optoacoustic signals. We demonstrated that the trained deep neural network could accurately remove electrical noise from

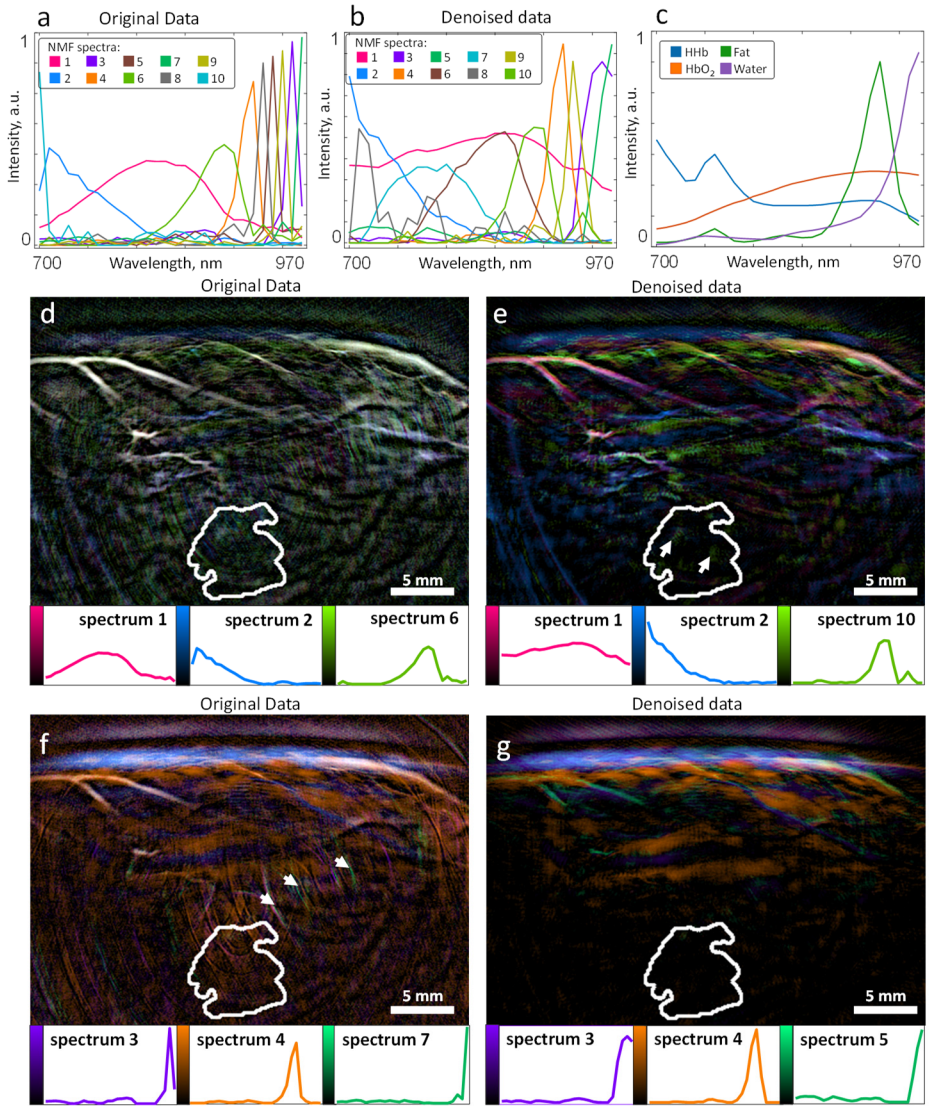

Fig. 5. Effects of denoising on the spectral content of optoacoustic images. a-b) NMF spectra that were obtained from the original (a) and the denoised (b) MSOT images of human breast lesions from Dataset-BC. c) Reference absorption spectra of the most prominent chromophores in breast tissue. d-g) Visualizations of the NMF decomposition of a representative MSOT image showing a malignant breast tumor at approximately 2 cm depth before (d, f) and after (e, g) denoising. The images color-encode the contributions of three spectra that respectively correlate with the absorption spectra of hemoglobin (d, e), fat, and water (f, g). The position of the tumor, obtained from ultrasound images, is demarcated by the white outlines.

in vivo scans of a MSOT system. The proposed signal processing technique offers a fast and accurate approach to improve the SNR of recorded optoacoustic sinograms, increase morphological image contrast, and enable rich spectral contrast at high resolution in handheld MSOT imaging.

The presented deep-learning-based denoising framework is effective because it can access the topology and the statistics of pure electrical noise and optoacoustic signal datasets. This structural information contained in large datasets has recently been made accessible by advances in computational power and methodology and is the driving force behind the increasing success of deep learning methods in medical imaging [39], [40]. We generated such a large and high-quality dataset by complementing the experimentally acquired pure noise data with simulated optoacoustic sinograms. The simulated data was obtained by applying a mathematical model of the imaging system to a general-feature image database, thereby incorporating prior knowledge about the imaging system without sacrificing general applicability of the method to any data acquired by the system. This is an example of the integration of a physical model into data-driven methods, which remains a major challenge in machine learning [27], [41], [42]. The proposed method allows a trade-off between model accuracy and generality. For example, one could potentially enhance denoising performance by selecting an optoacoustic signal dataset that more specifically reflects typical tissue responses. However, the method achieves accurate denoising and good generalization beyond the training data without any such specialization. To further evaluate the utility of the presented denoising approach, future research

may also focus on strategies to improve the interpretability of the employed deep neural network model [43].

Furthermore, the trained deep neural network model provides a means of fast denoising. Clinical optoacoustic imaging systems typically provide real-time feedback to the device operator on a built-in monitor. Due to the restricted processing times, these online images are usually much lower in quality than those produced offline, which can lead to longer imaging sessions and incorrect selection of regions of interest. We demonstrated that the method can denoise a full optoacoustic sinogram of the MSOT system in approximately 9 milliseconds, which is fast enough for real-time feedback during device operation. Improving instantaneous image quality enhances the dynamic imaging capabilities of MSOT [44] while decreasing examination times.

In addition to better image quality, the denoising method also enhances the fidelity of the obtained spectral information. MSOT enables molecular contrast by extending the high-resolution optical contrast of optoacoustic imaging to the spectral dimension [45]. However, previous clinical MSOT studies extracted spectral information mostly by averaging over larger areas in MSOT images [6]–[8], thereby sacrificing the superior resolution of optoacoustic imaging. In this work, we demonstrated that denoising overcomes the necessity to average over large tissue regions and enables spectral contrast down to the system resolution, which is $\sim$200 μm in the test system. High-resolution spectral contrast was highlighted by localizing hemoglobin contrast inside a 2 cm deep breast tumor. Spectral contrast is of the utmost interest for clinical applications of MSOT, since it, for example, enables detailed studies of local blood oxygenation and tissue metabolism.

Finally, the presented denoising framework is also applicable to other (optoacoustic) imaging systems. For example, optoacoustic mesoscopy [46] and microscopy systems [47] are beset with similar electrical noise, making the approach of acquiring pure noise measurements and simulating signals with a numerical model applicable to these systems without any major changes. Other noise sources, like speckle noise in ultrasound imaging [48] or optical coherence tomography [49] and shot noise in coherent diffraction imaging [50] can be modeled as independent multiplicative noise and can thus be approached by adapting the proposed method accordingly. More generally, the presented methodology can in any context disentangle two random fields that are mixed in a known way and whose distributions can be accessed by sampling. In particular, the denoising approach can also be applied to remove signal-dependent noise if samples are obtained from the conditional probability distribution of the noise $P(S_{\text{noise}}|S_{\text{OA}} = s_{\text{OA}})$. Then, the denoising network could be trained with signal and noise samples that are generated through the following two-step process: 1: Get sample s_{OA} from $S_{\text{OA}} \sim P_{\text{OA}}$. 2: Get sample s_{noise} from $(S_{\text{noise}}|S_{\text{OA}} = s_{\text{OA}}) \sim P(S_{\text{noise}}|S_{\text{OA}} = s_{\text{OA}})$.

In summary, the deep learning framework that we propose in this work is an efficient and flexible method for denoising optoacoustic tomography data. By significantly improving the data quality of the considered MSOT system, we move one step closer to the full potential of handheld MSOT imaging, which is dynamic high-resolution molecular contrast deep in tissue.

Source Code

The source code for the presented denoising framework is available at
https://github.com/juestellab/msot-sinogram-denoising.

Acknowledgment

The authors would like to thank Hong Yang for providing and scanning the phantom, Angelos Karlas and Korbinian Paul-Yuan for designing the study to acquire the breast scans, Stefan Metz for acquiring the breast scans, Jan Kukačka for his help with the analysis of the MSOT breast images, and Robert Wilson for his attentive reading and improvements of the manuscript.

References

[1] S. Tzoumas, A. Rosenthal, C. Lutzweiler, D. Razansky, and V. Ntziachristos, "Spatiospectral denoising framework for multispectral optoacoustic imaging based on sparse signal representation," *Med. Phys.*, vol. 41, no. 11, 2014, Art. no. 113301.

[2] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital Communication*. New York, NY, USA: Springer, 2004.

[3] B. T. Cox, J. G. Laufer, P. C. Beard, and S. R. Arridge, "Quantitative spectroscopic photoacoustic imaging: A review," *J. Biomed. Opt.*, vol. 17, no. 6, 2012, Art. no. 061202.

[4] A. Taruttis and V. Ntziachristos, "Advances in real-time multispectral optoacoustic imaging and its applications," *Nature Photon.*, vol. 9, no. 4, pp. 219–227, Apr. 2015.

[5] P. Beard, "Biomedical photoacoustic imaging," *Interface Focus*, vol. 1, no. 4, pp. 602–631, Aug. 2011.

[6] F. Knieling *et al.*, "Multispectral optoacoustic tomography for assessment of Crohn's disease activity," *New England J. Med.*, vol. 376, no. 13, pp. 1292–1294, Mar. 2017.

[7] A. P. Regensburger *et al.*, "Detection of collagens by multispectral optoacoustic tomography as an imaging biomarker for Duchenne muscular dystrophy," *Nature Med.*, vol. 25, no. 12, pp. 1905–1915, Dec. 2019.

[8] W. Roll *et al.*, "Multispectral optoacoustic tomography of benign and malignant thyroid disorders: A pilot study," *J. Nucl. Med.*, vol. 60, no. 10, pp. 1461–1466, Oct. 2019.

[9] K. Wang, R. Su, A. A. Oraevsky, and M. A. Anastasio, "Investigation of iterative image reconstruction in three-dimensional optoacoustic tomography," *Phys. Med. Biol.*, vol. 57, no. 17, pp. 5399–5423, Aug. 2012.

[10] A. Buehler, A. Rosenthal, T. Jetzfellner, A. Dima, D. Razansky, and V. Ntziachristos, "Model-based optoacoustic inversions with incomplete projection data," *Med. Phys.*, vol. 38, no. 3, pp. 1694–1704, 2011.

[11] H. Yang *et al.*, "Soft ultrasound priors in optoacoustic reconstruction: Improving clinical vascular imaging," *Photoacoustics*, vol. 19, Sep. 2020, Art. no. 100172.

[12] J. Kukačka *et al.*, "Image processing improvements afford second-generation handheld optoacoustic imaging of breast cancer patients," *Photoacoustics*, vol. 26, Jun. 2022, Art. no. 100343.

[13] S. Sethuraman *et al.*, "Intravascular photoacoustic imaging using an IVUS imaging catheter," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 54, no. 5, pp. 978–986, May 2007.

[14] J.-M. Yang *et al.*, "Photoacoustic endoscopy," *Opt. Lett.*, vol. 34, no. 10, pp. 1591–1593, 2009.

[15] L. Zeng, D. Xing, H. Gu, D. Yang, S. Yang, and L. Xiang, "High antinoise photoacoustic tomography based on a modified filtered backprojection algorithm with combination wavelet," *Med. Phys.*, vol. 34, no. 2, pp. 556–563, Jan. 2007.

[16] S. H. Holan and J. A. Viator, "Automated wavelet denoising of photoacoustic signals for circulating melanoma cell detection and burn image reconstruction," *Phys. Med. Biol.*, vol. 53, no. 12, pp. N227–N236, Jun. 2008.

[17] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Advances in Neural Information Processing Systems*, vol. 25, F. Pereira, C. J. C. Burges, L. Bottou, Eds. Red Hook, NY, USA: Curran Associates, 2012, pp. 341–349.

[18] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[19] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder–decoder networks with symmetric skip connections," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, Barcelona, Spain, 2016, pp. 2810–2818.

[20] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 154–169.

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

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

[23] D. Allman, A. Reiter, and M. A. L. Bell, "Photoacoustic source detection and reflection artifact removal enabled by deep learning," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1464–1477, Jun. 2018.

[24] H. Shan, G. Wang, and Y. Yang, "Accelerated correction of reflection artifacts by deep neural networks in photo-acoustic tomography," *Appl. Sci.*, vol. 9, no. 13, p. 2615, Jun. 2019.

[25] E. M. A. Anas, H. K. Zhang, J. Kang, and E. Boctor, "Enabling fast and high quality LED photoacoustic imaging: A recurrent neural networks based approach," *Biomed. Opt. Exp.*, vol. 9, no. 8, pp. 3852–3866, 2018.

[26] A. Hariri, K. Alipour, Y. Mantri, J. P. Schulze, and J. V. Jokerst, "Deep learning improves contrast in low-fluence photoacoustic imaging," *Biomed. Opt. Exp.*, vol. 11, no. 6, pp. 3360–3373, 2020.

[27] J. Gröhl, M. Schellenberg, K. Dreher, and L. Maier-Hein, "Deep learning for biomedical photoacoustic imaging: A review," *Photoacoustics*, vol. 22, Jun. 2021, Art. no. 100241.

[28] S. Prahl. (1999). *Tabulated Molar Extinction Coefficient for Hemoglobin in Water*. [Online]. Available: http://omlc.ogi.edu/spectra/hemoglobin/summary.html

[29] D. J. Segelstein, *The Complex Refractive Index of Water*. Kansas City, MO, USA: Univ. of Missouri-Kansas City, 1981.

[30] R. L. P. van Veen, H. J. C. M. Sterenborg, A. Pifferi, A. Torricelli, E. Chikoidze, and R. Cubeddu, "Determination of visible near-IR absorption coefficients of mammalian fat using time- and spatially resolved diffuse reflectance and transmission spectroscopy," *J. Biomed. Opt.*, vol. 10, no. 5, 2005, Art. no. 054004.

[31] K. B. Chowdhury, J. Prakash, A. Karlas, D. Justel, and V. Ntziachristos, "A synthetic total impulse response characterization method for correction of hand-held optoacoustic images," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3218–3230, Oct. 2020.

[32] K. B. Chowdhury, M. Bader, C. Dehner, D. Jüstel, and V. Ntziachristos, "Individual transducer impulse response characterization method to improve image quality of array-based handheld optoacoustic tomography," *Opt. Lett.*, vol. 46, no. 1, pp. 1–4, 2021.

[33] M. Everingham *et al.*, "The PASCAL visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, pp. 303–338, 2010.

[34] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, in Lecture Notes in Computer Science, 2015, pp. 234–241.

[35] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2015, *arXiv:1412.6980*.

[36] L. A. Kunyansky, "Explicit inversion formulae for the spherical mean radon transform," *Inverse Problems*, vol. 23, no. 1, pp. 373–383, Feb. 2007.

[37] P. Kuchment and L. Kunyansky, "Mathematics of photoacoustic and thermoacoustic tomography," in *Handbook of Mathematical Methods in Imaging*, O. Scherzer, Ed. New York, NY, USA: Springer, 2011, pp. 817–865.

[38] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. 13th Int. Conf. Neural Inf. Process. Syst.*, Denver, CO, USA, 2000, pp. 535–541.

[39] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153–1159, May 2016.

[40] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[41] G. Carleo *et al.*, "Machine learning and the physical sciences," *Rev. Mod. Phys.*, vol. 91, no. 4, 2019, Art. no. 045002.

[42] A. Hauptmann and B. Cox, "Deep learning in photoacoustic tomography: Current approaches and future directions," *J. Biomed. Opt.*, vol. 25, no. 11, Oct. 2020, Art. no. 112903.

[43] Q. Zhang, Y. N. Wu, and S.-C. Zhu, "Interpretable convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8827–8836.

[44] A. Karlas *et al.*, "Multispectral optoacoustic tomography of muscle perfusion and oxygenation under arterial and venous occlusion: A human pilot study," *J. Biophoton.*, vol. 13, no. 6, Jun. 2020, Art. no. e201960169.

[45] V. Ntziachristos and D. Razansky, "Molecular imaging by means of multispectral optoacoustic tomography (MSOT)," *Chem. Rev.*, vol. 110, no. 5, pp. 2783–2794, 2010.

[46] J. Aguirre *et al.*, "Precision assessment of label-free psoriasis biomarkers with ultra-broadband optoacoustic mesoscopy," *Nature Biomed. Eng.*, vol. 1, no. 5, p. 68, 2017.

[47] M. A. Pleitez *et al.*, "Label-free metabolic imaging by mid-infrared optoacoustic microscopy in living cells," *Nature Biotechnol.*, vol. 38, no. 3, pp. 293–296, Mar. 2020.

[48] A. Fenster, D. B. Downey, and H. N. Cardinal, "Three-dimensional ultrasound imaging," *Phys. Med. Biol.*, vol. 46, no. 5, pp. R67–R99, 2001.

[49] A. F. Fercher, W. Drexler, C. K. Hitzenberger, and T. Lasser, "Optical coherence tomography—Principles and applications," *Rep. Prog. Phys.*, vol. 66, no. 2, pp. 239–303, 2003.

[50] I. Robinson and R. Harder, "Coherent X-ray diffraction imaging of strain at the nanoscale," *Nature Mater.*, vol. 8, no. 4, pp. 291–298, 2009.