

---

# OADAT: Experimental and Synthetic Clinical Optoacoustic Data for Standardized Image Processing

---

**Firat Ozdemir<sup>1\*</sup>**   **Berkan Lafci<sup>2,3\*</sup>**   **Xosé Luís Deán-Ben<sup>2,3</sup>**  
**Daniel Razansky<sup>2,3</sup>**   **Fernando Perez-Cruz<sup>1,4</sup>**

<sup>1</sup>Swiss Data Science Center, ETH Zurich and EPFL, Zurich, Switzerland

<sup>2</sup>Institute of Pharmacology and Toxicology and Institute for Biomedical Engineering, Faculty of Medicine, University of Zurich, Switzerland

<sup>3</sup>Institute for Biomedical Engineering, Department of Information Technology and Electrical Engineering, ETH Zurich, Switzerland

<sup>4</sup>Institute for Machine Learning, Department of Computer Science, ETH Zurich, Switzerland

## Abstract

Optoacoustic (OA) imaging is based on excitation of biological tissues with nanosecond-duration laser pulses followed by subsequent detection of ultrasound waves generated via light-absorption-mediated thermoelastic expansion. OA imaging features a powerful combination between rich optical contrast and high resolution in deep tissues. This enabled the exploration of a number of attractive new applications both in clinical and laboratory settings. However, no standardized datasets generated with different types of experimental set-up and associated processing methods are available to facilitate advances in broader applications of OA in clinical settings. This complicates an objective comparison between new and established data processing methods, often leading to qualitative results and arbitrary interpretations of the data. In this paper, we provide both experimental and synthetic OA raw signals and reconstructed image domain datasets rendered with different experimental parameters and tomographic acquisition geometries. We further provide trained neural networks to tackle three important challenges related to OA image processing, namely accurate reconstruction under limited view tomographic conditions, removal of spatial undersampling artifacts and anatomical segmentation for improved image reconstruction. Specifically, we define 44 experiments corresponding to the aforementioned challenges as benchmarks to be used as a reference for the development of more advanced processing methods.

## 1 Introduction

Optoacoustic (OA) imaging is being established as a powerful method with increasing application areas in clinical [37, 38] and preclinical settings [21, 27]. Using nanosecond-duration pulsed lasers operating in the visible and near-infrared (NIR) optical wavelength range, biological tissues are thermoelastically excited. This excitation yields ultrasound (US) waves, from which OA images are tomographically reconstructed (Fig. 1a). The rich optical contrast from endogenous tissue chromophores such as blood, melanin, lipids and others are combined with high US resolution, i.e., tens of micrometers. This unique feature makes OA particularly suitable for molecular and functional imaging. Other important advantages such as the feasibility of hand-held operation, the fast acquisition performance (real time feedback) and the non-invasive safe contrast (i.e., non-ionizing radiation) further foster the wide use of OA in multiple biomedical studies. OA imaging has been shown to provide unique capabilities in studies with disease models e.g., of breast cancer [26, 9, 2],

---

\*Equal contribution. Author ordering determined by coin flip. For correspondence, please contact either firat.ozdemir@datascience.ch or berkan.lafci@uzh.ch.

as well as for the clinical assessment of Crohn’s disease [20], atherosclerotic carotid plaques [18] or skin cancer [7]. As the range of applications of OA imaging gets broader, the need for different data processing pipelines increases in parallel. Also, new methods are continuously being developed to provide an enhanced OA performance. Specific examples include increased temporal resolution with compressed/sparse data acquisitions [31], accurate image reconstruction algorithms [5], light fluence correction by segmenting the tissue boundaries [22] or enhanced spectral unmixing algorithms from multispectral data [39].

Three major challenges suitable for data-driven approaches in clinical OA imaging are summarized below:

**Sparse acquisition:** OA imaging provides a unique potential to monitor fast-changing events such as cardiac arrhythmias [41], neuronal activity [33] or indocyanine green clearance [12] *in vivo*. For this, ultra-fast imaging systems capable of capturing changes in living organisms occurring at up to millisecond temporal scales are required. The main limiting factor affecting the achievable frame rate is the data transfer capacity. This limitation can be eliminated by reducing the number of acquired channels (signals). Therefore, sparse or compressed sensing methods have been proposed both using conventional methods [31] and deep learning algorithms [3].

**Limited view reconstruction:** OA is inherently a tomographic imaging modality. Acquisition of pressure signals from different angles is essential to capture the information encoded in US waves traveling in a 3D medium in order to render accurate tomographic reconstructions. This further increases the image contrast, resolution and quantitativeness. However, tomographic coverage of the samples is often hindered by physical restrictions. Thereby, new image processing pipelines have been suggested to improve limited-view-associated challenges in OA imaging by using data-driven algorithms in the image domain [13], signal domain [19] and combination of both domains [4, 23].

**Segmentation:** Optimal OA reconstruction algorithms need to account for different optical and acoustic properties in biological tissues and in the coupling medium (water). For example, the speed of sound (SoS) depends on the elastic properties of the medium. Proper assignment of SoS values in tissues and in water requires accurate delineation of the tissue boundaries. Thereby, segmentation of structures [22, 35] in OA images has been shown to enhance the image reconstruction performance. Additionally, the optical fluence (intensity) also varies with depth across different tissues. This issue remains as one of the main factors affecting quantification in OA images [1] and can also be corrected with tissue segmentation [25].

As an emerging method, OA imaging requires standardization, open source code publication and data sharing practices to expedite the development of new application areas and data processing pipelines. Particularly, the aforementioned challenges associated to high data throughput, limited angular coverage, SoS assignment and fluence corrections require coordinated efforts between experimental and data science communities. Initial efforts to standardize data storage formats and image reconstruction algorithms have been undertaken [10]. However, [10] focus on standard reconstruction algorithms and propose data storage formats for acquisition related metadata. Data-driven image and signal processing methods require additional initiatives on fast access to large bulks of OA image and signal data and benchmarks for learning-based methods.

Here, we provide experimental data and simulations of forearm datasets as well as benchmark networks aiming at facilitating the development of new image processing algorithms and benchmarking. These “Experimental and Synthetic Clinical Optoacoustic Data (OADAT)” include, (i) large and varied clinical and simulated forearm datasets with paired subsampled or limited view image reconstruction counterparts, (ii) raw signal acquisition data of each such image reconstruction, (iii) definition of 44 experiments with gold standards focusing on the aforementioned OA challenges, (iv) pretrained model weights of the networks used for each task, and (v) user-friendly scripts to load and evaluate the networks on our datasets. The presented datasets and algorithms will expedite the research in OA image processing.

## 2 Background

For OA imaging, the objects are excited with the nanosecond-duration laser pulses in visible or NIR light wavelengths which result in thermoelastic expansion of the structures. This expansion generates pressure waves (US signals) that are detected by transducer arrays. Corresponding images are reconstructed by solving the OA inverse problem on the acquired signals. Below, we explain the transducer arrays used for data acquisition, the reconstruction algorithm used to generate images from

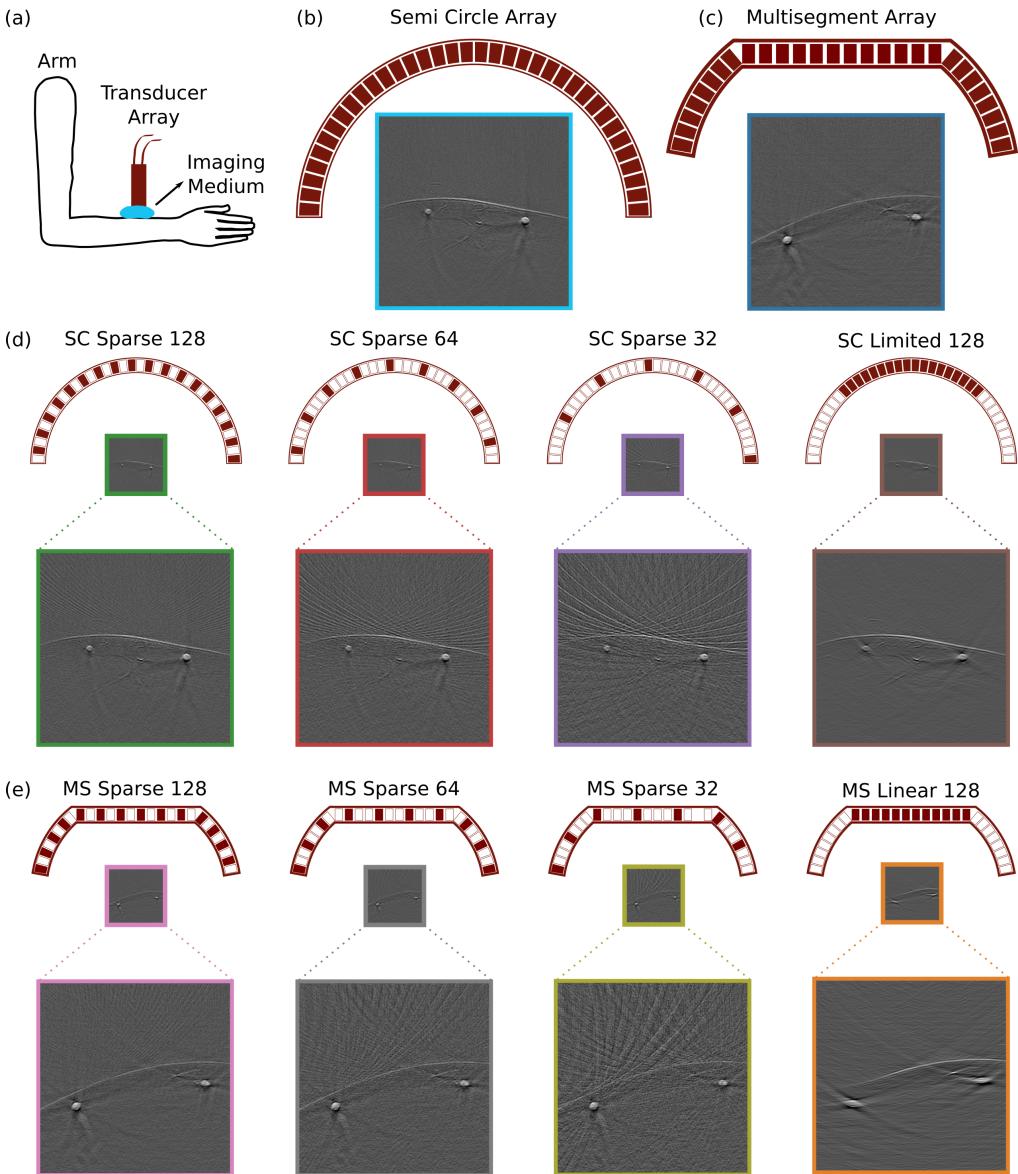


Figure 1: Experimental data acquisition, transducer arrays and resulting images. (a) Experimental setup for optoacoustic forearm imaging. (b) Semi circle array along with an example of acquired images. (c) Multisegment array along with an example of acquired images. (d) Uniform subsampling for the semi circle array (128, 64 and 32 elements) and limited view acquisition for the semi circle array with reduced angular coverage (128 elements). (e) Uniform subsampling for the multisegment array (128, 64 and 32 elements), and linear array acquisition for the multisegment array (128 elements). Transducer elements are shown as actively receiving (red) or off (white).

acquired signals and the sampling/acquisition techniques. Detailed explanation about OA imaging and used tools can be found in Appendix A.

## 2.1 Transducer arrays

**Semi circle** array contains 256 transducers elements distributed equidistantly over a semi circle (concave surface, Fig. 1b). **Multisegment** array is a combination of linear array in the center and concave parts on the right and left sides, designed to increase angular coverage, as shown in Fig. 1c. The linear part contains 128 elements and each of the concave parts consist of 64 elements, totaling to 256. **Linear** array is the central part of the multisegment array with 128 elements (Fig. 1c). The array geometry is optimized for US data acquisitions with planar waves. Hence, it produces OA images with limited view artifacts due to reduced angular coverage which is a limiting factor for OA image acquisitions. **Virtual circle** array is generated to simulate images with 360 degree angular coverage and yields artifact free reconstructions (Fig. 2a). It contains 1,024 transducer elements distributed over a full circle with equal distance. We also have a virtual multisegment array that correspond to its physical counterpart. Additional geometric and technical details are listed in Appendix A.2.

## 2.2 Reconstruction method

We use backprojection algorithm in this study to generate OA images from the acquired signals<sup>2</sup>. This algorithm is based on delay and sum beamforming approach [30] (see Appendix A.3 for details). First, a mesh grid is created to represent the imaged field of view. Then, the distance between the points of the mesh grid and array elements are calculated based on the known locations of the transducers. Time of flight is obtained through dividing distance by the SoS values that are assigned based on the temperature of the imaging medium and tissue properties. The clinical and simulated data are reconstructed with SoS of 1,510 m/s in this study as the simulations and the experiments were done at the corresponding imaging medium temperature. Unlike more sophisticated model-based reconstruction approaches [5], backprojection is parameterized using only SoS, making it a stable choice across all imaged scenes.

## 2.3 Sparse sampling

Sparse sampling yields streak artifacts on reconstructed images due to large inter-element pitch size. For a given angular coverage, i.e., transducer array geometry, using less transducer elements for reconstruction causes stronger artifacts due to increased inter-element pitch size. We define sparse sampled semi circle array acquisitions **semi circle sparse 128**, **semi circle sparse 64** and **semi circle sparse 32** when using 128, 64 and 32 elements out of the 256 of semi circle array (Fig. 1d first three columns), respectively. Similarly, we define sparse sampled virtual circle array acquisitions **virtual circle sparse 128**, **virtual circle sparse 64** and **virtual circle sparse 32** when using 128, 64 and 32 elements out of 1,024 of virtual circle array (Fig. 2c first three rows), respectively. In addition, we also define sparse sampled multisegment array acquisitions **multisegment sparse 128**, **multisegment sparse 64** and **multisegment sparse 32** when using 128, 64 and 32 elements out of the 256 elements of the multisegment array (Figs. 1e and 2d first three columns), respectively. All items correspond to uniform and hence equidistant subsampling of the corresponding transducer array signals.

## 2.4 Limited View

Limited view acquisitions lead to distorted geometry (e.g., elongated vessels) due to the reduced angular coverage (Figs. 1d,e & 2c,d last column, limited 128 and linear 128). To mimic commonly occurring limited view settings, we use a continuous subset of elements for a given transducer. This corresponds to retaining inter-element pitch size while reducing the angular coverage. We define a limited view acquisition for each transducer array as follows: (i) **Linear array** is the common practice in clinical settings for US data acquisition [17, 24]. Typically, the same linear geometry is combined with OA imaging to provide complementary information [28]. To model this clinically realistic scenario, we use the linear part of the multisegment array for OA image reconstruction. (ii) **Semi circle limited 128** uses half of the semi circle array; 128 transducer elements, yielding a quarter circle. The differences between linear and semi circle limited view array acquisitions are the

---

<sup>2</sup>Python module for OA reconstruction: [github.com/berkanlafci/pyoat](https://github.com/berkanlafci/pyoat).

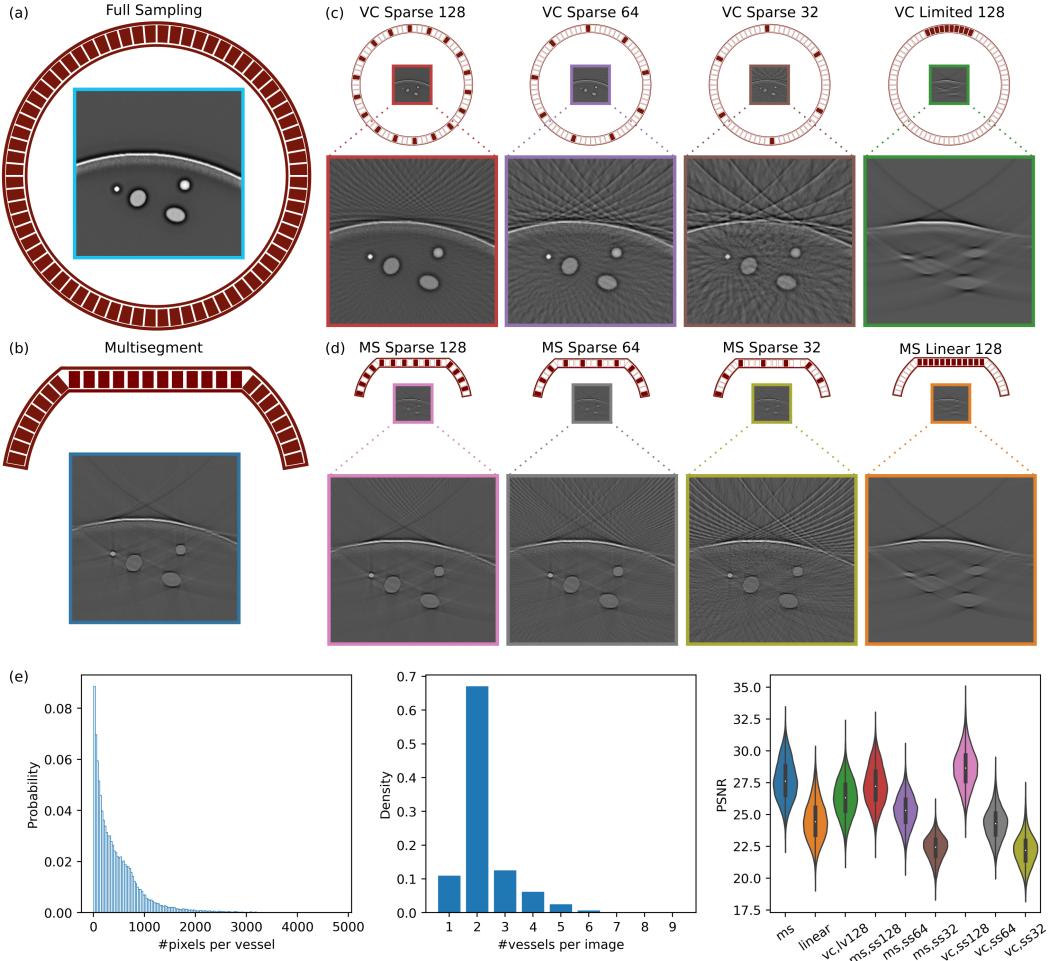


Figure 2: Overview of the simulated data. (a) Virtual circle array and an image reconstructed using 1,024 transducer elements. (b) Multisegment array and the corresponding image reconstructed using combined linear and concave parts of the transducer array. (c) Uniform subsampling of virtual circle array with 128, 64 and 32 elements and limited view acquisition with reduced angular coverage (12 elements). (d) Uniform subsampling of multisegment array with 128, 64 and 32 elements and linear array acquisition (128 elements). (e) Vessel size distribution (pixels per vessel), number of vessels per image, and peak signal-to-noise ratio of full sampling compared to other reconstructions (x axis naming conventions are explained in Sec. 3.3). Transducer elements are shown as actively receiving (red) or off (white).

Table 1: Publicly available OA datasets, supported tasks, provided data format(s), size, and content. [3] contains 274 mice and 469 phantom slices. [15] has 10 mice with 10 frames (100 slices) at 27 different wavelengths and 20 phantom slices.

Dataset	tasks			image reconstruction	raw signal	size (>5k instances)	clinical data
	limited view	sparse sampling	pixel annotations				
[3]	✗	✓	✗	✓	✗	✗	✗
[15]	✗	✗	✗	✗	✓	✗	✗
MSFD (ours)	✓	✗	✗	✓	✓	✓	✓
SWFD (ours)	✓	✓	✗	✓	✓	✓	✓
SCD (ours)	✓	✓	✓	✓	✓	✓	✗
OADAT-mini (ours)	✓	✓	✓	✓	✓	✓	✗ ∪ ✓

inter-element pitch size, focusing and geometry of the active area. (iii) **Virtual circle limited 128** uses 128 consecutive elements (one eighth of a circle) out of 1,024.

### 3 Datasets

We present four datasets <sup>3</sup> <sup>4</sup> (two experimental, one simulated, one fully annotated subset) where each has several subcategories for the purpose of tackling different challenges present in the domain. Raw signal acquisition data that is used to reconstruct all images are also provided with the datasets. Experimental datasets also include details about the volunteer Fitzpatrick skin phototype [14], which relates to the amount of melanin pigment in the skin (see Appendix B for distribution and further details). We also display a comparative overview of publicly available and our proposed OA datasets in Table 1. Please refer to Tables 5, 6, 7, 8, and 9 for summaries of the file contents of the datasets in the Appendix.

#### 3.1 Multispectral forearm dataset

Multispectral forearm dataset (MSFD) is collected using multisegment array (Sec. 2.1) from nine volunteers at six different wavelengths (700, 730, 760, 780, 800, 850 nm) for both arms. Selected wavelengths are particularly aimed for spectral decomposition aiming to separate oxy- and deoxy-hemoglobin [32]. All wavelengths are acquired consecutively, yielding almost identical scene being captured for a given slice across different wavelengths with slight displacement errors. For each of the mentioned category 1,400 slices are captured, creating a sum of  $9 \times 6 \times 2 \times 1,400 = 151,200$  unique signal matrices.

From this data, using backprojection algorithm, we reconstruct (i) linear array images  $\text{MSFD}_{\text{linear}}$ , (ii) multisegment array images  $\text{MSFD}_{\text{ms}}$ , (iii) multisegment sparse 128 images (Sec.2.3),  $\text{MSFD}_{\text{ms,ss128}}$ , (iv) multisegment sparse 64 images (Sec.2.3),  $\text{MSFD}_{\text{ms,ss64}}$ , and (v) multisegment sparse 32 images (Sec.2.3),  $\text{MSFD}_{\text{ms,ss32}}$ , where each dataset has 151,200 images of  $256 \times 256$  pixel resolution; totaling to 756,000 image instances.

#### 3.2 Single wavelength forearm dataset

Single wavelength forearm dataset (SWFD) is collected using both multisegment and semi circle arrays (Sec. 2.1) from 14 volunteers at a single wavelength (1,064 nm) for both arms. The choice of the wavelength is based on maximizing penetration depth for excitation light source (laser) [36]. Out of the 14 volunteers, eight of them have also participated in the MSFD experiment and their unique identifiers match across the dataset files. For each array, volunteer, and arm, we acquired 1,400 slices, creating a sum of  $2 \times 14 \times 2 \times 1,400 = 78,400$  unique signal matrices. It is important to note that despite the data being acquired from the same volunteers, signals between multisegment array and semi circle array are not paired due to physical constraints.

<sup>3</sup>Link to our datasets: doi.org/10.3929/ethz-b-000551512

<sup>4</sup>Repository for accessing and reading datasets: github.com/berkanlafci/oadat

From this data, using backprojection algorithm, we reconstruct (i) linear array images,  $\text{SWFD}_{\text{linear}}$ , (ii) multisegment array images,  $\text{SWFD}_{\text{ms}}$ , (iii) semi circle array images,  $\text{SWFD}_{\text{sc}}$ , (iv) semi circle array limited 128 images (Sec.2.4),  $\text{SWFD}_{\text{sc},\text{lv}128}$ , (v) semi circle sparse 128 images (Sec.2.3),  $\text{SWFD}_{\text{sc},\text{ss}128}$ , (vi) semi circle sparse 64 images (Sec.2.3),  $\text{SWFD}_{\text{sc},\text{ss}64}$ , (vii) semi circle sparse 32 images (Sec.2.3),  $\text{SWFD}_{\text{sc},\text{ss}32}$ , (viii) multisegment sparse 128 images (Sec.2.3),  $\text{SWFD}_{\text{ms},\text{ss}128}$ , (ix) multisegment sparse 64 images (Sec.2.3),  $\text{SWFD}_{\text{ms},\text{ss}64}$ , and (x) multisegment sparse 32 images (Sec.2.3),  $\text{SWFD}_{\text{ms},\text{ss}32}$ , where each dataset has 39,200 images of  $256 \times 256$  pixel resolution; totaling to 392,000 image instances.

### 3.3 Simulated cylinders dataset

Simulated cylinders dataset (SCD) is a group of synthetically generated 20,000 forearm acoustic pressure maps that we heuristically produced based on a range of criteria we observed in experimental images. The acoustic pressure maps are generated with  $256 \times 256$  pixel resolution where skin curves and afterwards a random amount of ellipses with different intensity profiles are generated iteratively for a given image (see Fig. 2). We explain details for the simulation algorithm<sup>5</sup> for generating acoustic pressure map in Appendix E.

Based on the acoustic pressure map, we generate its annotation map with three labels, corresponding to background, vessels, and skin curve. For each acoustic pressure map, we generate signal matrices for the geometries of linear, multisegment and virtual circle arrays. Using linear and multisegment array signals, we use backprojection algorithm to reconstruct (i) linear array images,  $\text{SCD}_{\text{linear}}$ , and (ii) multisegment array images,  $\text{SCD}_{\text{ms}}$ , (iii) multisegment sparse 128 images (Sec.2.3),  $\text{SCD}_{\text{ms},\text{ss}128}$ , (iv) multisegment sparse 64 images (Sec.2.3),  $\text{SCD}_{\text{ms},\text{ss}64}$ , and (v) multisegment sparse 32 images (Sec.2.3),  $\text{SCD}_{\text{ms},\text{ss}32}$ . From virtual circle array signals, we use backprojection algorithm to reconstruct (vi) virtual circle images,  $\text{SCD}_{\text{vc}}$ , (vii) virtual circle limited 128 images (Sec.2.4),  $\text{SCD}_{\text{vc},\text{lv}128}$ , (viii) virtual circle sparse 128 images (Sec.2.3),  $\text{SCD}_{\text{vc},\text{ss}128}$ , (ix) virtual circle sparse 64 images (Sec.2.3),  $\text{SCD}_{\text{vc},\text{ss}64}$ , and (x) virtual circle sparse 32 images (Sec.2.3),  $\text{SCD}_{\text{vc},\text{ss}32}$ , where each dataset has 20,000 images of  $256 \times 256$  pixel resolution; totaling to 200,000 image instances. All ten image reconstruction variations of SCD have corresponding pairs for each of the 20k image; i.e., produced from the same acoustic pressure map.

### 3.4 OADAT-mini dataset

Using a subset of 100 signal and corresponding reconstruction instances from each of the previously mentioned datasets, we also present OADAT-mini, which is a fragment of OADAT that is significantly smaller yet comprehensive for understanding the contents of OADAT. In addition, OADAT-mini contains manual annotation maps for vessels in the reconstructed images.

## 4 Tasks

Based on the datasets presented in Sec. 3, we define a list of experiments related to image translation to overcome (i) sparse sampling and (ii) limited view artifacts, and semantic segmentation of images.

### 4.1 Image translation

Through a list of permutations of our datasets, we can define several pairs of varying difficulty of image translation experiments where the target images are also available (see Table 2). We present sparse sampling and limited view reconstructions of SWFD, MSFD and SCD for all transducer arrays. Sparse sampling correction experiments learn mapping functions listed in Table 2, where the function notations denote the dataset used, the task of sparse sampling (ss) correction from the given number of elements used for image reconstruction and the array that is used to generate the input. Limited view correction experiments learn mapping functions, listed in Table 2, where the function notations denote the dataset used, the task of limited view (lv) correction from the given number of elements used for image reconstruction and the array that is used to generate the input.

---

<sup>5</sup>Python module for acoustic map simulation: [renkulab.io/gitlab/firat.ozdemir/oa-armsim](https://renkulab.io/gitlab/firat.ozdemir/oa-armsim).

Table 2: List of tasks and experiments we define on MSFD, SWFD, and SCD. Experiment names are built of (i) dataset name for translation tasks or seg for segmentation task, (ii) input data and corresponding number of active array elements; sparse sampling (ss), limited view (lv), virtual circle (vc), and (iii) input array type; semi circle (sc), virtual circle (vc), linear (li), multisegment (ms). Image data and annotation maps are represented with  $x$  and  $y$ , while predicted image and annotations are shown as  $x^*$  and  $y^*$ .

Limited view correction	Semantic segmentation
$f_{\text{MSFD\_lv128,li}}: x \sim \text{MSFD}_{\text{linear}} \rightarrow x^* \sim \text{MSFD}_{\text{ms}}$	$f_{\text{seg\_vc,vc}}: x \sim \text{SCD}_{\text{vc}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SWFD\_lv128,li}}: x \sim \text{SWFD}_{\text{linear}} \rightarrow x^* \sim \text{SWFD}_{\text{ms}}$	$f_{\text{seg\_lv128,li}}: x \sim \text{SCD}_{\text{linear}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SWFD\_lv128,sc}}: x \sim \text{SWFD}_{\text{sc},\text{lv128}} \rightarrow x^* \sim \text{SWFD}_{\text{sc}}$	$f_{\text{seg\_lv128,vc}}: x \sim \text{SCD}_{\text{vc},\text{lv128}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SCD\_lv128,li}}: x \sim \text{SCD}_{\text{linear}} \rightarrow x^* \sim \text{SCD}_{\text{vc}}$	$f_{\text{seg\_ss128,vc}}: x \sim \text{SCD}_{\text{vc},\text{ss128}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SCD\_lv128,vc}}: x \sim \text{SCD}_{\text{vc},\text{lv128}} \rightarrow x^* \sim \text{SCD}_{\text{vc}}$	$f_{\text{seg\_ss64,vc}}: x \sim \text{SCD}_{\text{vc},\text{ss64}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SCD\_lv256,ms}}: x \sim \text{SCD}_{\text{ms}} \rightarrow x^* \sim \text{SCD}_{\text{vc}}$	$f_{\text{seg\_ss32,vc}}: x \sim \text{SCD}_{\text{vc},\text{ss32}} \rightarrow y^* \sim \text{labels}$
<b>Sparse sampling correction</b>	
$f_{\text{SWFD\_ss128,sc}}: x \sim \text{SWFD}_{\text{sc},\text{ss128}} \rightarrow x^* \sim \text{SWFD}_{\text{sc}}$	$f_{\text{seg\_ss128,ms}}: x \sim \text{SCD}_{\text{ms},\text{ss128}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SWFD\_ss64,sc}}: x \sim \text{SWFD}_{\text{sc},\text{ss64}} \rightarrow x^* \sim \text{SWFD}_{\text{sc}}$	$f_{\text{seg\_ss64,ms}}: x \sim \text{MSFD}_{\text{ms},\text{ss64}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SWFD\_ss32,sc}}: x \sim \text{SWFD}_{\text{sc},\text{ss32}} \rightarrow x^* \sim \text{SWFD}_{\text{sc}}$	$f_{\text{seg\_ss32,ms}}: x \sim \text{MSFD}_{\text{ms},\text{ss32}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SCD\_ss128,vc}}: x \sim \text{SCD}_{\text{vc},\text{ss128}} \rightarrow x^* \sim \text{SCD}_{\text{vc}}$	$f_{\text{seg\_MSFD\_lv128,ms}}: x \sim \text{MSFD}_{\text{linear}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SCD\_ss64,vc}}: x \sim \text{SCD}_{\text{vc},\text{ss64}} \rightarrow x^* \sim \text{SCD}_{\text{vc}}$	$f_{\text{seg\_MSFD\_ss128,ms}}: x \sim \text{MSFD}_{\text{ms},\text{ss128}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SCD\_ss32,vc}}: x \sim \text{SCD}_{\text{vc},\text{ss32}} \rightarrow x^* \sim \text{SCD}_{\text{vc}}$	$f_{\text{seg\_MSFD\_ss64,ms}}: x \sim \text{MSFD}_{\text{ms},\text{ss64}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SWFD\_ss128,ms}}: x \sim \text{SWFD}_{\text{ms},\text{ss128}} \rightarrow x^* \sim \text{SWFD}_{\text{ms}}$	$f_{\text{seg\_MSFD\_ss32,ms}}: x \sim \text{MSFD}_{\text{ms},\text{ss32}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SWFD\_ss64,ms}}: x \sim \text{SWFD}_{\text{ms},\text{ss64}} \rightarrow x^* \sim \text{SWFD}_{\text{ms}}$	$f_{\text{seg\_SWFD\_lv128,ms}}: x \sim \text{SWFD}_{\text{linear}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SWFD\_ss32,ms}}: x \sim \text{SWFD}_{\text{ms},\text{ss32}} \rightarrow x^* \sim \text{SWFD}_{\text{ms}}$	$f_{\text{seg\_SWFD\_lv128,sc}}: x \sim \text{SWFD}_{\text{sc},\text{lv128}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SCD\_ss128,ms}}: x \sim \text{SCD}_{\text{ms},\text{ss128}} \rightarrow x^* \sim \text{SCD}_{\text{vc}}$	$f_{\text{seg\_SWFD\_ms,ms}}: x \sim \text{SWFD}_{\text{ms}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SCD\_ss64,ms}}: x \sim \text{SCD}_{\text{ms},\text{ss64}} \rightarrow x^* \sim \text{SCD}_{\text{vc}}$	$f_{\text{seg\_SWFD\_sc}}: x \sim \text{SWFD}_{\text{sc}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SCD\_ss32,ms}}: x \sim \text{SCD}_{\text{ms},\text{ss32}} \rightarrow x^* \sim \text{SCD}_{\text{vc}}$	$f_{\text{seg\_SWFD\_ss128,sc}}: x \sim \text{SWFD}_{\text{sc},\text{ss128}} \rightarrow y^* \sim \text{labels}$
$f_{\text{MSFD\_ss128,ms}}: x \sim \text{MSFD}_{\text{ms},\text{ss128}} \rightarrow x^* \sim \text{MSFD}_{\text{ms}}$	$f_{\text{seg\_SWFD\_ss64,sc}}: x \sim \text{SWFD}_{\text{sc},\text{ss64}} \rightarrow y^* \sim \text{labels}$
$f_{\text{MSFD\_ss64,ms}}: x \sim \text{MSFD}_{\text{ms},\text{ss64}} \rightarrow x^* \sim \text{MSFD}_{\text{ms}}$	$f_{\text{seg\_SWFD\_ss32,sc}}: x \sim \text{SWFD}_{\text{sc},\text{ss32}} \rightarrow y^* \sim \text{labels}$
$f_{\text{MSFD\_ss32,ms}}: x \sim \text{MSFD}_{\text{ms},\text{ss32}} \rightarrow x^* \sim \text{MSFD}_{\text{ms}}$	$f_{\text{seg\_SWFD\_ss128,ms}}: x \sim \text{SWFD}_{\text{ms},\text{ss128}} \rightarrow y^* \sim \text{labels}$
	$f_{\text{seg\_SWFD\_ss64,ms}}: x \sim \text{SWFD}_{\text{ms},\text{ss64}} \rightarrow y^* \sim \text{labels}$
	$f_{\text{seg\_SWFD\_ss32,ms}}: x \sim \text{SWFD}_{\text{ms},\text{ss32}} \rightarrow y^* \sim \text{labels}$

## 4.2 Semantic segmentation

SCD includes pixel annotations for skin curve, vessels and background. In addition to segmentation of these structures on the ideal reconstructions  $\text{SCD}_{\text{vc}}$ , we define this task on sparse sampling and limited view reconstructions that contain the relevant artifacts encountered in experimental data. Accordingly, we compose the nine segmentation experiments listed in Table 2, where the function notations denote the task segmentation (seg), type of the reconstructed input being used (virtual circle (vc), limited view (lv), and sparse sampling (ss)) and the array that is used to generate the input. All data is generated from SCD and the objective is to match the ground truth annotations of the acoustic pressure map. Different than SCD, experimental datasets under OADAT-mini include pixel annotations for vessels and consist of 14 segmentation experiments, also listed in Table 2.

## 5 Experiments and results

For all experiments we standardize the architecture that we based on UNet [34]. Specifically, we adopt the five spatial feature abstraction levels and use skip connections to concatenate with features of matching spatial dimension along the upsampling path. However, we make several distinct design choices that vary from vanilla UNet. First, we use attention gates [29] at the end of each skip connection. Second, we opt for residual convolutional blocks with batch normalization [16] at each level, shown in Fig. 3. Third, we use 2D bilinear upsampling instead of deconvolutions. Finally, we use half the number of convolutional kernels at each layer; e.g., start with 32 convolutional filters as opposed to 64. Full schematic as well as other implementation details are discussed in Appendix C. We refer to our modified UNet architecture as modUNet hereon.

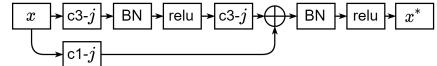


Figure 3: Residual convolutional block with batch normalization (BN).  $c_{i-j}$  conv. layer have  $i \times i$  kernels and  $j$  filters.

## 5.1 Data split and preprocessing

We standardize how we split each dataset into training and test sets regardless of the task in order to ensure consistency in our and future experiments. Out of the nine volunteers in MSFD, we use five for training (IDs: 2, 5, 6, 7, 9), one for validation (ID: 10) and three for testing (IDs: 11, 14, 15). Out of the 14 volunteers in SWFD, we use eight for training (IDs: 1, 2, 3, 4, 5, 6, 7, 8), one for validation (ID: 9) and five for testing (IDs: 10, 11, 12, 13, 14). Out of the 20k slices in SCD, we use the first 14k for training, following 1k for validation, and the last 5k for testing. For each experiment conducted on OADAT-mini, we use the first 75 samples for training, next 5 for validation and the last 20 for quantitative evaluation. This translates to six times the numbers for MSFD-mini, where there are 100 samples for each of the six wavelengths. As before, we conduct the MSFD-mini segmentation experiments as a single modUNet attempting segment any of the given six wavelength samples. As a preprocessing step, all data instances (except for annotation maps) are independently scaled by their maximum and then clipped at a minimum value of  $-0.2$  [8].

## 5.2 Results

Table 3: Image translation results of the proposed modUNet model reported as mean  $\pm$  std. Each row corresponds to the results of the experiment where input data is identified through (i) the input data and corresponding number of active transducer elements; sparse sampling (ss), limited view (lv) and (ii) the array type used for input; semi circle (sc), virtual circle (vc), linear (li), multisegment (ms).

		MAE	RMSE	SSIM	PSNR
<b>SCD</b>					
Limited view	lv128,li	$0.005 \pm 1.3e-3$	$0.012 \pm 3.3e-3$	$0.978 \pm 7.2e-3$	$39.909 \pm 2.15$
	lv128,vc	$0.005 \pm 1.3e-3$	$0.013 \pm 3.4e-3$	$0.978 \pm 6.2e-3$	$39.853 \pm 2.18$
	lv256,ms	$0.005 \pm 1.2e-3$	$0.009 \pm 3.0e-3$	$0.984 \pm 6.1e-3$	$42.459 \pm 2.52$
Sparse view	ss128,vc	$0.004 \pm 1.1e-3$	$0.008 \pm 2.0e-3$	$0.985 \pm 6.1e-3$	$44.068 \pm 2.12$
	ss64,vc	$0.005 \pm 1.3e-3$	$0.011 \pm 2.9e-3$	$0.981 \pm 6.9e-3$	$41.402 \pm 2.25$
	ss32,vc	$0.006 \pm 1.5e-3$	$0.013 \pm 3.7e-3$	$0.973 \pm 8.7e-3$	$39.309 \pm 2.23$
	ss128,ms	$0.005 \pm 1.1e-3$	$0.010 \pm 3.0e-3$	$0.984 \pm 6.1e-3$	$42.345 \pm 2.39$
	ss64,ms	$0.005 \pm 1.2e-3$	$0.010 \pm 3.0e-3$	$0.982 \pm 6.7e-3$	$41.525 \pm 2.19$
	ss32,ms	$0.005 \pm 1.3e-3$	$0.011 \pm 3.0e-3$	$0.979 \pm 7.3e-3$	$40.678 \pm 2.10$
<b>SWFD</b>					
Limited view	lv128,li	$0.028 \pm 1.6e-2$	$0.039 \pm 2.0e-2$	$0.613 \pm 1.4e-1$	$29.397 \pm 4.75$
	lv128,sc	$0.016 \pm 1.2e-2$	$0.021 \pm 1.5e-2$	$0.811 \pm 1.3e-1$	$36.791 \pm 5.30$
Sparse view	ss128,sc	$0.015 \pm 1.2e-2$	$0.019 \pm 1.5e-2$	$0.863 \pm 1.0e-1$	$38.233 \pm 5.35$
	ss64,sc	$0.019 \pm 1.5e-2$	$0.024 \pm 1.8e-2$	$0.769 \pm 1.6e-1$	$35.605 \pm 5.22$
	ss32,sc	$0.021 \pm 1.6e-2$	$0.028 \pm 2.0e-2$	$0.693 \pm 2.0e-1$	$33.852 \pm 5.37$
	ss128,ms	$0.023 \pm 1.5e-2$	$0.028 \pm 1.9e-2$	$0.784 \pm 9.7e-2$	$33.764 \pm 4.53$
	ss64,ms	$0.029 \pm 1.9e-2$	$0.037 \pm 2.3e-2$	$0.636 \pm 1.5e-1$	$31.311 \pm 4.80$
	ss32,ms	$0.033 \pm 2.1e-2$	$0.042 \pm 2.5e-2$	$0.521 \pm 1.8e-1$	$29.813 \pm 5.11$
<b>MSFD</b>					
limited view	lv128,li	$0.023 \pm 1.1e-2$	$0.035 \pm 1.4e-2$	$0.640 \pm 1.4e-1$	$29.731 \pm 3.97$
	ss128,ms	$0.017 \pm 9.7e-3$	$0.022 \pm 1.2e-2$	$0.839 \pm 8.1e-2$	$35.798 \pm 3.84$
Sparse view	ss64,ms	$0.022 \pm 1.2e-2$	$0.029 \pm 1.5e-2$	$0.719 \pm 1.4e-1$	$33.104 \pm 4.02$
	ss32,ms	$0.026 \pm 1.4e-2$	$0.036 \pm 1.7e-2$	$0.608 \pm 1.8e-1$	$30.873 \pm 4.22$

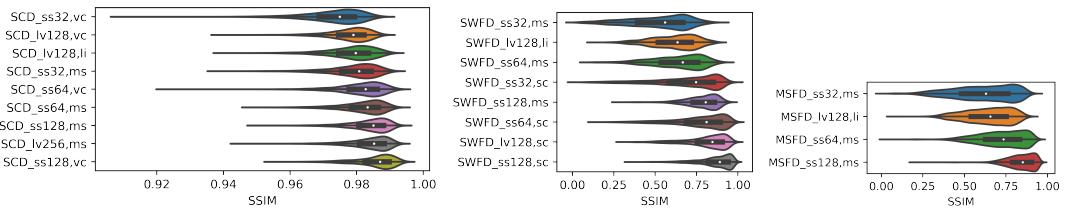


Figure 4: Distribution of modUNet SSIM performance on SCD(left), SWFD (middle) and MSFD (right) image translation experiments, sorted in ascending median test sample performance.

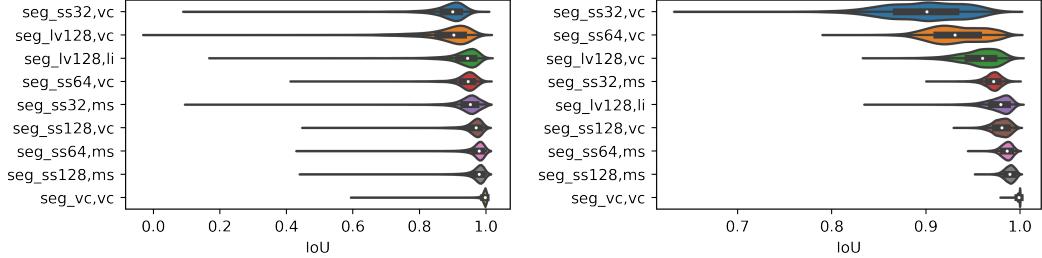


Figure 5: Distribution of modUNet IoU performance on SCD semantic segmentation experiments for vessel (left) and skin curve (right) labels, sorted in ascending median test sample performance.

We evaluate modUNet performance on the test sets using standard metrics. Namely, we report mean absolute error (MAE), root mean squared error (RMSE), structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR) for image translation experiments between modUNet predictions and targets. Segmentation task performance is reported using Dice coefficient (F1-score), intersection over union (IoU, i.e., Jaccard index) and 95-percentile Hausdorff distance (HD95) metrics between modUNet predictions and annotation maps for vessels and skin curves. HD95 is calculated in several steps: First, the set of pixels along the contour for each predicted (set A) and annotated (set B) target structures are found. For each set point, the closest point from the other set is determined based on  $l_2$  distance. Different from the standard Hausdorff distance, the 95-percentile distance value is taken as the directional distance from set A to B (and vice versa) instead of the maximum distance. Then the maximum of these two values is calculated as the symmetric HD95 for a given image. In OADAT-mini experiments, only vessel annotation maps are available. Since OADAT-mini consist of subsets of the other three datasets, we do not repeat image translation tasks. In Tables 3 & 4 we report modUNet results for mean and standard deviations aggregated over the corresponding test set images.

Using SSIM, we show performance across all our datasets in Fig. 4. Upon exploring the reason behind the long tails, we notice that most of the lower scores occur when acquisition noise and/or artifacts are more pronounced. Depending on the sample, this can imply either modUNet reduced the noise present in the target, or both input/output pair in the test set had low signal-to-noise ratio(SNR). Nevertheless, modUNet successfully corrects geometric distortions for limited view experiments. Given the low mean and standard deviations in MAE and RMSE, we can comment that modUNet can generalize well to previously unseen volunteer data. This is further corroborated with the narrow SSIM interquartile range in Fig. 4 violin plots.

Similarly, we plot the segmentation performance for IoU across different experiments on SCD in Fig. 5 for vessel and skin curve labels. While skin curve segmentation performance almost never drops below IoU of 0.80 for SCD, one can see that IoU can be drastically lower for vessel segmentation. We observed that this only happens when the size of vessel is small. For example, there are examples with ground truth vessels having as low as four pixels while the prediction has six, leading to an IoU of 0.6. As for experiments with OADAT-mini, vessel segmentation performance in experimental datasets are worse. Specifically, the worst scores are observed for MSFD-mini experiments. This is due to the movement artifact between different wavelengths of a given slice. All MSFD-mini instances have expert annotations for 800 nm wavelength. Slight movement across different wavelengths can yield poor quantitative metrics, particularly exacerbated when the observed vessels are small. Given the limited training and test sizes of OADAT-mini experiments, we believe that the quantitative results should be taken as a reference. The qualitative results in the Appendix can be more informative for gaining insight for modUNet performance when trained with a very limited amount of data. We provide qualitative results as well as conduct further analysis for all tasks in Appendix D.

## 6 Discussion

Major differences exist between simulated and experimental datasets. Even if the content is different, training and test samples of the simulated dataset are inherently sampled from the same distribution. On the other hand, experimental datasets feature shifts due to different volunteers being imaged, inherent noise from data acquisition system, and difference in directional sensitivity resulting from transducer alignment and positioning of the hand-held probe. Furthermore, despite the efforts to

Table 4: Segmentation results of our proposed modUNet model reported as mean  $\pm$  std. Each row corresponds to the results of the experiment where input data is identified through (i) the input data and corresponding number of active transducer elements; sparse sampling (ss), limited view (lv), virtual circle (vc), semi circle (sc), multisegment (ms) and (ii) the array type used for input; virtual circle (vc), semi circle (sc), multisegment (ms), and linear (li). SWFD- and MSFD-mini correspond to experiment conducted on OADAT-mini dataset.

		Dice		IoU		HD95	
		vessels	skin curve	vessels	skin curve	vessels	skin curve
<b>SCD</b>							
Full view	vc,vc	0.996 $\pm$ 9.8e-3	0.999 $\pm$ 8.7e-4	0.993 $\pm$ 1.8e-2	0.999 $\pm$ 1.7e-3	0.792 $\pm$ 1.0e+0	0.164 $\pm$ 2.0e+0
Limited view	lv128,li	0.963 $\pm$ 3.4e-2	0.988 $\pm$ 7.3e-3	0.931 $\pm$ 5.5e-2	0.976 $\pm$ 1.4e-2	2.734 $\pm$ 2.9e+0	2.196 $\pm$ 7.4e+0
	lv128,vc	0.933 $\pm$ 5.8e-2	0.978 $\pm$ 1.1e-2	0.878 $\pm$ 8.6e-2	0.957 $\pm$ 2.1e-2	4.310 $\pm$ 8.9e+0	7.372 $\pm$ 1.4e+1
	ss128,vc	0.980 $\pm$ 2.3e-2	0.990 $\pm$ 4.7e-3	0.961 $\pm$ 4.0e-2	0.980 $\pm$ 9.1e-3	2.499 $\pm$ 5.2e+0	3.897 $\pm$ 1.0e+1
	ss64,vc	0.967 $\pm$ 2.8e-2	0.965 $\pm$ 1.5e-2	0.937 $\pm$ 4.6e-2	0.933 $\pm$ 2.8e-2	2.738 $\pm$ 3.0e+0	9.634 $\pm$ 1.5e+1
	ss32,vc	0.938 $\pm$ 3.9e-2	0.946 $\pm$ 2.3e-2	0.886 $\pm$ 6.0e-2	0.899 $\pm$ 4.1e-2	3.074 $\pm$ 3.1e+0	12.220 $\pm$ 1.6e+1
	ss128,ms	0.983 $\pm$ 2.4e-2	0.994 $\pm$ 2.9e-3	0.968 $\pm$ 4.1e-2	0.989 $\pm$ 5.6e-3	2.289 $\pm$ 2.4e+0	5.311 $\pm$ 1.3e+1
Sparse view	ss64,ms	0.984 $\pm$ 2.4e-2	0.993 $\pm$ 3.1e-3	0.969 $\pm$ 4.1e-2	0.985 $\pm$ 6.1e-3	2.772 $\pm$ 8.2e+0	2.668 $\pm$ 8.5e+0
	ss32,ms	0.971 $\pm$ 2.8e-2	0.985 $\pm$ 4.8e-3	0.945 $\pm$ 4.5e-2	0.971 $\pm$ 9.2e-3	3.026 $\pm$ 7.0e+0	6.795 $\pm$ 1.5e+1
<b>SWFD-mini</b>							
Full view	sc,sc	0.857 $\pm$ 5.7e-2	N/A	0.755 $\pm$ 8.5e-2	N/A	17.225 $\pm$ 2.8e+1	N/A
	ms,ms	0.842 $\pm$ 5.9e-2	N/A	0.732 $\pm$ 8.6e-2	N/A	17.218 $\pm$ 2.6e+1	N/A
Limited view	lv128,li	0.794 $\pm$ 6.4e-2	N/A	0.662 $\pm$ 8.3e-2	N/A	28.685 $\pm$ 3.9e+1	N/A
	lv128,sc	0.843 $\pm$ 5.5e-2	N/A	0.732 $\pm$ 7.8e-2	N/A	24.444 $\pm$ 2.6e+1	N/A
	ss128,sc	0.864 $\pm$ 4.3e-2	N/A	0.763 $\pm$ 5.8e-2	N/A	23.146 $\pm$ 3.0e+1	N/A
	ss64,sc	0.841 $\pm$ 1.1e-1	N/A	0.737 $\pm$ 1.3e-1	N/A	20.216 $\pm$ 2.8e+1	N/A
	ss32,sc	0.864 $\pm$ 4.1e-2	N/A	0.762 $\pm$ 6.3e-2	N/A	26.832 $\pm$ 3.0e+1	N/A
	ss128,ms	0.836 $\pm$ 5.8e-2	N/A	0.722 $\pm$ 8.3e-2	N/A	17.909 $\pm$ 2.8e+1	N/A
Sparse view	ss64,ms	0.837 $\pm$ 5.6e-2	N/A	0.723 $\pm$ 8.0e-2	N/A	15.811 $\pm$ 2.3e+1	N/A
	ss32,ms	0.809 $\pm$ 6.0e-2	N/A	0.684 $\pm$ 8.2e-2	N/A	20.603 $\pm$ 3.1e+1	N/A
	ss128,ms	0.563 $\pm$ 1.0e-1	N/A	0.400 $\pm$ 1.1e-1	N/A	14.312 $\pm$ 1.9e+1	N/A
<b>MSFD-mini</b>	lv128,li	0.474 $\pm$ 1.3e-1	N/A	0.320 $\pm$ 1.2e-1	N/A	20.134 $\pm$ 3.3e+1	N/A
	ss128,ms	0.572 $\pm$ 1.2e-1	N/A	0.411 $\pm$ 1.3e-1	N/A	19.716 $\pm$ 2.1e+1	N/A
Sparse view	ss64,ms	0.639 $\pm$ 1.4e-1	N/A	0.485 $\pm$ 1.5e-1	N/A	11.499 $\pm$ 1.7e+1	N/A

avoid corrupted acquisitions during the data collection, experimental datasets still contain samples with relatively low signal-to-noise ratio. Such samples are expected to yield reduced performance metrics for image translation tasks due to significant mismatch between the predicted and noisy target images. In a clinical setting, a medical expert typically repeats an acquisition if they deem the signal quality is significantly lower than expected. However, beyond this subjective filtering step, one needs to make sure that even the worst results are either sufficiently good or their poor performance can be attributed to a cause. Accordingly, we further analyze some of the worst samples in the Appendix and believe that this should be a standard for future work. Provided dataset is limited to one body part of volunteers without any known health issues. The image reconstruction methods for OA imaging can be applied on any other body part or imaging setup as they solve the same physical inverse problem. The image translation algorithms for sparse acquisition and limited view problems can be adapted for different devices and acquisitions by another clinician/technician at another center, as the streak artifacts originating from sparse acquisition and limited view follow the same pattern.

We envision that future research will tackle additional challenges such as unsupervised or weakly supervised domain adaptation across the datasets provided in this work. There are initial studies to correct limited view artifacts in OA using transfer learning between simulated and experimental datasets after domain adaptation [19]. Similarly, transfer learning between simulated and experimental domains can enhance segmentation performance of the vessels and skin curve in clinical images. Using properties of the detected tissues for fluence correction and heterogeneous SoS image reconstructions would then yield more accurate and quantitative images [6]. We anticipate additional contributions in the field of representation learning using OADAT, e.g. through self-supervised learning, could allow overcoming bottlenecks for specialized downstream tasks with limited amount of task-specific available data. The multispectral datasets with paired images across multiple wavelength acquisitions are expected to facilitate the investigation of generative modeling of multispectral signals from a given wavelength. We also anticipate future work to explore novel multispectral unmixing approaches using MSFD, enabling more accurate quantification of oxygenation, melanin and lipid

content of the tissues. Finally, the provided raw signal data, not available in commercial devices, has a high value for data-driven research. For example, it can serve to benchmark methodologies in image reconstruction e.g., based on variational networks with loop unrolling [40], as well as ultra-fast imaging through adaptive channel sampling.

## 7 Conclusion

In this work, we provide experimental and synthetic clinical OA data covering a large variety of examples to be used as common ground to compare established and new data processing methods. The datasets correspond to samples from volunteers of varying skin types, different OA array geometries, and several illumination wavelengths. A subset of this experimental data is annotated by an expert. The dataset is supplemented with simulated samples containing ground truth acoustic pressure maps, annotations, and combine pairs of samples reconstructed with different OA array geometries. We define a set of 44 experiments tackling major challenges in the OA field and provide reconstructions of the images under these scenarios along with their corresponding ground truths. We propose and release 44<sup>6</sup> trained neural networks that achieve a good performance for all these examples which can be used as baselines for future improvements. Additional problems can further be defined with the data provided, such as the effects of random sparse sampling or the presence of noise in the signal matrices prior to reconstruction. We believe that these datasets and benchmarks will play a significant role in fostering coordinated efforts to solve major challenges in OA imaging.

## Broader Impact Statement

The dataset is anonymized by randomly assigning identity numbers for each volunteer. The true identities will not be accessible by any third party now or in the future. The dataset should not be used to draw any medical conclusions. The purpose of the dataset is to help researchers to develop new image processing methods and provide benchmark scores.

## Author Contributions

B.L. setup and acquired raw datasets, B.L. and F.O. analyzed and parsed datasets, F.O. conducted numerical experiments based on the defined tasks. All authors reviewed the manuscript.

## Acknowledgments

This work was supported by Swiss Data Science Center (grant C19-04).

## References

- [1] Frederic M. Brochu, Joanna Brunker, James Joseph, Michal R. Tomaszewski, Stefan Morscher, and Sarah E. Bohndiek. Towards quantitative evaluation of tissue absorption coefficients using light fluence correction in optoacoustic tomography. *IEEE Transactions on Medical Imaging*, 36(1):322–331, 2017. doi: 10.1109/TMI.2016.2607199.
- [2] Reni Butler, Philip T. Lavin, F. Lee Tucker, Lora D. Barke, Marcela Böhm-Vélez, Stamatia Destounis, Stephen R. Grobmyer, Janine Katzen, Kenneth A. Kist, Erini V. Makariou, Kathy J. Schilling, Catherine A. Young, Basak E. Dogan, and Erin I. Neuschler. Optoacoustic breast imaging: Imaging-pathology correlation of optoacoustic features in benign and malignant breast masses. *American Journal of Roentgenology*, 211(5):1155–1170, 2018. doi: 10.2214/AJR.17.18435. URL <https://doi.org/10.2214/AJR.17.18435>. PMID: 30106610.
- [3] Neda Davoudi, Xosé Luís Deán-Ben, and Daniel Razansky. Deep learning optoacoustic tomography with sparse data. *Nature Machine Intelligence*, 1(10):453–460, Oct 2019.

<sup>6</sup>Pretrained model weights and various scripts to train and evaluate modUNet are available at <https://renkulab.io/gitlab/firat.ozdemir/oadat-evaluate>.

- ISSN 2522-5839. doi: 10.1038/s42256-019-0095-3. URL <https://doi.org/10.1038/s42256-019-0095-3>.
- [4] Neda Davoudi, Berkcan Lafci, Ali Özbek, Xosé Luís Deán-Ben, and Daniel Razansky. Deep learning of image- and time-domain data enhances the visibility of structures in optoacoustic tomography. *Opt. Lett.*, 46(13):3029–3032, Jul 2021. doi: 10.1364/OL.424571. URL <http://opg.optica.org/ol/abstract.cfm?URI=ol-46-13-3029>.
  - [5] Xosé Luís Deán-Ben, Andreas Buehler, Vasilis Ntziachristos, and Daniel Razansky. Accurate model-based reconstruction algorithm for three-dimensional optoacoustic tomography. *IEEE Transactions on Medical Imaging*, 31:1922–1928, 2012.
  - [6] X. Luís Deán-Ben, Vasilis Ntziachristos, and Daniel Razansky. Effects of small variations of speed of sound in optoacoustic tomographic imaging. *Medical Physics*, 41(7):073301, 2014. doi: <https://doi.org/10.1111/1.4875691>. URL <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1111/1.4875691>.
  - [7] Xosé Luís Deán-Ben and Daniel Razansky. Optoacoustic imaging of the skin. *Experimental Dermatology*, 30(11):1598–1609, 2021. doi: <https://doi.org/10.1111/exd.14386>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/exd.14386>.
  - [8] Lu Ding, X Luís Deán-Ben, Christian Lutzweiler, Daniel Razansky, and Vasilis Ntziachristos. Efficient non-negative constrained model-based inversion in optoacoustic tomography. *Physics in Medicine and Biology*, 60(17):6733–6750, aug 2015. doi: 10.1088/0031-9155/60/17/6733. URL <https://doi.org/10.1088/0031-9155/60/17/6733>.
  - [9] Gael Diot, Stephan Metz, Aurelia Noske, Evangelos Liapis, Barbara Schroeder, Saak V. Ovsepian, Reinhard Meier, Ernst Rummery, and Vasilis Ntziachristos. Multispectral Optoacoustic Tomography (MSOT) of Human Breast Cancer. *Clinical Cancer Research*, 23(22):6912–6922, 11 2017. ISSN 1078-0432. doi: 10.1158/1078-0432.CCR-16-3200. URL <https://doi.org/10.1158/1078-0432.CCR-16-3200>.
  - [10] Janek Gröhl, Lina Hacker, Ben T. Cox, Kris K. Dreher, Stefan Morscher, Avotra Rakotondrainibe, François Varay, Lawrence C.M. Yip, William C. Vogt, and Sarah E. Bohndiek. The ipasc data format: A consensus data format for photoacoustic imaging. *Photoacoustics*, 26:100339, 2022. ISSN 2213-5979. doi: <https://doi.org/10.1016/j.pacs.2022.100339>. URL <https://www.sciencedirect.com/science/article/pii/S2213597922000106>.
  - [11] E. Grüneisen. Theorie des festen zustandes einatomiger elemente, January 1912. URL <https://doi.org/10.1002/andp.19123441202>.
  - [12] Lisanne Grünherz, Epameinondas Gousopoulos, Carlotta Barbon, Semra Uyulmaz, Berkcan Lafci, Daniel Razansky, Andreas Boss, Pietro Giovanoli, and Nicole Lindenblatt. Preoperative mapping of lymphatic vessels by multispectral optoacoustic tomography. *Lymphatic Research and Biology*, 0(0):null, 2022. doi: 10.1089/lrb.2021.0067. URL <https://doi.org/10.1089/lrb.2021.0067>. PMID: 35230197.
  - [13] Steven Guan, Amir A. Khan, Siddhartha Sikdar, and Parag V. Chitnis. Limited-view and sparse photoacoustic tomography for neuroimaging with deep learning. *Scientific Reports*, 10(1):8510, May 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-65235-2. URL <https://doi.org/10.1038/s41598-020-65235-2>.
  - [14] Vishal Gupta and Vinod Kumar Sharma. Skin typing: Fitzpatrick grading and others. *Clinics in Dermatology*, 37(5):430–436, 2019. ISSN 0738-081X. doi: <https://doi.org/10.1016/j.clindermatol.2019.07.010>. URL <https://www.sciencedirect.com/science/article/pii/S0738081X1930121X>. The Color of Skin.
  - [15] Shan Huang, Andreas Blutke, Annette Feuchtinger, Uwe Klemm, Robby Zachariah Tom, Susanna M Hofmann, Andre C Stiel, and Vasilis Ntziachristos. Functional multispectral optoacoustic tomography imaging of hepatic steatosis development in mice. *EMBO Molecular Medicine*, 13(9):e13490, 2021. doi: <https://doi.org/10.15252/emmm.202013490>. URL <https://www.embopress.org/doi/abs/10.15252/emmm.202013490>.

- [16] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.
- [17] Jørgen Arendt Jensen. Medical ultrasound imaging. *Progress in Biophysics and Molecular Biology*, 93(1):153–165, 2007. ISSN 0079-6107. doi: <https://doi.org/10.1016/j.biopharmolbio.2006.07.025>. URL <https://www.sciencedirect.com/science/article/pii/S0079610706000861>. Effects of ultrasound and infrasound relevant to human health.
- [18] Angelos Karlas, Michael Kallmayer, Michael Bariotakis, Nikolina-Alexia Fasoula, Evangelos Liapis, Fabien Hyafil, Jaroslav Pelisek, Moritz Wildgruber, Hans-Henning Eckstein, and Vasilis Ntziachristos. Multispectral optoacoustic tomography of lipid and hemoglobin contrast in human carotid atherosclerosis. *Photoacoustics*, 23:100283, 2021. ISSN 2213-5979. doi: <https://doi.org/10.1016/j.pacs.2021.100283>. URL <https://www.sciencedirect.com/science/article/pii/S2213597921000446>.
- [19] Anna Klimovskaia, Berkan Lafci, Firat Ozdemir, Neda Davoudi, Xose Luis Deán-Ben, Fernando Perez-Cruz, and Daniel Razansky. Signal domain learning approach for optoacoustic image reconstruction from limited view data. In *Medical Imaging with Deep Learning*, 2022. URL <https://openreview.net/forum?id=9N0yrfUBtx1>.
- [20] Ferdinand Knieling, Clemens Neufert, Arndt Hartmann, Jing Claussen, Alexander Urich, Cornelia Egger, Marcel Vetter, Sarah Fischer, Lukas Pfeifer, Alexander Hagel, Christian Kielißch, Rüdiger S. Götz, Dane Wildner, Matthias Engel, Jens Röther, Wolfgang Uter, Jürgen Siebler, Raja Atreya, Wolfgang Rascher, Deike Strobel, Markus F. Neurath, and Maximilian J. Waldner. Multispectral optoacoustic tomography for assessment of crohn’s disease activity. *New England Journal of Medicine*, 376(13):1292–1294, 2017. doi: 10.1056/NEJMc1612455. URL <https://doi.org/10.1056/NEJMc1612455>. PMID: 28355498.
- [21] Berkan Lafci, Elena Merčep, Joaquin L. Herraiz, Xosé Luís Deán-Ben, and Daniel Razansky. Noninvasive multiparametric characterization of mammary tumors with transmission-reflection optoacoustic ultrasound. *Neoplasia*, 22(12):770–777, 2020. ISSN 1476-5586. doi: <https://doi.org/10.1016/j.neo.2020.10.008>. URL <https://www.sciencedirect.com/science/article/pii/S1476558620301639>.
- [22] Berkan Lafci, Elena Merčep, Stefan Morscher, Xosé Luís Deán-Ben, and Daniel Razansky. Deep learning for automatic segmentation of hybrid optoacoustic ultrasound (opus) images. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 68(3):688–696, 2021. doi: 10.1109/TUFFC.2020.3022324.
- [23] Hengrong Lan, Daohuai Jiang, Changchun Yang, and Fei Gao. Y-net: A hybrid deep learning reconstruction framework for photoacoustic imaging in vivo, 2019. URL <https://arxiv.org/abs/1908.00975>.
- [24] Ramona De Luca, Tommaso Dattoma, Leonardo Forzoni, Jeffrey Bamber, Paolo Palchetti, and Alessandro Gubbini. Diagnostic ultrasound probes: a typology and overview of technologies. *Current Directions in Biomedical Engineering*, 4(1):49–53, 2018. doi: [doi:10.1515/cdbme-2018-0013](https://doi.org/10.1515/cdbme-2018-0013). URL <https://doi.org/10.1515/cdbme-2018-0013>.
- [25] Subhamoy Mandal, Xosé Luís Deán-Ben, and Daniel Razansky. Visual quality enhancement in optoacoustic tomography using active contour segmentation priors. *IEEE Transactions on Medical Imaging*, 35(10):2209–2217, 2016. doi: 10.1109/TMI.2016.2553156.
- [26] Srirang Manohar and Maura Dantuma. Current and future trends in photoacoustic breast imaging. *Photoacoustics*, 16:100134, 2019. ISSN 2213-5979. doi: <https://doi.org/10.1016/j.pacs.2019.04.004>. URL <https://www.sciencedirect.com/science/article/pii/S2213597918300442>.
- [27] Elena Merčep, Neal C. Burton, Jing Claussen, and Daniel Razansky. Whole-body live mouse imaging by hybrid reflection-mode ultrasound and optoacoustic tomography. *Opt. Lett.*, 40(20):4643–4646, Oct 2015. doi: 10.1364/OL.40.004643. URL <http://opg.optica.org/o1/abstract.cfm?URI=o1-40-20-4643>.

- [28] Elena Merčep, Xosé Luís Deán-Ben, and Daniel Razansky. Combined pulse-echo ultrasound and multispectral optoacoustic tomography with a multi-segment detector array. *IEEE Transactions on Medical Imaging*, 36(10):2129–2137, 2017. doi: 10.1109/TMI.2017.2706200.
- [29] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Matthias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.
- [30] Ali Ozbek, X. L. Deán-Ben, and Daniel Razansky. Realtime parallel back-projection algorithm for three-dimensional optoacoustic imaging devices. In Vasilis Ntziachristos and Charles P. Lin, editors, *Opto-Acoustic Methods and Applications*, volume 8800, pages 90 – 95. International Society for Optics and Photonics, SPIE, 2013. doi: 10.1117/12.2033376. URL <https://doi.org/10.1117/12.2033376>.
- [31] Ali Özbek, Xosé Luís Deán-Ben, and Daniel Razansky. Optoacoustic imaging at kilohertz volumetric frame rates. *Optica*, 5(7):857–863, Jul 2018. doi: 10.1364/OPTICA.5.000857. URL <http://opg.optica.org/optica/abstract.cfm?URI=optica-5-7-857>.
- [32] Valeriya Perekatova, Pavel Subochev, Mikhail Kleshnin, and Ilya Turchin. Optimal wavelengths for optoacoustic measurements of blood oxygen saturation in biological tissues. *Biomed. Opt. Express*, 7(10):3979–3995, Oct 2016. doi: 10.1364/BOE.7.003979. URL <http://opg.optica.org/boe/abstract.cfm?URI=boe-7-10-3979>.
- [33] Justine Robin, Richard Rau, Berkan Lafci, Aileen Schroeter, Michael Reiss, Xosé-Luís Deán-Ben, Orcun Goksel, and Daniel Razansky. Hemodynamic response to sensory stimulation in mice: Comparison between functional ultrasound and optoacoustic imaging. *NeuroImage*, 237: 118111, 2021. ISSN 1053-8119. doi: <https://doi.org/10.1016/j.neuroimage.2021.118111>. URL <https://www.sciencedirect.com/science/article/pii/S1053811921003888>.
- [34] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [35] Melanie Schellenberg, Kris K. Dreher, Niklas Holzwarth, Fabian Isensee, Annika Reinke, Nicholas Schreck, Alexander Seitl, Minu D. Tizabi, Lena Maier-Hein, and Janek Gröhl. Semantic segmentation of multispectral photoacoustic images using deep learning. *Photoacoustics*, 26:100341, 2022. ISSN 2213-5979. doi: <https://doi.org/10.1016/j.pacs.2022.100341>. URL <https://www.sciencedirect.com/science/article/pii/S2213597922000118>.
- [36] Arunima Sharma, Srishti Srishti, Vijitha Periyasamy, and Manojit Pramanik. Photoacoustic imaging depth comparison at 532-, 800-, and 1064-nm wavelengths: Monte Carlo simulation and experimental validation. *Journal of Biomedical Optics*, 24(12):1 – 10, 2019. doi: 10.1117/1.JBO.24.12.121904. URL <https://doi.org/10.1117/1.JBO.24.12.121904>.
- [37] Idan Steinberg, David M. Huland, Ophir Vermesh, Hadas E. Frostig, Willemieke S. Tummers, and Sanjiv S. Gambhir. Photoacoustic clinical imaging. *Photoacoustics*, 14:77–98, Jun 2019. ISSN 2213-5979. doi: 10.1016/j.pacs.2019.05.001. URL <https://pubmed.ncbi.nlm.nih.gov/31293884/>. 31293884[pmid].
- [38] Jimmy L. Su, Bo Wang, Katheryne E. Wilson, Carolyn L. Bayer, Yun-Sheng Chen, Seungsoo Kim, Kimberly A. Homan, and Stanislav Y. Emelianov. Advances in clinical and biomedical applications of photoacoustic imaging. *Expert opinion on medical diagnostics*, 4(6):497–510, Nov 2010. ISSN 1753-0067. doi: 10.1517/17530059.2010.529127. URL <https://pubmed.ncbi.nlm.nih.gov/21344060/>. 21344060[pmid].
- [39] Stratis Tzoumas and Vasilis Ntziachristos. Spectral unmixing techniques for optoacoustic imaging of tissue pathophysiology. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 375(2107):20170262, Nov 2017. ISSN 1471-2962. doi: 10.1098/rsta.2017.0262. URL <https://pubmed.ncbi.nlm.nih.gov/29038385/>. 29038385[pmid].
- [40] Valery Vishnevskiy, Richard Rau, and Orcun Goksel. Deep variational networks with exponential weighting for learning computed tomography. In Dinggang Shen, Tianming Liu, Terry M. Peters, Lawrence H. Staib, Caroline Essert, Sean Zhou, Pew-Thian Yap, and Ali Khan, editors,

*Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pages 310–318, Cham, 2019. Springer International Publishing. ISBN 978-3-030-32226-7.

- [41] Çağla Özsoy, Ali Özbek, Michael Reiss, Xosé Luís Deán-Ben, and Daniel Razansky. Ultrafast four-dimensional imaging of cardiac mechanical wave propagation with sparse optoacoustic sensing. *Proceedings of the National Academy of Sciences*, 118(45):e2103979118, 2021. doi: 10.1073/pnas.2103979118. URL <https://www.pnas.org/doi/abs/10.1073/pnas.2103979118>.

## A Optoacoustic imaging background

Optoacoustic (OA) imaging setups contain light source, transducer array, data acquisition system and workstation PC. Usually, nanosecond-duration pulsed lasers are used as light sources. Alternatively, light emitting diodes (LEDs) can be used with high repetition rates (shorter pulses) as cheaper alternatives to lasers. The excitation of objects/tissues with short, pulsed light sources results in heating and expansion (thermoelastic expansion) of the materials. This expansion generates OA signals (US waves) that propagate to various directions in a 3D imaging medium (see A.1 for detailed explanation of pressure waves). Transducer arrays are used to capture the propagating waves in the imaging medium. Piezocomposite elements in these arrays convert mechanical pressure waves into electrical signals (see A.2 for transducer array specifications). After the detection and conversion, the electrical signals are digitized by the data acquisition systems at the defined sampling rate. The signals are acquired and digitized by all the elements in a transducer array simultaneously. The OA acquisition generates time domain (raw) signals acquired from each transducer element. Image reconstruction is performed by solving the inverse problem of signal to image domain conversion. In this study, we used backprojection method to reconstruct OA images (see A.3 for the detailed explanation of the method and used algorithms). Overall data acquisition system and experimental design are explained in A.4.

### A.1 Optoacoustic wave equations

OA imaging is based on thermoelastic expansion of the tissues which results in propagation of pressure waves in imaging medium depending on spatial and temporal changes. The OA wave equation can be written as follows

$$\frac{\partial^2 p(r,t)}{\partial t^2} - c^2 \nabla^2 p(r,t) = \Gamma H(r,t) \frac{\partial \delta(t)}{\partial t}$$

where  $r$  and  $t$  are the spatial and temporal variables, respectively.  $\Gamma$  is the Grüneisen constant [11].  $c$  stands for speed of sound.  $H(r, t)$  is the absorbed energy field based on the location and time of the sample.  $\delta(t)$  stands for temporal laser light intensity change based on the illumination.  $p(r, t)$  represents the pressure wave dependent on spatial and temporal variables. The Poisson solution of OA wave equation for pressure wave can be written as

$$p(r, t) = \frac{\Gamma}{4\pi c} \frac{\partial}{\partial t} \int_{S'} \frac{H(r')}{|r-r'|} dS'$$

where  $S'$  is the time dependent spherical surface defined by  $|r-r'| = ct$ . This equation represents OA forward model which inverse problem of reconstruction can be derived. OA images are reconstructed by absorbed energy field  $H(r')$  at specific location based on measured pressure waves.  $H(r')$  is calculated from detected pressure waves at the surface  $S$  as follows

$$H(r') = \frac{1}{\Gamma} \int_{\Omega} \frac{d\Omega}{\Omega} \left[ 2p(r, t) - 2t \frac{\partial p(r, t)}{\partial t} \right]_{t=\frac{|r-r'|}{c}}.$$

The constants at this equation can be omitted. After omitting the constants in the formula, the equation is discretized as

$$H(r'_j) = \sum \left[ p(r_i, t_{ij}) - t_{ij} \frac{\partial p(r_i, t_{ij})}{\partial t} \right]$$

where  $r'_j$  is the  $j$ -th point on the defined reconstruction grid,  $r_i$  is the position of  $i$ -th transducer and  $t_{ij} = |r_i - r'_j|/c$ . In summary, the equation calculates the distance between a point on the defined grid and an element of the transducer array. Then, it finds the corresponding wave intensity in signal (or defined as surface) based on the time of flight calculated using speed of sound in the imaging medium.

### A.2 Transducer array details

**Semi circle:** The array contains 256 piezocomposite transducers distributed over a semi circle (concave surface) equidistantly with the radius of 40 mm (Fig. 1b, main manuscript). The single transducer elements have dimensions of 0.37 mm × 15 mm with inter-element distance of 0.10 mm.

This configuration of transducer elements results in cylindrical (toroidal) focusing at 38 mm (close to the center of the array). The central peak frequency of array is 5 MHz with 60% bandwidth at -6dB.

**Multisegment:** The array is formed by the combination of a linear detector array and concave parts on the right and left sides as shown in Fig. 1c (main manuscript). The linear part contains 128 elements distributed on a linear surface with inter-element pitch size of 0.25 mm. Both of the concave parts include 64 elements which make the total number of elements equal to 256 (128 linear + 128 concave). The inter-element pitch size of concave part is 0.6 mm with 40 mm radius of curvature. The height of all elements are equal to 10 mm. Concave parts are designed to increase angular coverage in OA imaging. This configuration results in a cylindrical focusing at 38 mm close to the center of the array. The array has 7.5 MHz central frequency with 70% bandwidth at -6 dB.

**Linear array:** The array is central part of the multisegment array with 128 transducer elements distributed over a line with pitch size of 0.25 mm (Fig. 1c, main manuscript). Similar to concave parts, the linear array has 7.5 MHz central frequency with 70% bandwidth at -6 dB. The linear array is optimized for US data acquisitions with planar waves. Hence, the array produces OA images with limited view artifacts due to reduced angular coverage which is a limiting factor for OA image acquisitions.

**Virtual circle:** The array is generated to simulate images with 360 degree angular coverage which results in artifact free reconstructions (Fig. 2a, main manuscript). It contains 1,024 transducer elements distributed over a full circle with equal distance. The radius of the transducer array is kept equal to semi circle array (40 mm) to allow comparison between simulations and experimental acquisitions.

### A.3 OA image reconstruction

A Python package “pyoat” is presented with the datasets to reconstruct images from raw signals. The library uses “pip” package manager to install and use the functions. The package provides functions to bandpass filter and normalize raw signals as preprocessing step. Implementation of backprojection algorithm is also included in the package (Sec. 2.2, main manuscript). The forward model operator is implemented to simulate signals from acoustic pressure maps. Data readers and savers are integrated into the package for loading the raw data and saving the reconstructed images, respectively. The package provides examples to use different functions. The examples are easy to use Python scripts which require only raw signal data paths as input. Positions of elements in all transducer arrays used in this study are included in the library to enable reconstruction from raw signals.

### A.4 Experimental data acquisition setup

Signal acquisition is performed with OA imaging setup that combines four main components, namely, transducer arrays, nanosecond pulsed lasers, data acquisition system and workstation PC (Fig. 1, main manuscript). SWFD dataset was acquired with multisegment and semi circle transducer arrays (Sec. 2.1, main manuscript) using a nanosecond laser at 1,064 nm with repetition rate of 10 Hz. MSFD dataset was acquired with only multisegment array at six different wavelengths (700, 730, 760, 780, 800, 850 nm) using the laser with repetition rate of 50 Hz. The increased repetition rate guarantees that displacement within the scene between the frames of different wavelengths are minimal. Data acquisition system is used to digitize signals acquired by the transducer arrays. A sampling rate of 40 mega-samples-per-second was used in all experiments. Then, digital signals are sent to workstation PC to store the data and display both raw signals and reconstructed images in real time with lower resolution. The high resolution images presented in this work were reconstructed after the acquisition (offline) using the backprojection method (Sec. 2.2, main manuscript). The real time feedback helps to position the transducer arrays. The workstation PC also synchronizes all the imaging setup components by setting delays for data acquisition system and triggers for laser pulses. Imaging medium was filled with water to increase coupling efficiency between the transducer arrays and the skin. Water has low attenuation at the wavelengths used in this study. Hence, the signal attenuation in the medium was low enough to be neglected. Transducer arrays are held orthogonal to the forearm surface throughout acquisition and swiftly moved from elbow towards wrist. All participants joined the experiments voluntarily and were informed about details of the experiments.

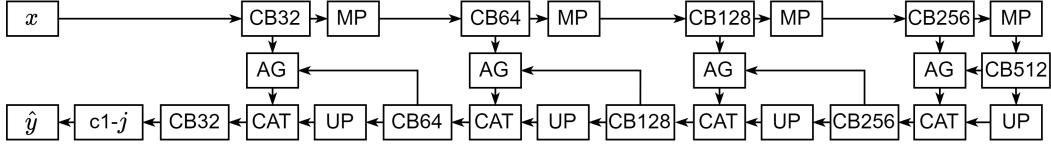


Figure 7: Schematic of the proposed modUNet architecture.  $\text{CB}j$  represents the residual 2D convolutional block with batch normalization shown in Fig. 3 (main manuscript) where each convolution has  $j$  filters. Other abbreviations correspond to 2D-maxpooling (MP) of poolsize 2, 2D bilinear upsampling (UP) by a factor of 2, concatenation (CAT), and attention gates (AG) [29]. Finally,  $c1-j$  represents a convolutional layer of  $j$  filters (1 for image translation and 3 for semantic segmentation experiments) and  $1 \times 1$  kernels without activation.

## B Fitzpatrick skin prototype in experimental datasets

Fitzpatrick skin prototype is a metric to quantify the amount of melanin pigment in the skin of a subject [14]. The metric ranges from one to six, going from pale white skin to black skin color. This is relevant for various OA applications due to different skin types (melanin concentration) lead to varying amount of contrast (absorption) at skin surfaces in the acquired images. Accordingly, the distribution of the skin types of the volunteers across SWFD and MSFD are shown in Fig. 6.

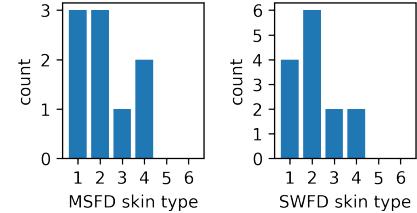


Figure 6: Fitzpatrick skin phototype [14] distribution of volunteers in datasets.

## C Architecture and implementation details

We show the schematic of the proposed modUNet architecture in Fig. 7.

We use categorical cross entropy loss and mean squared error loss for segmentation and image translation experiments, respectively. We add an additional  $l_1 = l_2 = 0.01$  regularization weight for each learned model parameter.

For all experiments, we use Adam optimizer and scale the learning rate with exponential decay (decay rate of 0.98 and decay steps of 1,000) from a peak of 0.0001 following a linear warmup of 10,000 steps. Except for experiments on OADAT-mini, all models are trained for 150 epochs with a mini-batch size of 25 and best validation set loss is used as the early stopping criteria for which test set performance metrics are presented. Experiments done on OADAT-mini are optimized for 10,000 epochs. For each experiment, we used an Nvidia P100 or Titan X GPU with 12GB memory to train the model. We do not augment data during training. Experiments on MSFD, SWFD, and SCD took about, 61, 32 and 15 hours, respectively. The training time variation across datasets is due to the difference in size of the datasets.

## D Qualitative results

Despite the impressive performance of the modUNet, there is a long tail in almost all tasks for SSIM or IoU distributions as seen in Figs. 4 and 5 (main manuscript). Accordingly, we qualitatively examine some of the worst samples at the bottom of these tails for each task.

### D.1 Image translation

We showcase and analyze a selection of the worst performing samples from the most challenging limited view and sparse sampling reconstruction experiments on SWFD and report SSIM for these samples with respect to the target sample.

In Figs. 8 & 9 (experiments  $f_{\text{SWFD\_lv128,li}}$  and  $f_{\text{SWFD\_lv128,sc}}$ ) one can see that some of the worst samples based on SSIM correspond to images with little to no signal content. The remaining sample from the 10th percentile in Fig. 8 suggests that low SSIM in multisegment array images can be

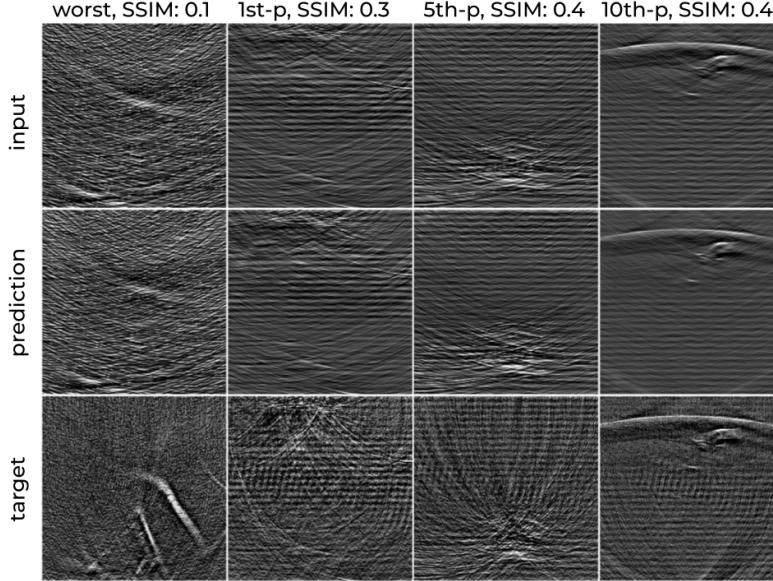


Figure 8: We showcase the worst (1st column), 1st- (2nd column), 5th- (3rd column), and 10th-percentile (4th column) SSIM samples based on modUNet predictions for experiment  $f_{SWFD\_lv128,li}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. Red arrows indicate some of the distorted vessel geometries at input getting corrected at modUNet predictions.

deceiving due to different set of artifacts that appear on the target (multisegment) array. In Fig. 9, where the target is the semi circle array, the artifacts are not as intense. Nevertheless, given the accurate vessel geometry correction, low SSIMs in 1st, 5th and 10th percentiles can be attributed to other local noise patterns, such as the circular noise that exist on the target image. Thanks to the majority of samples in our datasets having high SNR, most of the undesirable artifacts that seldomly occur in test samples are removed. For example, in Figs. 9 & 10, columns with relatively higher signal content are either very close to the target or even more artifact-free.

In Figs. 11 & 12, we show some of the samples with the worst SSIM from the experiments  $f_{SWFD\_ss32,sc}$  and  $f_{SWFD\_ss32,ms}$ , respectively. The samples in the left two columns for both experiments have very poor SNR. For all other samples, modUNet predictions from the heavily subsampled image reconstructions indicate higher SNR than the target samples thanks to removing most of the noise pattern (especially visible within the water medium prior to/above the forearm) while keeping all soft tissue structures intact and faithful to their counterparts in the target image. Unfortunately, mismatch of the noise patterns causes low SSIM for such samples with low SNR target images.

For completeness, we show the worst performing samples for SSIM in image translation experiments using SCD in Fig. 13. One can see that the rounded SSIM already reaches 1.0 by the 1st- (right figure) or 5th- (left figure) percentile performing sample in the test set. In the left figure 2nd column, one can see that the strong artifact passing through the small vessel caused modUNet to separate the vessel into two pieces, shown with a red arrow.

In addition, we also showcase some of the best performing samples for SSIM in Figs. 14, 15, 16, 17 & 18. Beyond correcting distorted vessel geometry, it can be seen in some samples that certain vessels that are barely visible get accurately redrawn to match the target (e.g., Fig. 9 3rd and 4th columns, Fig. 11 2nd, 3rd and 4th column, Fig. 12 3rd and 4th columns, and Fig. 18 4th column).

## D.2 Semantic segmentation

Similar to image translation task, we look at some of the samples from SCD with the worst vessel segmentation performance with respect to IoU metric in Figs. 19 & 20 (experiments  $f_{seg\_lv128,li}$  and

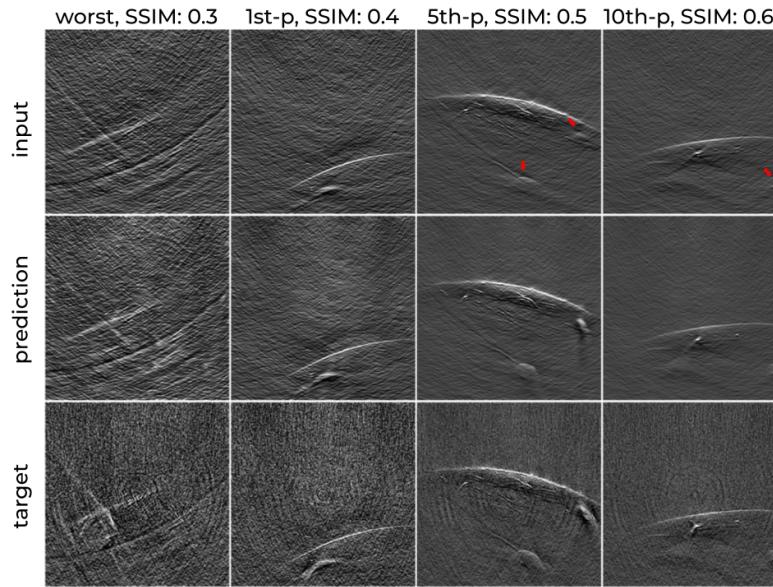


Figure 9: We showcase the worst (1st column), 1st- (2nd column), 5th- (3rd column), and 10th-percentile (4th column) SSIM samples based on modUNet predictions for experiment  $f_{\text{SWFD\_lv128,sc}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. Red arrows indicate some of the distorted vessel geometries at input getting corrected at modUNet predictions.

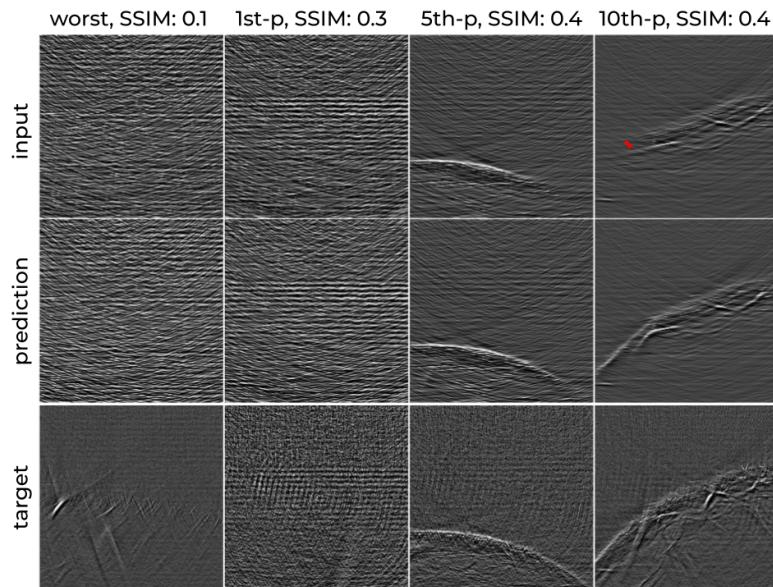


Figure 10: We showcase the worst (1st column), 1st- (2nd column), 5th- (3rd column), and 10th-percentile (4th column) SSIM samples based on modUNet predictions for experiment  $f_{\text{MSFD\_lv128,li}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. Red arrows indicate some of the distorted vessel geometries at input getting corrected at modUNet predictions.

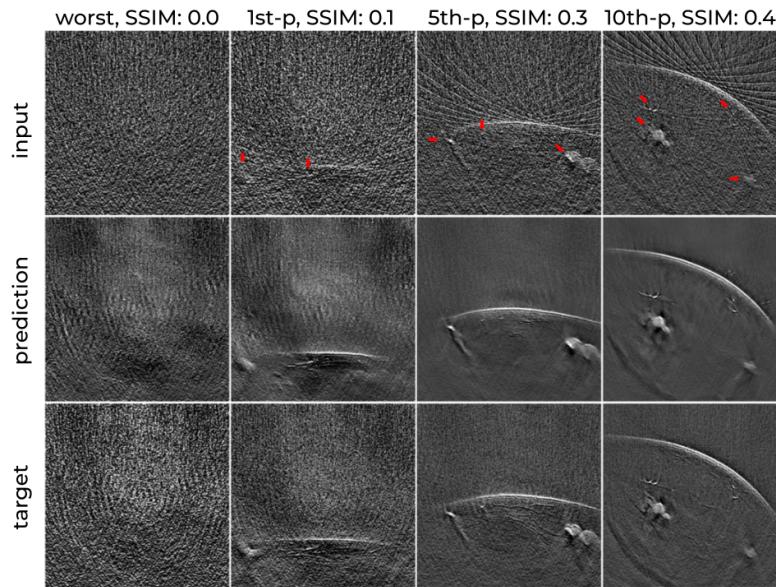


Figure 11: We showcase the worst (1st column), 1st- (2nd column), 5th- (3rd column), and 10th-percentile (4th column) SSIM samples based on modUNet predictions for experiment  $f_{\text{SWFD\_ss32,sc}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs.

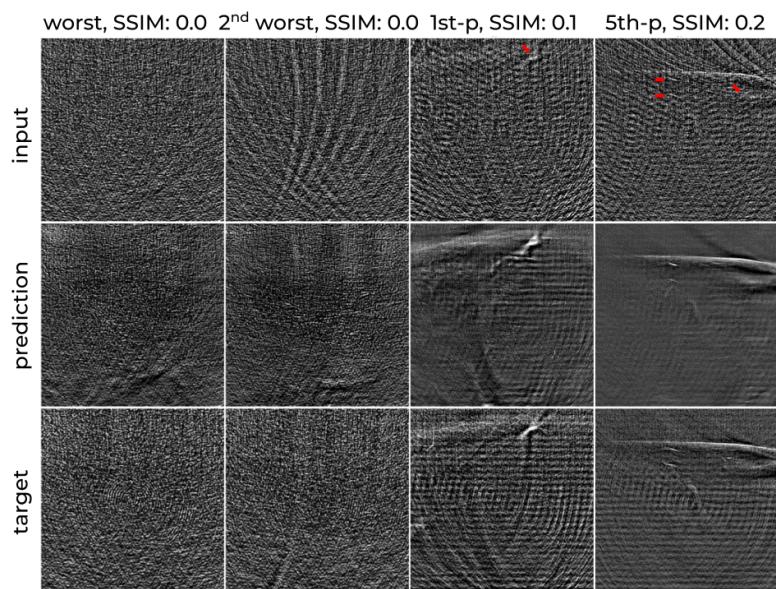


Figure 12: We showcase the worst (1st column), 1st- (2nd column), 5th- (3rd column), and 10th-percentile (4th column) SSIM samples based on modUNet predictions for experiment  $f_{\text{SWFD\_ss32,ms}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. Red arrows indicate some of the nearly invisible vessels at input getting corrected at modUNet predictions.

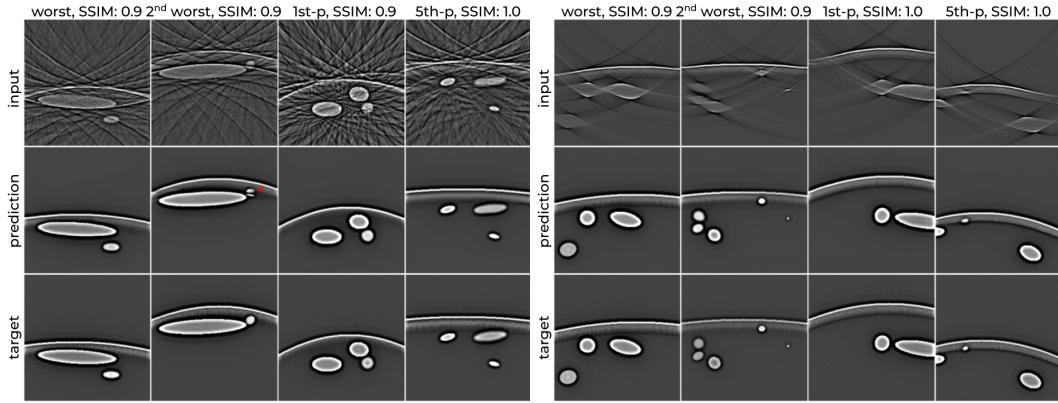


Figure 13: We showcase the worst (1st column), 2nd worst (2nd column), 1st- (3rd column), and 5th-percentile (4th column) SSIM samples based on modUNet predictions for experiments  $f_{SCD\_ss32,vc}$  (left) and  $f_{SCD\_lv128,li}$  (right) with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs.

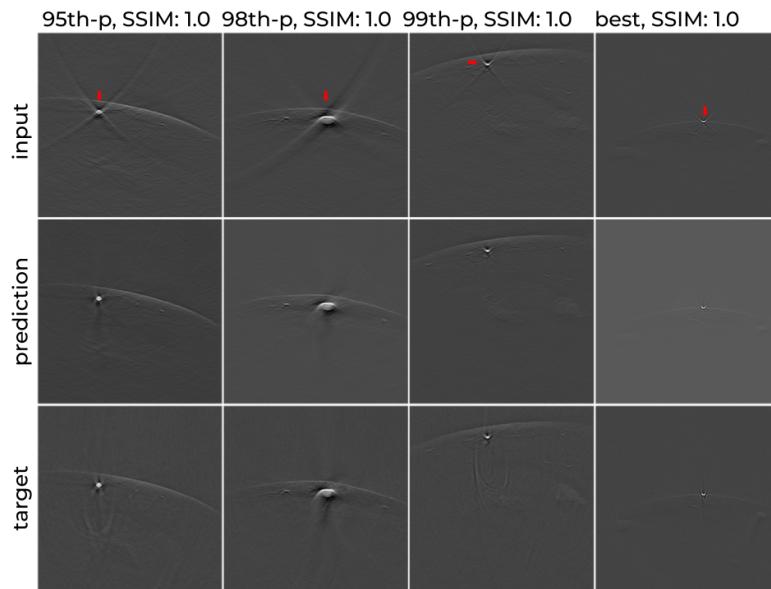


Figure 14: We showcase 95th- (1st column), 98th- (2nd column), 99th-percentile (3rd column), and the best (4th column) SSIM samples based on modUNet predictions for experiment  $f_{SWFD\_lv128,sc}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs.

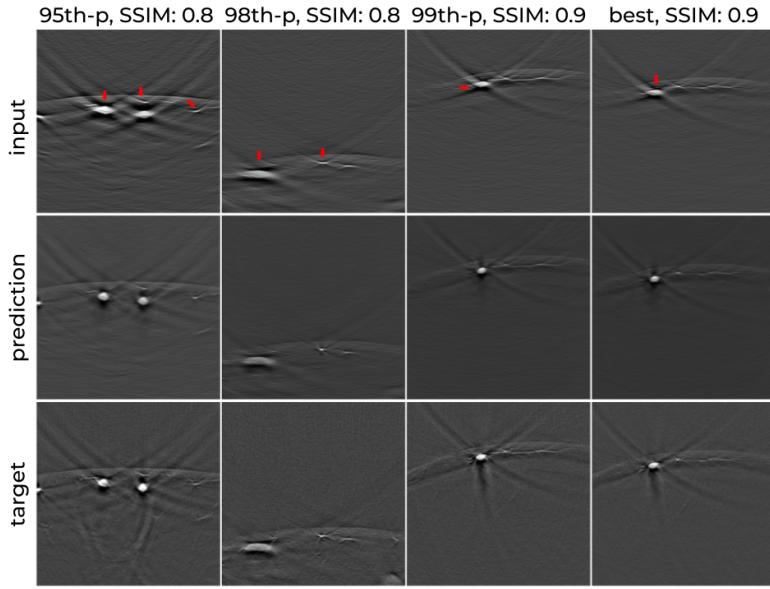


Figure 15: We showcase 95th- (1st column), 98th- (2nd column), 99th-percentile (3rd column), and the best (4th column) SSIM samples based on modUNet predictions for experiment  $f_{\text{SWFD\_lv128,li}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. Red arrows indicate some of the distorted vessel geometries at input getting corrected at modUNet predictions.

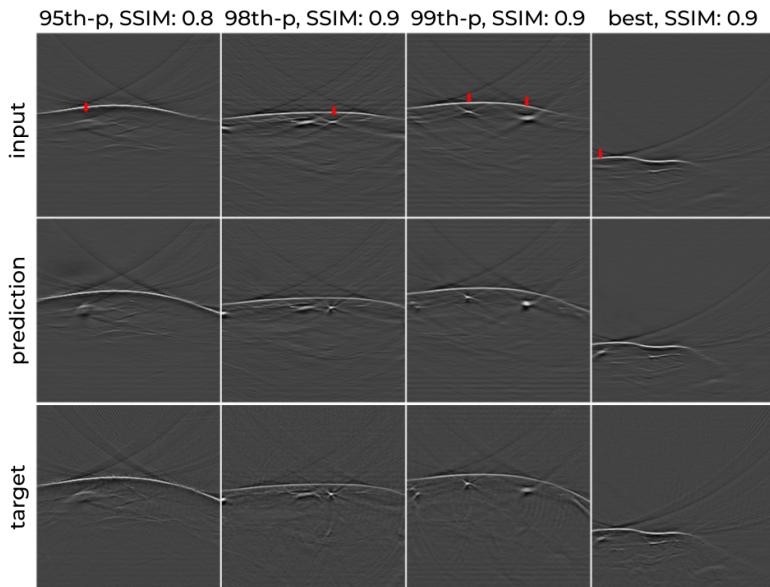


Figure 16: We showcase 95th- (1st column), 98th- (2nd column), 99th-percentile (3rd column), and the best (4th column) SSIM samples based on modUNet predictions for experiment  $f_{\text{MSFD\_lv128,li}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. Red arrows indicate some of the distorted vessel geometries at input getting corrected at modUNet predictions.

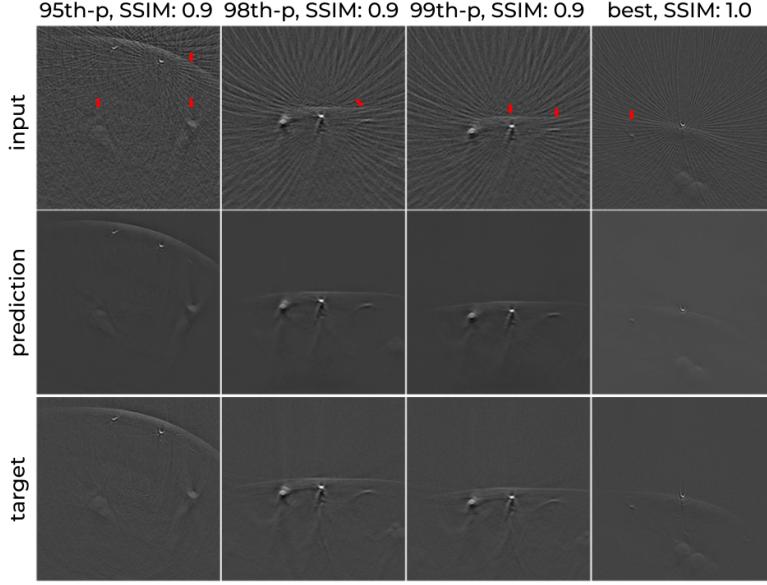


Figure 17: We showcase 95th- (1st column), 98th- (2nd column), 99th-percentile (3rd column), and the best (4th column) SSIM samples based on modUNet predictions for experiment  $f_{\text{SWFD\_ss32,sc}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs.

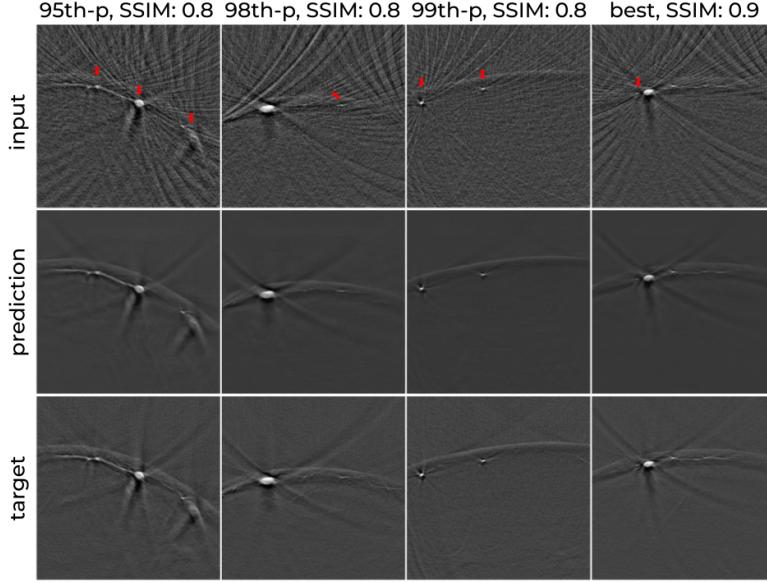


Figure 18: We showcase 95th- (1st column), 98th- (2nd column), 99th-percentile (3rd column), and the best (4th column) SSIM samples based on modUNet predictions for experiment  $f_{\text{SWFD\_ss32,ms}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs.

$f_{\text{seg\_ss32,vc}}$ ). As a result of SCD having flawless ground truth annotations, modUNet can achieve near perfect segmentation performance for all classes in all experiments of both tasks. The decrease of IoU seems to be exclusive for under- or over-segmenting little vessels by a handful of pixels. Despite most vessels being small (see Fig. 2e in the main manuscript, distribution of number of pixels per vessel), low IoU is limited to only a handful of extreme cases. This is also corroborated by the fact that skin curve segmentation distribution (see Fig. 5 right in the main manuscript) having a significantly shorter tail, barely falling below an IoU of 0.95 in most experiments. In order to understand the single experiment with a considerable performance drop, we look at experiment  $f_{\text{seg\_ss32,vc}}$  in Fig. 21. It can be observed that under heavy sparse sampling, the skin curve often gets ambiguous, rarely leading

to suboptimal segmentation performance in these cases. Nevertheless, for most such subsampled reconstructions having intense duplicates of skin curve (Fig. 20 first row), modUNet successfully segments the right curve. Considering OADAT-mini segmentation experiments, in Fig. 22, we show the worst performing four samples from the test set of MSFD-mini. In Fig 23, it can be seen that such poor IoU is not observed in the worst performing samples for SWFD-mini experiments.

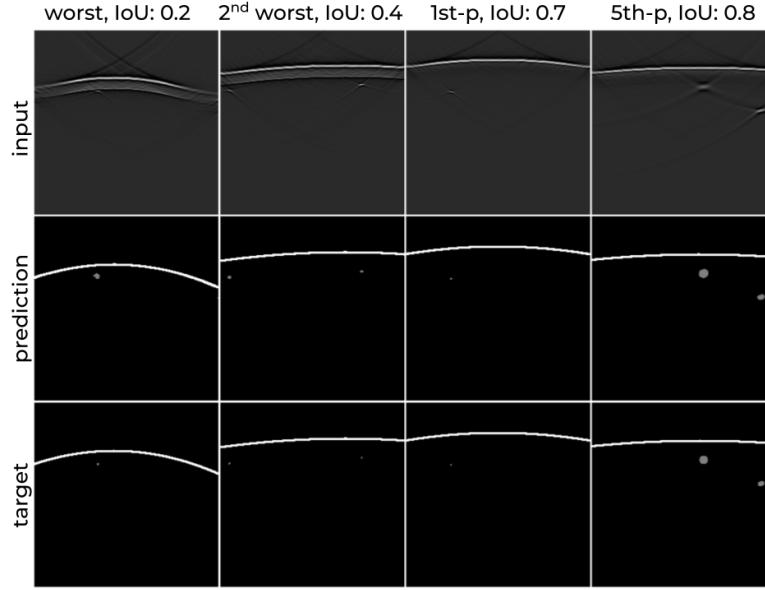


Figure 19: We showcase the worst (1st column), 2nd worst (2nd column), 1st- (3rd column), and 5th-percentile (4th column) vessel IoU samples based on modUNet predictions for experiment  $f_{\text{seg\_lv128,li}}$  with input (1st row), modUNet prediction (2nd row), and ground truth (3rd sample) pairs.

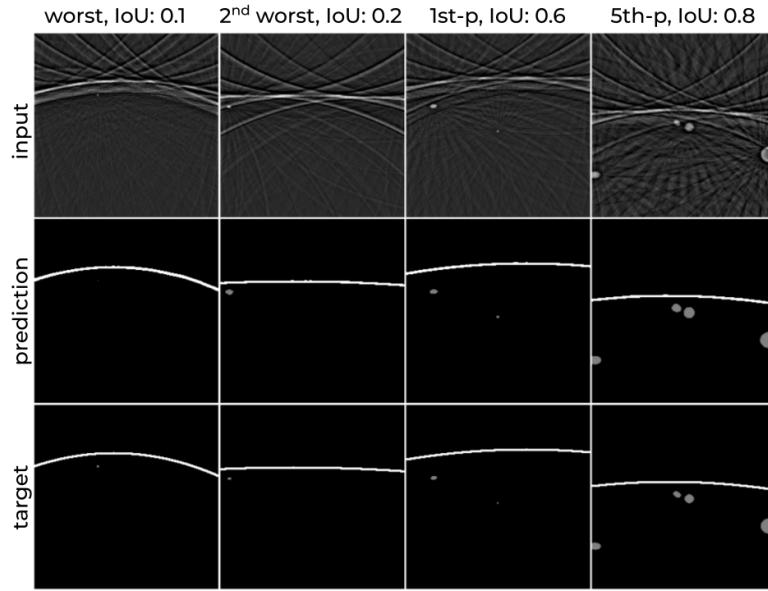


Figure 20: We showcase the worst (1st column), 2nd worst (2nd column), 1st- (3rd column), and 5th-percentile (4th column) vessel IoU samples based on modUNet predictions for experiment  $f_{\text{seg\_ss32,vc}}$  with input (1st row), modUNet prediction (2nd row), and ground truth (3rd sample) pairs.

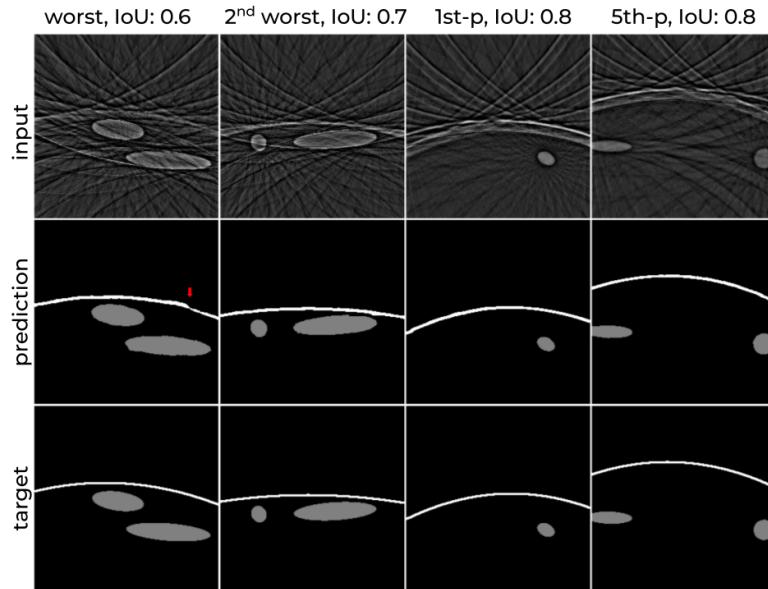


Figure 21: We showcase the worst (1st column), 2nd worst (2nd column), 1st- (3rd column), and 5th-percentile (4th column) skin curve IoU samples based on modUNet predictions for experiment  $f_{\text{seg\_ss32,vc}}$  with input (1st row), modUNet prediction (2nd row), and ground truth (3rd sample) pairs.

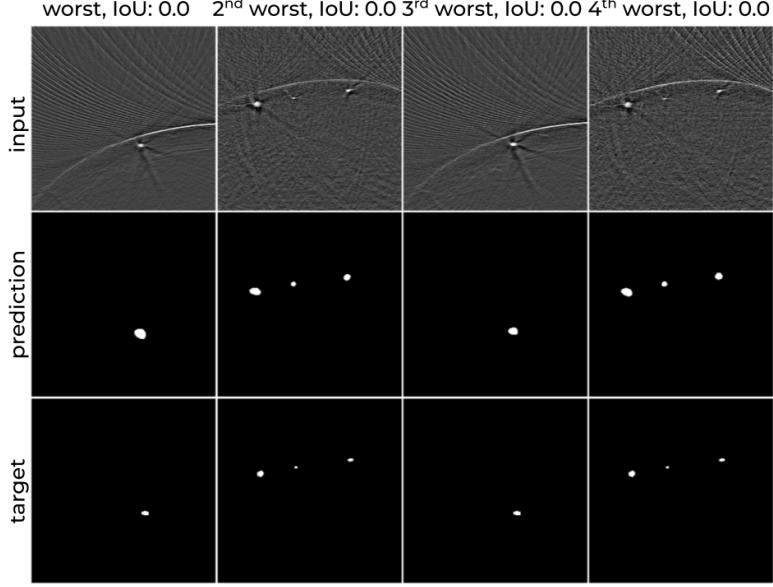


Figure 22: We showcase the worst (1st column), 2nd- (2nd column), 3rd- (3rd column), and 4th-worst (4th column) IoU samples based on modUNet predictions for experiment  $f_{\text{seg\_MSFD\_ss64,ms}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs.

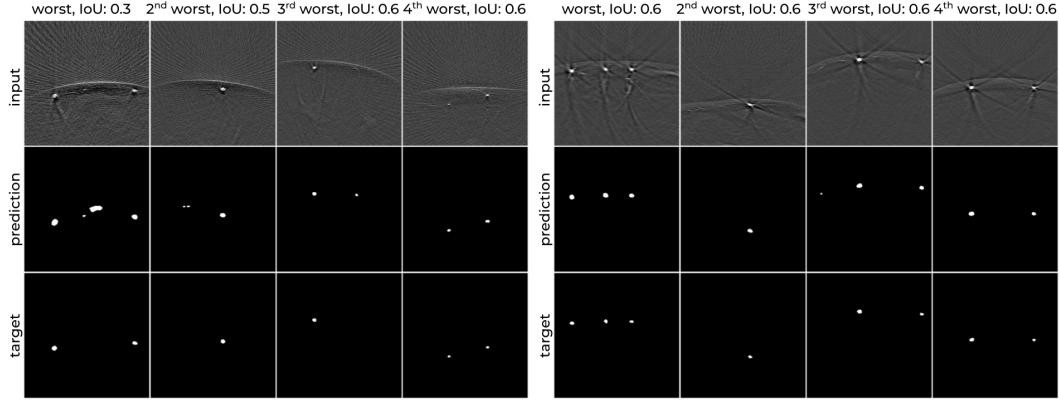


Figure 23: We showcase the worst (1st column), 2nd- (2nd column), 3rd- (3rd column), and 4th-worst (4th column) IoU samples based on modUNet predictions for experiments  $f_{\text{seg\_SWFD\_ss64,sc}}$  (left) and  $f_{\text{seg\_SWFD\_ss128,ms}}$  (right) with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs.

## E Simulated cylinder dataset generation

Our proposed simulated dataset, SCD, follows a group of heuristics we derived from observations on experimental images. For any given sample, we first generate an acoustic pressure map, from which we also generate ground truth annotations. We then apply certain post processing steps to imitate other phenomena such as patterns under the skin surface and different vessel textures. Given the geometry of the transducers we want to simulate, we then apply forward transform that gives the raw signals. Finally, we reconstruct these signals using backprojection algorithm to generate the images used in various SCD experiments in this manuscript.

The acoustic pressure map generation consists of initially drawing the curve that represents the laser pulse absorption on the skin surface mainly due to melanin. Given that experimental data is acquired with making sure that forearm is roughly at a certain distance range from the arrays, we also limit the drawn skin curve distance. We define the skin curve as a 2nd degree polynomial that is fitted to three points randomly sampled at the two horizontal edges and the center of the image at varying

heights. As a post-processing step to mimic experimental data, the curve is first smoothed with a Gaussian filter. Then, an exponential decay of randomized length is applied under the curve along vertical axis. Finally, a non-structured uniform normal noise is multiplied with the aforementioned exponential decay region. For vessel generation, the number of cylinders to be drawn is sampled based on a coin flip. Based on the outcome, either two cylinders drawn or the number of cylinders is sampled from Poisson distribution (see Fig. 2e, distribution of number of vessels per image, main manuscript). Each vessel is initially represented by a cylinder orthogonal to the image plane (z-axis) with a randomly sampled radius. We then randomly rotate the cylinder around x- and y- axes. The vessel is determined as the cross-section of the cylinder at the imaging plane, yielding ellipses based on the final angle of the cylinders. As a post-processing step, we flip a coin to determine whether the vessel has a homogeneous intensity profile or has a linearly decreasing intensity from its center. We then apply a Gaussian filter on the vessel to smooth its edges. Finally, based on a coin flip, we decide whether or not to multiply the intensity profile of the vessel with uniform normal noise. The same process is iteratively repeated until desired number of non-overlapping vessels are generated. All parameters used for the aforementioned steps are empirically selected based on our observations on the experimental datasets. We provide the script to simulate acoustic pressure maps that we used for SCD.

## F Organization of datasets

The datasets and required algorithms to read raw signals and reconstructed images are summarized in Fig. 24 along with the codes to rerun and evaluate all the presented experiments. All datasets are stored as HDF5 files and corresponding indices across datasets within an HDF5 file have the same scene as raw signal or reconstructed image. Similarly, metadata such as *patientID*, *side* (i.e., left vs right arm), *skin type*, and *sliceID* contain the information for the corresponding indices. Some of these metadata such as *patientID*, which corresponds to the unique anonymized identifier of the volunteer, only exists in experimental datasets. Matching identifiers across experimental datasets (MSFD and SWFD) correspond to the same volunteer.

Each sample (raw signal or image reconstruction) is chunked in a single piece for optimal compression/decompression overhead when reading individual images and storing datasets. For example, for a given HDF5 dataset of shape (number of instances, height, width), each block of (1, height, width) is compressed individually. In our experiments, this ensured zero additional idle GPU time reading random training sample from a compressed HDF5 file when compared with its non-compressed HDF5 counterpart.

Raw signal datasets have the shape of (number of instances, temporal acquisition axis, receiving element axis). Temporal acquisition axis is fixed to 2,030 sampling points across all transducer arrays. Receiving element axis depends on the number of transducer elements that is used for signal acquisition. For sparse sampling and limited view reconstructions, we retain the shape of the raw signal data, however, we switch off the receiving elements that are not active, hence they are padded with zeros. This convention also allows for direct compatibility with the *pyoat* reconstruction package.

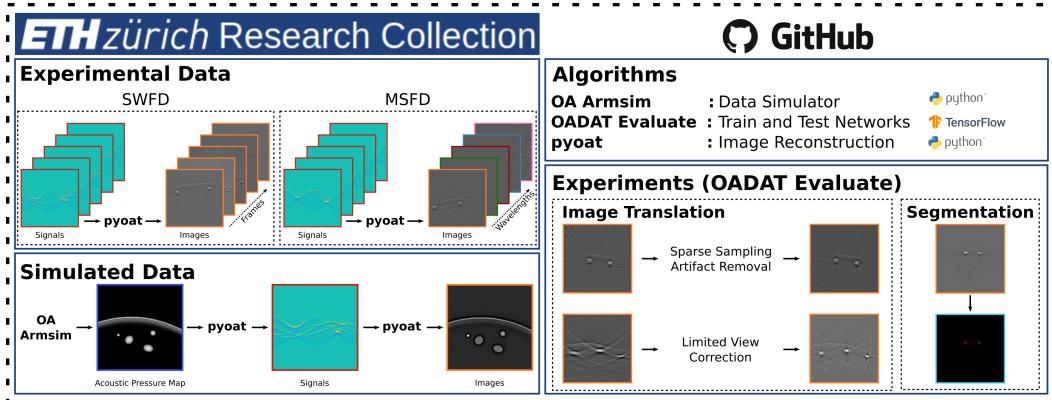


Figure 24: Pipeline figure to summarize data storage and algorithms.

Table 5: Contents of the MSFD file.

Name	Shape	Content	Name	Shape	Content
patientID	(25200,)	(2, 5, 6, 7, 9, 10, 11, 14, 15)	ms_ss64_BP_w700	(25200, 256, 256)	image reconstructions
side	(25200,)	(left, right)	ms_ss64_BP_w730	(25200, 256, 256)	image reconstructions
skin_type	(25200,)	(1, 2, 3, 4)	ms_ss64_BP_w760	(25200, 256, 256)	image reconstructions
sliceID	(25200,)	(1, ..., 1400)	ms_ss64_BP_w780	(25200, 256, 256)	image reconstructions
linear_BP_w700	(25200, 256, 256)	image reconstructions	ms_ss64_BP_w800	(25200, 256, 256)	image reconstructions
linear_BP_w730	(25200, 256, 256)	image reconstructions	ms_ss64_BP_w850	(25200, 256, 256)	image reconstructions
linear_BP_w760	(25200, 256, 256)	image reconstructions	ms_ss128_BP_w700	(25200, 256, 256)	image reconstructions
linear_BP_w780	(25200, 256, 256)	image reconstructions	ms_ss128_BP_w730	(25200, 256, 256)	image reconstructions
linear_BP_w800	(25200, 256, 256)	image reconstructions	ms_ss128_BP_w760	(25200, 256, 256)	image reconstructions
linear_BP_w850	(25200, 256, 256)	image reconstructions	ms_ss128_BP_w780	(25200, 256, 256)	image reconstructions
ms_BP_w700	(25200, 256, 256)	image reconstructions	ms_ss128_BP_w800	(25200, 256, 256)	image reconstructions
ms_BP_w730	(25200, 256, 256)	image reconstructions	ms_ss128_BP_w850	(25200, 256, 256)	image reconstructions
ms_BP_w760	(25200, 256, 256)	image reconstructions	ms_ss32_raw_w700	(25200, 2030, 256)	raw signals
ms_BP_w780	(25200, 256, 256)	image reconstructions	ms_ss32_raw_w730	(25200, 2030, 256)	raw signals
ms_BP_w800	(25200, 256, 256)	image reconstructions	ms_ss32_raw_w760	(25200, 2030, 256)	raw signals
ms_BP_w850	(25200, 256, 256)	image reconstructions	ms_ss32_raw_w780	(25200, 2030, 256)	raw signals
linear_raw_w700	(25200, 2030, 256)	raw signals	ms_ss32_raw_w800	(25200, 2030, 256)	raw signals
linear_raw_w730	(25200, 2030, 256)	raw signals	ms_ss32_raw_w850	(25200, 2030, 256)	raw signals
linear_raw_w760	(25200, 2030, 256)	raw signals	ms_ss64_raw_w700	(25200, 2030, 256)	raw signals
linear_raw_w780	(25200, 2030, 256)	raw signals	ms_ss64_raw_w730	(25200, 2030, 256)	raw signals
linear_raw_w800	(25200, 2030, 256)	raw signals	ms_ss64_raw_w760	(25200, 2030, 256)	raw signals
linear_raw_w850	(25200, 2030, 256)	raw signals	ms_ss64_raw_w780	(25200, 2030, 256)	raw signals
ms_raw_w700	(25200, 2030, 256)	raw signals	ms_ss64_raw_w760	(25200, 2030, 256)	raw signals
ms_raw_w730	(25200, 2030, 256)	raw signals	ms_ss64_raw_w780	(25200, 2030, 256)	raw signals
ms_raw_w760	(25200, 2030, 256)	raw signals	ms_ss64_raw_w800	(25200, 2030, 256)	raw signals
ms_raw_w780	(25200, 2030, 256)	raw signals	ms_ss64_raw_w850	(25200, 2030, 256)	raw signals
ms_raw_w800	(25200, 2030, 256)	raw signals	ms_ss128_raw_w700	(25200, 2030, 256)	raw signals
ms_raw_w850	(25200, 2030, 256)	raw signals	ms_ss128_raw_w730	(25200, 2030, 256)	raw signals
ms_ss32_BP_w700	(25200, 256, 256)	image reconstructions	ms_ss128_raw_w760	(25200, 2030, 256)	raw signals
ms_ss32_BP_w730	(25200, 256, 256)	image reconstructions	ms_ss128_raw_w780	(25200, 2030, 256)	raw signals
ms_ss32_BP_w760	(25200, 256, 256)	image reconstructions	ms_ss128_raw_w800	(25200, 2030, 256)	raw signals
ms_ss32_BP_w780	(25200, 256, 256)	image reconstructions	ms_ss128_raw_w850	(25200, 2030, 256)	raw signals
ms_ss32_BP_w800	(25200, 256, 256)	image reconstructions			
ms_ss32_BP_w850	(25200, 256, 256)	image reconstructions			

We provide a header dataset file “OADAT.h5” as a convenient access point to each other HDF5 dataset file. For convenience, we also provide a header dataset file “OADAT\_v2.h5” that allows more intuitive access to addendums of MSFD, SWFD, and SCD with multisegment sparse sampling datasets. Additionally, a header dataset file “OADAT-mini.h5” is provided as an access point to reach all relevant OADAT-mini HDF5 files. These header files also contain metadata module of the Dataset Nutrition Label, which should help making the dataset self explanatory.

## F.1 MSFD

Contents of MSFD are listed in Table 5. The contents of the dataset are stored in the files “MSFD\_multisegment\_RawBP.h5” and “MSFD\_multisegment\_ss\_RawBP.h5”. Provided *sliceID* corresponds to the time index of the slice acquired from a given volunteer for a given acquisition. For example, *sliceID*  $i + 1$  is recorded right after  $i$  for a given *patientID* and *side*. This information can be relevant for future work that does not treat each slice independently, but exploit correlations from consecutive slice acquisitions. Nevertheless, a *sliceID*  $i \in [1, \dots, 1400]$  does not necessarily correspond to the same position on the forearm across the volunteers.

## F.2 SWFD

Contents of SWFD are split into two separate HDF5 files with contents listed in Tables 6 & 7 (“SWFD\_semicircle\_RawBP.h5” and “SWFD\_multisegment\_RawBP.h5”). Despite having the same number of instances per dataset category across the two files, the indices are not paired across semi circle and multisegment array. Hence we decided to split the dataset into two files. Furthermore, we include two additional HDF5 files “SWFD\_semicircle\_ss\_RawBP.h5” and “SWFD\_multisegment\_ss\_RawBP.h5” that contain sparse sampling image and raw data for multisegment array. As an artifact from discarding first few slices to filter out low SNR instances (both in MSFD and SWFD), SWFD acquisitions have 1,401 slices, having *sliceID* in the range  $[100, \dots, 1500]$ , yielding  $2 \times 14 \times 1,401 = 39,228$  samples per dataset category as opposed to 39,200.

Table 6: Contents of SWFD semi circle file.

Name	Shape	Content	Name	Shape	Content
patientID	(39228,)	(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14)	sc_lv128_raw	(39228, 2030, 256)	raw signals
side	(39228,)	(left, right)	sc_ss128_raw	(39228, 2030, 256)	raw signals
skin_type	(39228,)	(1, 2, 3, 4)	sc_ss32_raw	(39228, 2030, 256)	raw signals
sliceID	(39228,)	(100, ..., 1500)	sc_ss64_raw	(39228, 2030, 256)	raw signals
sc_lv128_BP	(39228, 256, 256)	image reconstructions	sc_raw	(39228, 2030, 256)	raw signals
sc_ss128_BP	(39228, 256, 256)	image reconstructions			
sc_ss32_BP	(39228, 256, 256)	image reconstructions			
sc_ss64_BP	(39228, 256, 256)	image reconstructions			
sc_BP	(39228, 256, 256)	image reconstructions			

Table 7: Contents of SWFD multisegment file.

Name	Shape	Content	Name	Shape	Content
patientID	(39228,)	(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14)	ms_ss128_BP	(39228, 256, 256)	image reconstructions
side	(39228,)	(left, right)	ms_ss32_BP	(39228, 256, 256)	image reconstructions
skin_type	(39228,)	(1, 2, 3, 4)	ms_ss64_BP	(39228, 256, 256)	image reconstructions
sliceID	(39228,)	(100, ..., 1500)	ms_ss128_raw	(39228, 2030, 256)	raw signals
linear_BP	(39228, 256, 256)	image reconstructions	ms_ss32_raw	(39228, 2030, 256)	raw signals
ms_BP	(39228, 256, 256)	image reconstructions	ms_ss64_raw	(39228, 2030, 256)	raw signals
linear_raw	(39228, 2030, 256)	raw signals			
ms_raw	(39228, 2030, 256)	raw signals			

### F.3 SCD

Contents of SCD are listed in Table 8 and stored in the file “SCD\_RawBP.h5”. Different than the experimental datasets, SCD also contains the ground truth acoustic pressure map and annotations (labels). Annotations are encoded as 0: background, 1: vessel, and 2: skin curve.

### F.4 OADAT-mini

OADAT-mini consists of a subset of 100 instances of each category of OADAT along with their annotation maps for vessels. Accordingly, its contents are split into multiple files, one for each transducer array and SCD: “SWFD\_semicircle\_RawBP-mini.h5”, “SWFD\_multisegment\_RawBP-mini.h5”, “MSFD\_multisegment\_RawBP-mini.h5”, “SCD\_RawBP-mini.h5”. We list the contents of each of these files in Table 9.

Table 8: Contents of SCD files.

Name	Shape	Content	Name	Shape	Content
sliceID	(20000,)	(0, ..., 19999)	vc_lv128_raw	(20000, 2030, 1024)	raw signals
ground_truth	(20000, 256, 256)	acoustic pressure map	vc_ss128_raw	(20000, 2030, 1024)	raw signals
labels	(20000, 256, 256)	(0, 1, 2)	vc_ss32_raw	(20000, 2030, 1024)	raw signals
linear_BP	(20000, 256, 256)	image reconstructions	vc_ss64_raw	(20000, 2030, 1024)	raw signals
ms_BP	(20000, 256, 256)	image reconstructions	vc_raw	(20000, 2030, 1024)	raw signals
linear_raw	(20000, 2030, 256)	raw signals	ms_ss128_BP	(20000, 256, 256)	image reconstructions
ms_raw	(20000, 2030, 256)	raw signals	ms_ss32_BP	(20000, 256, 256)	image reconstructions
vc_lv128_BP	(20000, 256, 256)	image reconstructions	ms_ss64_BP	(20000, 256, 256)	image reconstructions
vc_ss128_BP	(20000, 256, 256)	image reconstructions	ms_ss128_raw	(20000, 2030, 256)	raw signals
vc_ss32_BP	(20000, 256, 256)	image reconstructions	ms_ss32_raw	(20000, 2030, 256)	raw signals
vc_ss64_BP	(20000, 256, 256)	image reconstructions	ms_ss64_raw	(20000, 2030, 256)	raw signals
vc_BP	(20000, 256, 256)	image reconstructions			

Table 9: Contents of OADAT-mini files.

Name	Shape	Content	Name	Shape	Content
<b>SCD - mini</b>					
sliceID	(100,)	$\subset (0, \dots, 19999)$	vc_lv128_raw	(100, 2030, 1024)	raw signals
ground_truth	(100, 256, 256)	acoustic pressure map	vc_ss128_raw	(100, 2030, 1024)	raw signals
labels	(100, 256, 256)	(0, 1, 2)	vc_ss32_raw	(100, 2030, 1024)	raw signals
linear_BP	(100, 256, 256)	image reconstructions	vc_ss64_raw	(100, 2030, 1024)	raw signals
ms_BP	(100, 256, 256)	image reconstructions	vc_raw	(100, 2030, 1024)	raw signals
linear_raw	(100, 2030, 256)	raw signals	ms_ss128_BP	(100, 256, 256)	image reconstructions
ms_raw	(100, 2030, 256)	raw signals	ms_ss32_BP	(100, 256, 256)	image reconstructions
vc_lv128_BP	(100, 256, 256)	image reconstructions	ms_ss64_BP	(100, 256, 256)	image reconstructions
vc_ss128_BP	(100, 256, 256)	image reconstructions	ms_ss128_raw	(100, 2030, 256)	raw signals
vc_ss32_BP	(100, 256, 256)	image reconstructions	ms_ss32_raw	(100, 2030, 256)	raw signals
vc_ss64_BP	(100, 256, 256)	image reconstructions	ms_ss64_raw	(100, 2030, 256)	raw signals
vc_BP	(100, 256, 256)	image reconstructions			
<b>SWFD semi circle - mini</b>					
patientID	(100,)	(1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13)	sc_lv128_raw	(100, 2030, 256)	raw signals
side	(100,)	(left, right)	sc_ss128_raw	(100, 2030, 256)	raw signals
skin_type	(100,)	(1, 2, 3, 4)	sc_ss32_raw	(100, 2030, 256)	raw signals
sliceID	(100,)	$\subset (100, \dots, 1500)$	sc_ss64_raw	(100, 2030, 256)	raw signals
labels	(100, 256, 256)	(0, 1)	sc_raw	(100, 2030, 256)	raw signals
sc_lv128_BP	(100, 256, 256)	image reconstructions			
sc_ss128_BP	(100, 256, 256)	image reconstructions			
sc_ss32_BP	(100, 256, 256)	image reconstructions			
sc_ss64_BP	(100, 256, 256)	image reconstructions			
sc_BP	(100, 256, 256)	image reconstructions			
<b>SWFD multisegment - mini</b>					
patientID	(100,)	(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12)	ms_ss128_BP	(100, 256, 256)	image reconstructions
side	(100,)	(left, right)	ms_ss32_BP	(100, 256, 256)	image reconstructions
skin_type	(100,)	(1, 2, 3, 4)	ms_ss64_BP	(100, 256, 256)	image reconstructions
sliceID	(100,)	$\subset (100, \dots, 1500)$	ms_ss128_raw	(100, 2030, 256)	raw signals
labels	(100, 256, 256)	(0, 1)	ms_ss32_raw	(100, 2030, 256)	raw signals
linear_BP	(100, 256, 256)	image reconstructions	ms_ss64_raw	(100, 2030, 256)	raw signals
ms_BP	(100, 256, 256)	image reconstructions			
linear_raw	(100, 2030, 256)	raw signals			
ms_raw	(100, 2030, 256)	raw signals			
<b>MSFD multisegment - mini</b>					
patientID	(100,)	(2, 5, 6, 7, 9, 10, 11, 14, 15)	ms_ss64_BP_w700	(100, 256, 256)	image reconstructions
side	(100,)	(left, right)	ms_ss64_BP_w730	(100, 256, 256)	image reconstructions
skin_type	(100,)	(1, 2, 3, 4)	ms_ss64_BP_w760	(100, 256, 256)	image reconstructions
sliceID	(100,)	$\subset (1, \dots, 1400)$	ms_ss64_BP_w780	(100, 256, 256)	image reconstructions
labels	(100, 256, 256)	(0, 1)	ms_ss64_BP_w800	(100, 256, 256)	image reconstructions
linear_BP_w700	(100, 256, 256)	image reconstructions	ms_ss64_BP_w850	(100, 256, 256)	image reconstructions
linear_BP_w730	(100, 256, 256)	image reconstructions	ms_ss128_BP_w700	(100, 256, 256)	image reconstructions
linear_BP_w760	(100, 256, 256)	image reconstructions	ms_ss128_BP_w730	(100, 256, 256)	image reconstructions
linear_BP_w780	(100, 256, 256)	image reconstructions	ms_ss128_BP_w760	(100, 256, 256)	image reconstructions
linear_BP_w800	(100, 256, 256)	image reconstructions	ms_ss128_BP_w780	(100, 256, 256)	image reconstructions
linear_BP_w850	(100, 256, 256)	image reconstructions	ms_ss128_BP_w800	(100, 256, 256)	image reconstructions
ms_BP_w700	(100, 256, 256)	image reconstructions	ms_ss128_BP_w850	(100, 256, 256)	image reconstructions
ms_BP_w730	(100, 256, 256)	image reconstructions	ms_ss32_raw_w700	(100, 2030, 256)	raw signals
ms_BP_w760	(100, 256, 256)	image reconstructions	ms_ss32_raw_w730	(100, 2030, 256)	raw signals
ms_BP_w780	(100, 256, 256)	image reconstructions	ms_ss32_raw_w760	(100, 2030, 256)	raw signals
ms_BP_w800	(100, 256, 256)	image reconstructions	ms_ss32_raw_w780	(100, 2030, 256)	raw signals
ms_BP_w850	(100, 256, 256)	image reconstructions	ms_ss32_raw_w800	(100, 2030, 256)	raw signals
linear_raw_w700	(100, 2030, 256)	raw signals	ms_ss32_raw_w850	(100, 2030, 256)	raw signals
linear_raw_w730	(100, 2030, 256)	raw signals	ms_ss64_raw_w700	(100, 2030, 256)	raw signals
linear_raw_w760	(100, 2030, 256)	raw signals	ms_ss64_raw_w730	(100, 2030, 256)	raw signals
linear_raw_w780	(100, 2030, 256)	raw signals	ms_ss64_raw_w760	(100, 2030, 256)	raw signals
linear_raw_w800	(100, 2030, 256)	raw signals	ms_ss64_raw_w780	(100, 2030, 256)	raw signals
linear_raw_w850	(100, 2030, 256)	raw signals	ms_ss64_raw_w800	(100, 2030, 256)	raw signals
ms_raw_w700	(100, 2030, 256)	raw signals	ms_ss64_raw_w760	(100, 2030, 256)	raw signals
ms_raw_w730	(100, 2030, 256)	raw signals	ms_ss64_raw_w780	(100, 2030, 256)	raw signals
ms_raw_w760	(100, 2030, 256)	raw signals	ms_ss64_raw_w800	(100, 2030, 256)	raw signals
ms_raw_w780	(100, 2030, 256)	raw signals	ms_ss64_raw_w850	(100, 2030, 256)	raw signals
ms_raw_w800	(100, 2030, 256)	raw signals	ms_ss128_raw_w700	(100, 2030, 256)	raw signals
ms_raw_w850	(100, 2030, 256)	raw signals	ms_ss128_raw_w730	(100, 2030, 256)	raw signals
ms_ss32_BP_w700	(100, 256, 256)	image reconstructions	ms_ss128_raw_w760	(100, 2030, 256)	raw signals
ms_ss32_BP_w730	(100, 256, 256)	image reconstructions	ms_ss128_raw_w780	(100, 2030, 256)	raw signals
ms_ss32_BP_w760	(100, 256, 256)	image reconstructions	ms_ss128_raw_w800	(100, 2030, 256)	raw signals
ms_ss32_BP_w780	(100, 256, 256)	image reconstructions	ms_ss128_raw_w850	(100, 2030, 256)	raw signals
ms_ss32_BP_w800	(100, 256, 256)	image reconstructions			
ms_ss32_BP_w850	(100, 256, 256)	image reconstructions			

## **G Considerations when using our datasets**

As the volunteers in experimental datasets are considered to be healthy, anonymization is done one way, and true identities of volunteers are not possible to trace. Furthermore, to the best of our knowledge, usage of our proposed datasets cannot pose threat to the volunteers, even with malicious intent. However, as with all clinical data, it should be acknowledged that our dataset would represent a particular subset of all potential forearm images collected at respective wavelengths and transducer arrays. Accordingly, any subsequent work that makes use of our datasets for validation purposes need to ensure that the diversity of our datasets is sufficient for their target application.

## **H Persistence of proposed datasets**

All datasets are hosted on Libdrive at ETH Zurich Research Collection repository. Accordingly, the datasets have a landing page with DOI and metadata. Data will be freely accessible with no restriction. As per ETH Zurich Research Collection documentation, the dataset is guaranteed to have a retention period for 10 years. Based on the frequency of usage by the community, the Research Collection may continue to host the data beyond the 10 year period.

## **I Author statement on data license**

The dataset is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC-BY-NC).

## **J Training and evaluation code**

In oadat-evaluate repository, along with saved model weights, we provide all necessary scripts to train modUNet from scratch for all 44 experiments as well as evaluating them over the whole test set. We provide standalone examples on the repository landing page to show how to;

- (i) load a pretrained model, for example, to do inference or finetuning it,
- (ii) train modUNet from scratch for either of the two tasks for a given experiment,
- (iii) evaluate a serialized model, whether it is one of the pretrained models we provide as is, or any other Tensorflow model that is already serialized, and
- (iv) use our provided data loader to read any sample from OADAT.