MELBA Journal Machine Learning for Biomedical Imaging

Learning Neural Parametric 3D Breast Shape Models for Metrical Surface Reconstruction From Monocular RGB Videos

Maximilian **Weiherer** [1,2], Antonia **von Riedheim** [3], Vanessa **Brébant** [3], Bernhard **Egger** [1*], Christoph **Palm** [2*]

1 Visual Computing Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
2 Regensburg Medical Image Computing (ReMIC), OTH Regensburg, Regensburg, Germany
3 Department for Plastic, Hand and Reconstructive Surgery, University Hospital Regensburg, Regensburg, Germany
* *equal contribution*

Abstract

We present a neural parametric 3D breast shape model and, based on this model, introduce a low-cost and accessible 3D surface reconstruction pipeline capable of recovering accurate breast geometry from a monocular RGB video. In contrast to widely used, commercially available yet expensive 3D breast scanning solutions *and* existing low-cost alternatives, our method requires neither specialized hardware nor proprietary software and can be used with any device that is able to record RGB videos. The key building blocks of our pipeline are a state-of-the-art, off-the-shelf Structure-from-Motion pipeline, paired with a parametric breast model for robust surface reconstruction. Our model, similarly to the recently proposed implicit Regensburg Breast Shape Model (iRBSM), leverages implicit neural representations to model breast shapes. However, unlike the iRBSM, which employs a single *global* neural Signed Distance Function (SDF), our approach—inspired by recent state-of-the-art face models—decomposes the implicit breast domain into multiple smaller regions, each represented by a *local* neural SDF anchored at anatomical landmark positions. When incorporated into our surface reconstruction pipeline, the proposed model, dubbed liRBSM (short for *localized* iRBSM), significantly outperforms the iRBSM in terms of reconstruction quality, yielding more detailed surface reconstruction than its global counterpart. Overall, we find that the introduced pipeline is able to recover high-quality and metrically correct 3D breast geometry within an error margin of less than 2 mm. Our method is fast (requires less than six minutes), fully transparent and open-source, and together with the model publicly available at `https://rbsm.re-mic.de/local-implicit`.

Keywords

3D Reconstruction, Shape Modeling, Implicit Neural Representations, Breast Surface Reconstruction, 3D Breast Imaging

Article informations

https://doi.org/10.59275/j.melba.2026-8b23

Volume 2026, Received: 07/2025, Published 02/2026

Corresponding author: maximilian.weiherer@fau.de

Special issue: MELBA–BVM 2025 Special Issue
Guest editors: Andreas Maier, Thomas Deserno, Heinz Handels, Klaus Maier-Hein, Christoph Palm, Thomas Tolxdorff, Katharina Breininger

1. Introduction

Despite recent advances and emerging applications, research evolving around 3D parametric models for the female breast remains rare. Since the seminal work of Seo et al. (2007), who developed the first classical mesh-based statistical breast shape model, the field has seen relatively little progress—especially when compared to the face domain (Egger et al., 2020)—likely due to the sensitivity of the data and the resulting lack of publicly available 3D breast scan datasets. Nevertheless, over the years, 3D parametric breast models have found use in a range of medical applications, including automatic breast volume estimation from surface scans (Seo et al., 2007; Göpper et al., 2020), plastic surgery simulation (Kim et al., 2008), model-based breast segmentation in Magnetic Resonance Imaging (MRI) volumes (Gallego and Martel, 2011), and surface reconstruction from 2D photographs (Ruiz et al., 2018) or point clouds obtained using depth cameras (Mazier et al., 2021). Potential future use cases beyond the clinical domain include virtual try-on and custom bra design. A common limitation shared by *all* of these approaches is the absence of publicly available code and trained models, which prevents broader

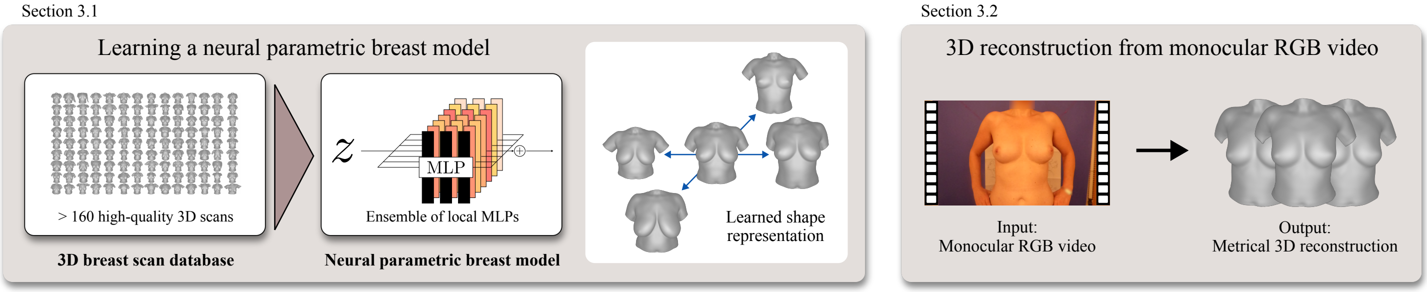

Figure 1: **Contributions.** We propose a 3D parametric breast model along with a model-based surface reconstruction pipeline that is able to accurately reconstruct 3D breast surfaces from a single monocular RGB video. Our model, trained on over 160 breast scans, builds upon a *localized* neural implicit representation and represents breast geometry as an ensemble of local MLPs instead of a single global network, significantly improving the level of detail.

adoption and hinders future research. As a consequence and as of today, none of the above-mentioned methods is effectively accessible to clinicians or other researchers, and only two publicly available 3D breast models exist: the Regensburg Breast Shape Model (RBSM; Weiherer et al. (2023)) and the recently proposed state-of-the-art *implicit* RBSM (iRBSM; Weiherer et al. (2025)). While the RBSM is a traditional mesh-based statistical shape model built by applying Principal Component Analysis (PCA) to a set of 110 non-rigidly registered 3D breast scans, the iRBSM takes a fundamentally different approach by leveraging implicit neural representations to learn a realistic and expressive parametric breast model from over 160 subjects. Instead of using triangular meshes as surface representation, the iRBSM represents breast shapes as the zero-level set of a neural Signed Distance Function (SDF), modeled with a simple coordinate-based Multi-Layer Perceptron (MLP). Once trained, triangular meshes can be extracted from the implicit volume using marching cubes algorithm (Lorensen and Cline, 1987). A major advantage of implicit representations over PCA-based models is that they eliminate the need for training data to be in correspondence, entirely removing the reliance on computationally demanding and error-prone non-rigid surface registration—a task that is particularly difficult in feature-less and partially occluded breast shapes as discussed in (Weiherer et al., 2025).

In this work, we build upon the neural implicit representation of the iRBSM and propose a novel open-access 3D parametric breast model. Our model is inspired by recent state-of-the-art face models (Zheng et al., 2022; Giebenhain et al., 2023; Potamias et al., 2025) and employs a decomposition of the implicit breast domain into multiple smaller regions, each of which is modeled using a *local* neural SDF anchored at anatomical landmark positions instead of a single global network. Compared to the iRBSM, our model provides a significantly higher level of detail, enabling the recovery of fine anatomical structures such as skin folds and nipples. Due to its local nature, we refer to our model as liRBSM (short for *localized* iRBSM).

Largely independent from these developments—with the exception of (Ruiz et al., 2018; Mazier et al., 2021)—3D breast scanning (or imaging) has become standard practice in plastic surgery over the past two decades, often used to digitally assess breast volume (Kovacs et al., 2007; Lee et al., 2016; Seoud et al., 2017; Gouveia et al., 2021) or symmetry (Eder et al., 2012; Brébant et al., 2022; Noisser et al., 2022; Bai et al., 2023), perform anthropometric measurements (Hartmann et al., 2021; Leusink et al., 2021; Wang et al., 2025b), or simulate surgical outcomes (Kim et al., 2008; Georgii et al., 2014). To create surface scans of the breast, most researchers and clinicians rely on commercially available 3D scanning solutions, typically requiring special hardware and proprietary software, such as Artec's handheld Eva[1] or Canfield's portable Vectra H2[2] or static Vectra XT[3] system. Although established, well validated for numerous breast-related applications, and known to produce highly accurate 3D breast scans, due to their high purchase price (currently ∼10,000 Euros for Artec Eva; around 20,000 Dollars for Vectra H2), these systems are typically only available to large hospitals or clinical facilities. As a result, a recent line of research started to investigate various cost-effective alternatives based on *consumer-grade* hardware, most notably the iPhone (Pinto et al., 2022; Han et al., 2023; Rudy et al., 2024; Behrens et al., 2024; Dijkman et al., 2024; Kyriazidis et al., 2025; Chrobot et al., 2025), in combination with freely available software such as the 3D Scanner App[4]. While these methods offer more accessibility than Artec's or Canfield's systems, they are still limited to certain devices and rely on closed-source software, restricting transparency and full control over the

---

[1] `https://www.artec3d.com/portable-3d-scanners/artec-eva`

[2] `https://www.canfieldsci.com/imaging-systems/vectra-h2-body`

[3] `https://www.canfieldsci.com/imaging-systems/vectra-xt-3d-imaging-system`

[4] `https://www.3dscannerapp.com`

surface reconstruction process which ultimately raises privacy concerns and imposes laborious workflows (Chrobot et al., 2025).

To tackle these shortcomings and based on the proposed parametric breast model, we introduce a new low-cost, accessible, and accurate 3D breast surface reconstruction method that only requires a monocular RGB video as input. Our method produces accurate metrical 3D reconstructions, is fast (takes less than six minutes to run on a standard computer with a consumer-grade graphics card), well-documented, and fully open-source. Moreover, it can be used in combination with any capturing device that is able to record RGB videos. From a technical perspective, our pipeline combines a recent state-of-the-art and off-the-shelf Structure-from-Motion pipeline with robust model-based surface reconstruction.

Our key contributions are outlined in Figure 1 and can be summarized as follows:

- We present a new 3D parametric breast model that builds upon a *localized* neural implicit representation, significantly increasing the level of detail.
- We propose a low-cost and accessible 3D surface reconstruction pipeline that is able to accurately recover metrical breast surfaces from just a single monocular RGB video, captured using commodity hardware.
- To encourage further research and widespread adoption, we publicly release our model and surface reconstruction pipeline along with an easy-to-use graphical user interface that runs on all common operating systems and (optionally) without a graphics card. Both available at `https://rbsm.re-mic.de/local-implicit`.

2. Related Work

We begin with a comprehensive overview of existing parametric 3D breast shape models, followed by a review of related work on 3D breast surface reconstruction methods. To contextualize our work within the broad landscape of shape modeling and (model-based) surface reconstruction from RGB images, we finally briefly summarize related literature on full-body models and reconstruction techniques.

2.1 Parametric 3D Breast Models

With the exception of (Gallego and Martel, 2011), all existing models have been trained on 3D breast scans acquired in a standing position.

Seo et al. (2007) were the first to build a parametric breast model from 28 scans to analyze breast volume and surface measurements. They assumed symmetric breasts by simply mirroring the right breast.

Study	Representation	Method	Training Data	Pose	#Subjects	Public?
Seo et al. (2007)	Mesh-based	PCA	3D breast scans	Standing	28	✗
Gallego and Martel (2011)	Mesh-based	PCA	Segmented MRIs	Prone	415	✗
Ruiz et al. (2018)	Mesh-based	PCA	3D breast scans	Standing	310	✗
Mazier et al. (2021)	Mesh-based	PCA	Blendshapes	Standing	55	✗
Weiherer et al. (2023)	Mesh-based	PCA	3D breast scans	Standing	110	✓
Weiherer et al. (2025)	Implicit (Global)	Deep	3D breast scans	Standing	168	✓
Ours	Implicit (Local)	Deep	3D breast scans	Standing	168	✓

Table 1: **Overview of existing 3D parametric breast models.** Our model is the first to use a *localized* neural implicit representation.

Gallego and Martel (2011) learned a PCA-based model from 415 MRI-extracted 3D breast surfaces captured in *prone* position. Their model is specifically tailored for automatic, model-based breast segmentation in MRI volumes; hence, it was constructed using data only from a single breast rather than the full thoracic region.

Ruiz et al. (2018) built a 3D breast model from 310 scans and subsequently fit their model onto 3D breast scans or a set of three unconstrained 2D photographs taken from frontal and lateral views (± 90 degrees).

Mazier et al. (2021) proposed a *rigged* 3D breast model built from 55 artist-created blendshapes to transfer surgical reference patterns drawn on the model's mean shape onto any patient in any position by non-rigidly registering the model to a patient's 3D breast scan. Due to the synthetic nature of the data, their model is likely to generate rather unnatural-looking breast shapes, and the overall accuracy measured in terms of mean absolute error and landmark error of the surface registrations is moderate.

Weiherer et al. (2023) published the first publicly available PCA-based model of the female breast, trained on 110 high-quality 3D breast scans. While representing a notable advancement towards making 3D parametric breast models accessible, their model exhibits correspondence errors as a result of occluded underbusts in large or sagged breasts—a common problem in classical PCA-based *breast* models built from surface-only 3D breast scans (Seo et al., 2007).

To account for this, Weiherer et al. (2025) recently introduced the iRBSM, the first *implicit* 3D breast shape model. Based on correspondence-free neural implicit representations (representing breast shapes as the zero-level set of implicit surfaces parametrized with an MLP) and trained on 168 breast scans, their model is able to generate diverse yet plausible breast shapes. However, when fitted onto point clouds, the global nature of the iRBSM misses fine details such as skin folds and nipples.

Only the RBSM (Weiherer et al., 2023) and iRBSM (Weiherer et al., 2025) are publicly available. A compact summary of all models is given in Table 1.

Study	Input Data	Capturing Device	Software (SW)	Compatible With	Handheld?	Public SW?	Open-source SW?
Costa et al. (2014)	RGB-D	Microsoft Kinect	Kinect SDK	Any RGB-D camera*	✗	✓	✗
Henseler et al. (2014)	RGB-D	Microsoft Kinect	Kinect SDK + Custom	Any RGB-D camera*	✗	✗	✗
Pöhlmann et al. (2014)	RGB-D	Microsoft Kinect	Kinect SDK	Any RGB-D camera*	✗	✓	✗
Wheat et al. (2014)	RGB-D	2 Microsoft Kinects	Kinect SDK + Custom	Any RGB-D camera*	✗	✗	✗
Lacher et al. (2015)	RGB-D	Microsoft Kinect	Kinect SDK + Custom	Any RGB-D camera*	✗	✗	✗
Henseler et al. (2016)	RGB-D	Microsoft Kinect	Kinect SDK + Custom	Any RGB-D camera*	✗	✗	✗
Pöhlmann et al. (2017)	RGB-D	Microsoft Kinect	Kinect SDK	Any RGB-D camera*	✓†	✓	✗
Koban et al. (2018)	RGB-D	3D scanner (Sense)	Supplied	Nothing	✓	✓	✗
Lacher et al. (2019)	RGB-D	Microsoft Kinect	Kinect SDK + Custom	Any RGB-D camera*	✗	✗	✗
Oranges et al. (2019)	RGB-D	3D scanner (Structure Sensor 3D)	Supplied	Nothing	✓	✓	✗
Tong et al. (2020)	RGB-D	2 mod. 3D scanners (HP Pro S3)	Supplied	Nothing	✗	✗	✗
Luu et al. (2021)	RGB-D	3D scanner (Custom)	Custom	Nothing	✗	✗	✗
Pinto et al. (2022)	RGB-D	iPhone 11 (TrueDepth)	Scandy Pro	iPhone & iPad w\ TrueDepth	✗	✓	✗
Han et al. (2023)	RGB-D	iPhone 12 Pro (LiDAR)	Custom	iPhone & iPad w\ LiDAR‡	✓	✗	✗
Behrens et al. (2024)	RGB-D	iPhone 11 Pro Max (TrueDepth)	3D Scanner App	iPhone & iPad w\ TrueDepth	✗	✓	✗
Dijkman et al. (2024)	RGB-D	iPhone XR (TrueDepth)	Heges	iPhone & iPad w\ TrueDepth	✓	✓	✗
Fu et al. (2024)	RGB-D	Intel RealSense	3D Slicer + Custom	Any RGB-D camera*	✓	✗	✗
Rudy et al. (2024)	RGB-D	iPhone X (TrueDepth)	Scandy Pro	iPhone & iPad w\ TrueDepth	✓	✓	✗
Kyriazidis et al. (2025)	RGB-D	iPhone 15 Pro (LiDAR)	3D Scanner App	iPhone & iPad w\ LiDAR	✓	✓	✗
de Heras Ciechomski et al. (2012)	Photo	Not specified	Custom	Any RGB camera	✓	✗	✗
Henseler et al. (2013)	Photo	8 DSLR cameras	Dimensional Imaging 3D	Nothing	✗	✗	✗
Ruiz et al. (2018)	Photo	Not specified	Custom	Any RGB camera	✓	✗	✗
Chrobot et al. (2025)	Photo	iPhone 15	3D Scanner App	Newer iPhone & iPad	✓	✓	✗
Ours	Video	iPhone 12 Mini	Custom	Any RGB camera	✓	✓	✓

Table 2: **Overview of existing low-cost 3D breast surface reconstruction methods.** The list comprises all works that propose or use affordable alternatives to commercial 3D scanning systems from companies such as Artec and Canfield, all of which rely on *commodity* and *non-medical* hardware. Among these, our proposed method stands out as the only pipeline that is (i) compatible with the widest range of devices—that is, any device capable of recording a standard RGB video, (ii) handheld and publicly available without requiring the purchase of specialized hardware, and (iii) based entirely on open-source software. We highlight in green the most accessible pipelines, which we define as pipelines that are handheld (hence do not require extra equipment such as tripods) and rely only on publicly available software. *Compatibility might require significant adaptation or exchange of the Kinect SDK; †Multiple setups are proposed, and only some of them are handheld; ‡Compatibility unclear due to self-developed software (software was developed under iOS, hence compatibility is assumed for iPhone and iPad).

2.2 Breast Surface Reconstruction

Next, we revisit 3D breast surface reconstruction pipelines, limiting our scope to low-cost approaches that utilize commodity hardware (*i.e.*, methods proposing alternatives to commercial systems from Artec or Canfield). A comprehensive overview of these works is provided in Table 2.

RGB-D. The vast majority of the existing pipelines (19 out of 23) use RGB and depth data (referred to as *RGB-D*), acquired using RGB-D cameras such as Microsoft Kinect (Costa et al., 2014; Henseler et al., 2014; Pöhlmann et al., 2014; Wheat et al., 2014; Lacher et al., 2015; Henseler et al., 2016; Lacher et al., 2019) or Intel RealSense (Fu et al., 2024). In this context, Lacher et al. (2017) presented a systematic comparison between Microsoft Kinect v1 and v2 and established 3D scanning solutions (Artec Eva and 3dMD[5] stereophotogrammetry system) based solely on open-source software. They found that both devices produce satisfactory 3D breast surface reconstructions within an error margin of 3 mm compared to the ground truth.

Lacher et al. (2019) introduced a *non-rigid*, template-free method for surface reconstruction from an RGB-D stream captured with a Microsoft Kinect. While being able to recover clinical-quality 3D breast scans without motion or breathing artifacts, long runtimes of 1–2 hours prohibit their method from widespread use in daily clinical practice.

Recently, Fu et al. (2024) proposed a portable setup involving the Intel RealSense D415 camera. Although their method reportedly produces 3D breast scans with an average landmark error of less than 1.5 mm, they employ custom software not publicly available.

Besides RGB-D cameras, researchers explored 3D breast surface reconstruction pipelines based on affordable (non-medical) structured light scanners, either commercially available and handheld (Koban et al., 2018; Oranges et al., 2019) or custom-built and static (Tong et al., 2020; Luu et al., 2021). All of these methods rely on proprietary software and require purchasing special hardware.

A third recent line of work investigated the use of the iPhone's LiDAR sensor (Han et al., 2023; Kyriazidis et al., 2025) and TrueDepth camera (Pinto et al., 2022; Rudy et al., 2024; Behrens et al., 2024; Dijkman et al., 2024) for 3D breast surface reconstruction. Both studies that utilize the iPhone's LiDAR sensor report very similar findings, evaluating surface reconstruction quality based on anthropometric distances measured between anatomical landmarks. Most measurements show reasonable agree-

---

[5] `https://www.3dmd.com`

ment with ground truth values obtained via measuring tape, with the notable exception of the nipple-to-inframammary fold distance. This discrepancy is due to the low spatial resolution of the iPhone's LiDAR sensor, which captures only coarse surface geometry and completely fails to represent high-frequency details—an issue clearly visible in the poor reconstructions shown in Kyriazidis et al. (2025). On the other hand, the iPhone's TrueDepth camera appears capable of reconstructing high-fidelity surfaces. Rudy et al. (2024) report an average surface-to-surface distance of approximately 1.5 mm between 3D breast reconstructions obtained with the TrueDepth camera and ground truth Vectra H2 scans. These results confirm earlier findings by Pinto et al. (2022); however, Dijkman et al. (2024) and Behrens et al. (2024) report only sufficient reconstruction quality when assessed in the context of breast volume estimation.

In summary, among all low-cost depth-based pipelines, only the Microsoft Kinect, when paired with custom software (Lacher et al., 2019), and the iPhone's TrueDepth camera demonstrate the ability to reconstruct high-fidelity 3D breast surfaces, positioning them as a real alternative to commercial 3D scanning solutions. Nevertheless, both approaches require special hardware and depend on either non-publicly available or proprietary software.

Photo. Only a few methods have been proposed for reconstructing 3D breast surfaces from multi-view images. de Heras Ciechomski et al. (2012) presented a web-based application for 3D breast surface reconstruction using three 2D photographs taken from frontal and lateral views, along with user-provided anthropometric measurements. They reported a mean surface-to-surface reconstruction error ranging between 2 and 4 mm. The method requires the user to select landmarks on *all* three images (22 in total) and is not publicly available.

Henseler et al. (2013) designed a static setup using eight DSLR cameras and employed photogrammetry to reconstruct the 3D surface of the breast. Their setup is fairly large, complex, and impractical for everyday use, requiring significant effort and space to assemble.

Based on a custom-built 3D parametric breast model, Ruiz et al. (2018) proposed a 3D surface reconstruction technique that operates on three 2D photographs captured from frontal and lateral perspectives (± 90 degrees). The method fits the parametric model to a sparse 3D point cloud reconstructed via triangulation of an automatically detected set of corresponding 2D landmarks across the three images. Prior to model fitting, estimated camera parameters required for triangulation and the back-projected 3D landmarks are jointly refined using bundle adjustment. From a technical perspective, their pipeline is very similar to our method, except for the fact that we obtain the sparse point cloud by applying Structure-from-Motion (SfM) to a set of multi-view images extracted from a video sequence instead of three images (note that they essentially also use SfM, but with a custom keypoint detector). On the downside, they only evaluate their method based on the re-projection error of the 3D landmarks, leaving the actual quality of the reconstructed geometry unclear. Furthermore, unlike ours, their pipeline is not publicly available.

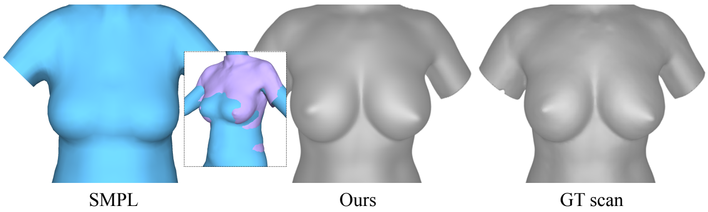

Figure 2: **Body model vs. breast-specific model**. We fit the female version of the popular SMPL model (Loper et al., 2015) to one of our 3D breast scans, demonstrating that existing full-body models, despite including the chest region, can not accurately represent nude breast shapes.

Recently, Chrobot et al. (2025) employed an iPhone 15 and the 3D Scanner App's photogrammetry mode to reconstruct 3D breast surfaces from a series of 2D photographs, automatically taken every 0.8 seconds as controlled by the application. They report only moderate accuracy, assessed using 14 anthropometric distances measured on the smartphone-based 3D reconstructions and ground truth Vectra H2 scans.

In contrast to the aforementioned methods, our pipeline uses a simple monocular RGB video as input. We chose to work with video sequences because they are easier and faster to acquire than individual photographs, thereby minimizing capture time and reducing motion and breathing artifacts. Similar to Ruiz et al. (2018), we adopt a model-based 3D surface reconstruction approach, which is particularly robust in noisy settings, as expected due to the degraded image quality of video frames. However, unlike their method, we employ a *neural implicit* 3D breast model, enabling more detailed and accurate surface reconstructions. Our approach is publicly available, relies solely on open-source software, and does not require specialized hardware—any standard RGB camera can be used.

2.3 Human Body Models and Reconstruction

We finally review related literature on human body models (that naturally include the breast region) and recent methods for neural implicit surface reconstruction from a sequence of multi-view RGB images.

Models. Probably the most well-known full-body model is SMPL (Loper et al., 2015), a PCA-based parametric human pose and shape model that uses Linear Blend Skinning to allow for pose-dependent shape deformations and whose

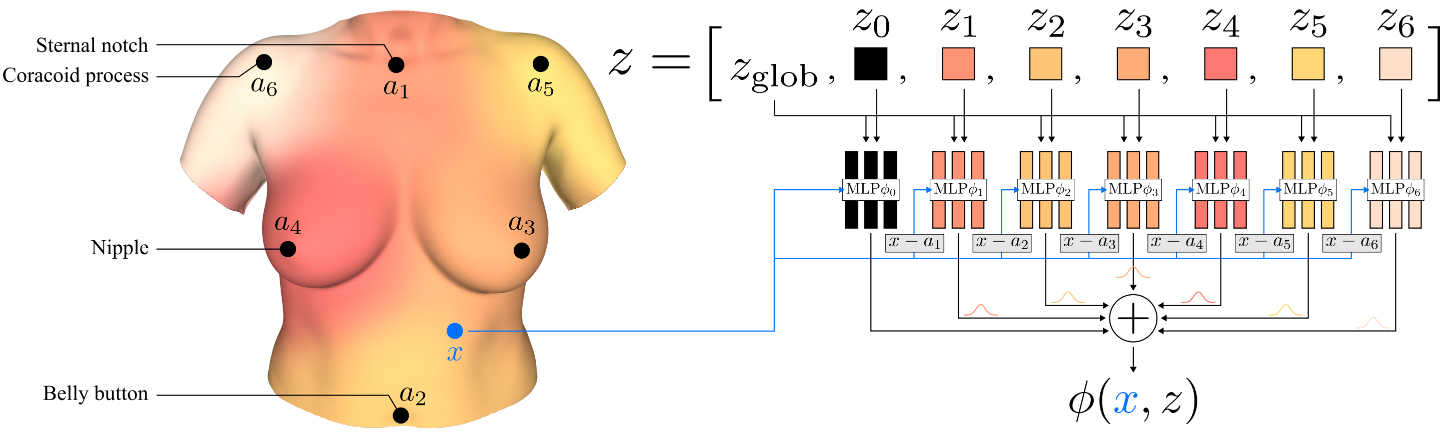

Figure 3: **Overview of our model's architecture.** Instead of using a single global MLP to represent breast shapes, following Giebenhain et al. (2023), we partition the implicit domain into six smaller regions, centered around anchor points $\{a_1, a_2, \ldots, a_6\}$ placed at anatomical landmark positions. Each region is represented by a shallow MLP ϕ_i, conditioned on a local latent code z_i. To account for areas distant from any anchor, we introduce an additional MLP ϕ_0 that shares the same representation as the local MLPs. The final SDF $\phi(x, z)$ at a query point x is obtained by blending the outputs of all MLPs using a Gaussian weighting scheme. An illustration of the spatial influence of each local MLP is shown on the left, visualized on the resulting model's mean shape.

shape space is learned from the CAESAR (Robinette et al., 2002) dataset. While the female version of SMPL does include the chest region, as demonstrated in Figure 2, it is not able to accurately reconstruct nude breast shapes as it has been trained on 3D scans of women wearing a light bra. Numerous follow-ups improved upon SMPL, sticking to its mesh-based representation. Notably, GHUM (Xu et al., 2020) replaces the PCA-based, linear shape space with a non-linear shape space learned using a Variational Auto Encoder. Introduced by Osman et al. (2020), STAR constrains SMPL's deformations to be more realistic, SUPR (Osman et al., 2022) uses a factorized representation based on part models, and SKEL (Keller et al., 2023) re-rigs SMPL with a biomechanics skeleton.

With the rise of neural implicit representations, Neural Parametric Models (NPMs; Palafox et al. (2021)) propose a full-body model with disentangled human shape and pose, representing shape in canonical pose as the zero-level set of a latent-conditioned neural SDF, which is deformed through a learned deformation field to yield shapes in posed space. NPMs differ from other neural implicit approaches (Deng et al., 2020; Mihajlovic et al., 2021; Alldieck et al., 2021; Mihajlovic et al., 2022; Palafox et al., 2022; Mihajlovic et al., 2025) in that they do *not* require any domain-specific annotations (such as a kinematic chain, skeleton, or part segmentations) or knowledge, therefore can be easily adapted to other parts of the human body. Existing neural implicit body models are typically trained on human motion datasets captured from clothed subjects; hence, these models can not accurately represent breast shapes.

Neural Parametric Head Models (NPHMs; Giebenhain et al. (2023)) built upon NPMs and proposed a part-based head model, decomposing the implicit head domain into multiple smaller regions each of which is represented by a *local* neural SDF instead of a single global network. Facial expressions are modeled as deformations of shape in canonical space, analogously to pose in NPMs. Conceptually similar to (Zheng et al., 2022), this strategy significantly increases the level of detail, reaching state-of-the-art results in dynamic head reconstruction from monocular RGB videos (Giebenhain et al., 2024).

In this work, we adopt NPHMs' architecture and use it to model breast shapes. We note that, in this context, the previously introduced iRBSM (Weiherer et al., 2025) follows the architecture of NPMs without pose deformations, whereas our proposed model, liRBSM, resembles NPHMs' architecture but omits the facial expression network.

Reconstruction. Most of the state-of-the-art neural surface reconstruction methods employ a *prior-free* approach, combining implicit surface representations (such as occupancy functions or SDFs) with differentiable volume rendering (Niemeyer et al., 2020; Yariv et al., 2021; Wang et al., 2021; Oechsle et al., 2021; Yu et al., 2022; Li et al., 2023; Wang et al., 2023; Cao and Taketomi, 2024). Per-scene optimization methods typically suffer from long run-times and, due to missing geometric priors, degenerate quickly if only a sparse set of images is available.

In contrast, *prior-based* approaches work well in a few-shot scenario. More recent methods combine differentiable volume rendering with neural parametric models (Ramon et al., 2021; Giebenhain et al., 2024) and optional displacement fields for improved geometrical detail (Grassal et al., 2022; Caselles et al., 2025) and are primarily being developed in the head domain. Fitting a model to a set of images is then done in an analysis-by-synthesis fashion, and requires a careful initialization, (facial) landmarks and/or silhouette masks detected in *every* image, and photometric losses to supervise the optimization process.

Contrary to rendering-based 3D reconstruction methods and inspired by the recent success of differentiable SfM pipelines (Wang et al., 2024b,a; Duisterhof et al., 2025; Wang et al., 2025a) in terms of robustness against photometric changes and few-shot capabilities, we opted for a more classical approach and fit our 3D neural parametric breast model directly onto a *point cloud* obtained from SfM. This model-based surface reconstruction strategy is robust against noise but retains the few-shot capabilities of prior-based approaches while avoiding cumbersome per-image landmark detection, which is difficult in the breast domain due to the lack of automatic landmark detectors. Moreover, it bypasses the necessity of large paired 2D-3D datasets common in human mesh recovery (Tian et al., 2023), as no such data is available in the breast domain. As we will show later, our method achieves state-of-the-art results and works well with as few as three images extracted from the input video.

3. Methods

We will now describe our model and the proposed 3D surface reconstruction pipeline in detail.

3.1 Localized Neural Implicit Breast Model

Our model adopts the neural implicit representation of the iRBSM, which represents breast shapes as the zero-level set of a latent-conditioned neural SDF. However, inspired by recent state-of-the-art face models (Zheng et al., 2022; Giebenhain et al., 2023; Potamias et al., 2025) and to increase the level of detail, we further employ a space partitioning approach and decompose the implicit breast domain into multiple smaller regions, each represented by a *local* SDF anchored at anatomical landmark positions. Specifically, as illustrated in Figure 3, instead of a single global MLP that models the entire breast geometry, we use an ensemble of K *local* MLPs

$$\begin{aligned} \phi_k : \mathbb{R}^3 \times \mathbb{R}^{L_{\text{glob}}} \times \mathbb{R}^{L_{\text{loc}}} &\to \mathbb{R} \\ (x, z_{\text{glob}}, z_k) &\mapsto \phi_k(x - a_k, z_{\text{glob}}, z_k) \end{aligned} \tag{1}$$

centered around anchor points $a_k \in \mathbb{R}^3$ and blended to yield a parametric model of the form

$$\phi(x, z) = \sum_{k=1}^{K} w_k(x)\phi_k(x - a_k, z_{\text{glob}}, z_k), \tag{2}$$

where $z_{\text{glob}} \in \mathbb{R}^{L_{\text{glob}}}$ is a global (per-shape) latent code, $z_k \in \mathbb{R}^{L_{\text{loc}}}$ are local latent codes for the individual MLPs, and $z := [z_{\text{glob}}, z_1, z_2, \ldots, z_K]$. The blending weights $w_k(x)$ in Eq. (2) are defined as

$$w_k(x) = \frac{w_k^*(x)}{\sum_{k'=1}^{K} w_{k'}^*(x)}, \tag{3}$$

where

$$w_k^*(x) = \exp\left(-\frac{\|x - a_k\|_2^2}{2h^2}\right) \tag{4}$$

is an isotropic Gaussian kernel with bandwidth $h > 0$. Anchor points $a = [a_1, a_2, \ldots, a_K]$ are predicted using a small MLP $\phi_{\text{anc}} : \mathbb{R}^{L_{\text{glob}}} \to \mathbb{R}^{K\times 3}$ conditioned on the global latent code, *i.e.*, $a = \phi_{\text{anc}}(z_{\text{glob}})$.

Finally, since each of the local MLPs focuses on a specific part of the breast, following Giebenhain et al. (2023), we add an additional MLP to our formulation in Eq. (2) that captures global context far away from the anchor points (effectively resulting in $K + 1$ regions in which we divide the domain). This MLP shares the same representation as the local MLPs introduced in Eq. (1), is anchored at the origin, and uses a constant weight response. We refer to this network as *background MLP*.

3.1.1 Training

We train our model in an auto-decoder fashion (Park et al., 2019), jointly optimizing the network's parameters and latent codes. Given a dataset of 3D breast scans, each of which is represented as normalized and centered point cloud $X = \{x_1, x_2, \ldots, x_n\} \subset [-1, 1]^3$ with normals $N = \{n_1, n_2, \ldots, n_n\} \subset \mathbb{R}^3$ and ground truth anchor points $a_{\text{gt}} \in \mathbb{R}^{K\times 3}$ placed at expert-annotated anatomical landmark positions, we minimize the same loss function as in (Giebenhain et al., 2023), namely,

$$\mathcal{L} = \mathcal{L}_{\text{SDF}} + \lambda_5\mathcal{L}_{\text{anc}} + \lambda_6\|z\|_2, \tag{5}$$

where

$$\begin{aligned} \mathcal{L}_{\text{SDF}} = \sum_{i=1}^{n} \lambda_1|\phi(x_i, z)| + \lambda_2\|\nabla_{x_i}\phi(x_i, z) - n_i\|_2 + \\ \mathbb{E}_{x\sim D}\left[\lambda_3|\|\nabla_x\phi(x, z)\|_2 - 1| + \lambda_4\exp(-\alpha|\phi(x, z)|)\right] \end{aligned} \tag{6}$$

ensures that ϕ represents a valid SDF,

$$\mathcal{L}_{\text{anc}} = \|\phi_{\text{anc}}(z_{\text{glob}}) - a_{\text{gt}}\|_2 \tag{7}$$

is used to supervise anchor predictions, and the last term regularizes latent codes. The first term in Eq. (6) enforces that the predicted SDF is zero at the given surface points and that its gradients on the zero-level set match the ground truth surface normals. The second term serves as a regularizer, consisting of an Eikonal constraint that encourages the SDF's gradient to have unit norm almost everywhere, and a volume loss (with parameter $\alpha \gg 1$) that penalizes

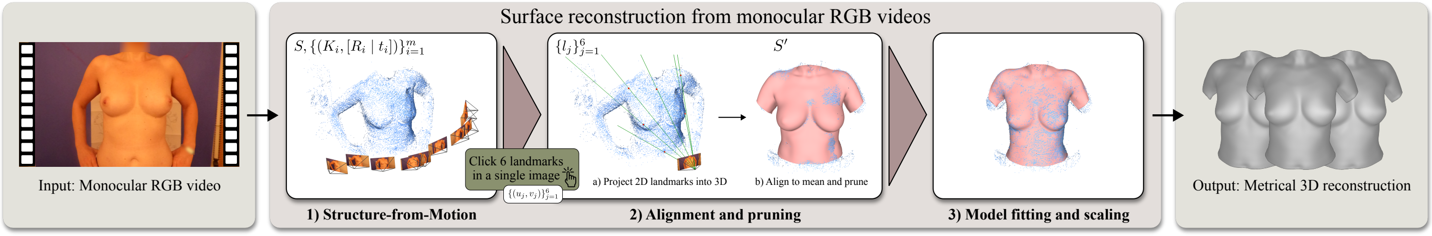

Figure 4: **Surface reconstruction from monocular RGB videos.** We present a method for low-cost *and* accurate metrical breast surface reconstruction from just a single monocular RGB video, acquired using commodity hardware (such as smartphones). Our pipeline starts by applying Structure-from-Motion to a set of frames extracted from the input video, resulting in a sparse point cloud and camera parameters. We then align the estimated point cloud to our model's mean shape, prune away background points, fit our model to the resulting point cloud, and finally recover real-world scale. Depending on the available hardware, the entire reconstruction can be executed in under six minutes.

off-surface points for creating SDF values close to zero. Off-surface points $D \subset \mathbb{R}^3$ are obtained by randomly offsetting the surface points and adding some points uniformly sampled in $D := [-1,1]^3 \cup \{x_1 + \epsilon_1, x_2 + \epsilon_2, \ldots, x_n + \epsilon_n\}$ with $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. We refer the reader to (Gropp et al., 2020) and (Sitzmann et al., 2020) for further details about this loss.

3.1.2 Inference

Given a 3D breast scan as *unoriented* point cloud (without normals) $X' = \{x'_1, x'_2, \ldots, x'_{n'}\}$ uniformly scaled to $[-1,1]^3$ and centered at the origin, at test-time, we obtain the corresponding latent code z^* via maximum a posteriori estimation as proposed in (Park et al., 2019). In particular, we fix the model's parameters and optimize

$$z^* = \operatorname*{arg\,min}_{z \in \mathbb{R}^L} \left\{ \sum_{i=1}^{n'} |\phi(x'_i, z)| + \lambda \|z\|_2 \right\}, \tag{8}$$

where $L = L_{\text{glob}} + (K+1)L_{\text{loc}}$ denotes the total latent dimensionality of the model, and $\lambda \geq 0$ can be used to filter out noise. We then apply the inverse scale and translation to the extracted mesh in order to recover its original position and scale.

3.2 Surface Reconstruction From Monocular RGB Videos

Given a sequence of RGB images $I = \{I_1, I_2, \ldots, I_m\}$ extracted from the input video, along with a set of 2D landmarks clicked in a *single* frame, our goal is to 3D reconstruct a metrically correct surface mesh, *i.e.*, infer the model's parameters z that best explains the underlying surface. We approach this problem by first applying Structure-from-Motion (SfM) to I, and then robustly fitting our model to the resulting (potentially noisy and) arbitrarily scaled sparse point cloud by solving Eq. (8).

In particular, our pipeline, as shown in Figure 4, is as follows. We begin by extracting a set I of m RGB images from the input video. To ensure that the selected frames are both sharp and evenly distributed across the video's timeline, we adopt a window-based selection strategy. Specifically, we first select m temporally equidistant candidate frames (regardless of image quality) and then search within a local neighborhood around each candidate for a frame that ranks within the top 25% of the overall sharpest frames. If no such frame is found, we adaptively relax the sharpness threshold over three iterations before falling back to simply selecting the sharpest frame within the window. This approach ensures good temporal coverage while maintaining consistent image sharpness within *and* across local windows.

Next, we use a state-of-the-art and off-the-shelf SfM pipeline, VGGSfM (Wang et al., 2024a), to reconstruct camera parameters and a sparse point cloud $S \subset \mathbb{R}^3$ from the given images, I.

After that, we align the point cloud to the model's mean shape using a set of K 2D landmarks annotated in just a *single* image. Let this set of landmarks be denoted as $\{(u_1, v_1), (u_2, v_2), \ldots, (u_K, v_K)\} \subset \mathbb{N}^2$, and let us further assume that these landmarks have been annotated in the i-th image of I. We obtain corresponding 3D landmark positions l_j by back-projecting (u_j, v_j) based on the estimated camera parameters. To do so, we cast rays $r_j(t) = o + td_j$ from the camera's origin $o \in \mathbb{R}^3$ through pixel (u_j, v_j) into the scene, selecting the point in S as back-projected landmark that first hits the point cloud. More formally, we choose the point in S that is closest to the camera and near the ray, setting

$$l_j = \operatorname*{arg\,min}_{x \in S} \left\{ t_x^j = (x - o) \cdot d_j : d(x, r_j) \leq \delta \right\} \tag{9}$$

for all $j \in \{1, 2, \ldots, K\}$, where t_x^j denotes the distance from x to the camera along the ray r_j (*i.e.*, the depth),

$$d(x, r_j) = \|(x - o) \times d_j\|_2 \tag{10}$$

is the distance from point x to ray r_j, and $\delta > 0$. Moreover,

$$o = -R_i^\top t_i, \quad d_j = \frac{\hat{d}_j}{\|\hat{d}_j\|_2}, \quad \hat{d}_j = R_i^\top K_i^{-1} \begin{bmatrix} u_j \\ v_j \\ 1 \end{bmatrix}. \quad (11)$$

Here, $K_i \in \mathbb{R}^{3\times3}$ denotes the estimated camera intrinsics of the i-th image, and $[R_i \mid t_i] \in \mathbb{R}^{3\times4}$ are the extrinsic parameters, formed by rotation $R_i \in \mathrm{SO}(3)$ and translation $t_i \in \mathbb{R}^3$. Finally, we globally align the estimated point cloud S to the model's mean shape by computing a similarity transformation between the back-projected 3D landmarks $\{l_1, l_2, \ldots, l_K\}$ and the corresponding model's average anchor points using the method in (Umeyama, 1991). Once aligned, we prune away points in S that are further away from the mean shape than a pre-selected threshold, $\tau > 0$:

$$S' = \{x \in S : \|x - c_M(x)\|_2 \leq \tau\}, \quad (12)$$

where $c_M(x) = \arg\min_{x' \in M}\{\|x - x'\|_2\}$ is the closest point to x in the mean shape, $M \subset \mathbb{R}^3$. This effectively removes unwanted points in the background.

As a last step, we fit our model to the aligned and pruned point cloud S' as detailed in Section 3.1.2 (without the additional scaling and translation step), resulting in a 3D surface mesh that lives in $[-1, 1]^3$. For a lot of applications, including anthropometry and breast volume estimation, however, a metrical (*i.e.*, real-world scale) reconstruction is required. To recover the correct scale, we propose to use either of the following two strategies:

1. Scale the reconstructed mesh based on a known landmark distance (we use the nipple-to-nipple distance) measured on the real subject. This yields an *exact metrical reconstruction*, but requires additional data collection.

2. Scale the reconstructed mesh with the inverse average scaling factor obtained from our real-world scale training data when normalized to $[-1, 1]^3$. This approach does *not* yield exact metrical reconstructions on a per-instance level—it merely results in a statistically *approximate metrical reconstruction*; at the same time, however, it avoids any additional data collection.

As we will show in our experimental evaluation, both strategies result in nearly the same reconstruction accuracy, indicating that approximate metrical reconstructions are of sufficient quality in practice.

3.3 Implementation Details

Model. We use $K = 6$ anchor points, distributed as shown in Figure 3. Furthermore, our model employs a global latent dimension of $L_{\text{glob}} = 128$, and a local latent dimension of $L_{\text{loc}} = 64$ (as such, our model has a total latent dimension of $L = 128 + (6+1)64 = 576$). The bandwidth h of the Gaussian kernel in Eq. (4) is set to 0.25, and a constant weight response of 0.2 is used for the background MLP. The remaining model and training parameters closely resemble those from (Giebenhain et al., 2023). In particular, local MLPs ϕ_k are fully connected and have four 200-dimensional hidden layers with a skip connection to the middle layer. We use the geometric initialization scheme proposed in (Gropp et al., 2020). The final SDF value is regressed by applying the softplus activation function. The MLP that predicts anchor positions has a single, 256-dimensional hidden layer with ReLU activation. Latent codes are initialized from a zero-mean normal distribution with variance 10^{-4}. We trained until convergence, but not longer than 15,000 epochs, using a batch size of 16 and the AdamW optimizer (Loshchilov and Hutter, 2019) with a weight decay of 0.01 and a learning rate of 5×10^{-4} for the model parameters and 10^{-3} for latent codes. Both learning rates are decayed by a factor of 0.5 every 3,000 epochs. We use 500 on-surface points and 500 off-surface points sampled from D (obtained using $\sigma^2 = 0.01$) for each shape. We empirically set $\lambda_1 = 2$, $\lambda_2 = 0.3$, $\lambda_3 = 0.1$, $\lambda_4 = 0.01$, $\lambda_5 = 7.5$, $\lambda_6 = 0.01$, and $\alpha = 10$. Gradients are clipped with a cut-off value of 0.1. Training our model took about 18 hours on a single NVIDIA A40 with 40 GB of VRAM.

At inference, we optimize Eq. (8) for 1,000 iterations using Adam (Kingma and Ba, 2015) with a learning rate of 10^{-2}. Additionally, we decay the learning rate by a factor of 0.5 every 200 iterations and divide λ by 3 after 200 iterations and by 10 after 600 iterations. Inference takes about 15 seconds, measured on a single NVIDIA RTX A4000 with 20 GB of VRAM and a resolution of 256^3.

Surface Reconstruction. We assume shared camera intrinsics across all images during SfM, which is safe to accept since our images are extracted from a video sequence. Moreover, we slightly adapted VGGSfM's default parameters (see discussion below). We selected six landmarks in a single image, corresponding to the model's anchor points placed at the anatomical locations and described below. Furthermore, δ is set such that the constraint in Eq. (9) considers the five closest points to a ray. Lastly, $\tau = 0.2$ in Eq. (12), effectively discarding points that are further away from the mean shape than ~10 cm.

4. Experiments and Results

We performed extensive experiments to validate our new model and the proposed 3D surface reconstruction pipeline.

Data. Our model is trained on the pre-processed dataset of the iRBSM (Weiherer et al., 2025), which includes 168 consistently oriented 3D breast scans. All scans have been taken in a standing position using Canfield's Vectra H2 system. Subsequently, after data acquisition, the following six anatomical landmarks have been expert-annotated in each 3D breast scan: (1) sternal notch, (2) belly button, (3) left nipple, (4) right nipple, (5) left coracoid process, and (6) right coracoid process, see also Figure 3.

To evaluate our model and the proposed surface reconstruction pipeline, we collected a test dataset of ten 3D breast scans acquired in a standing position using the Vectra H2 system, along with corresponding monocular RGB video sequences. The videos were captured using an iPhone 12 Mini while moving in a 180-degree circular arc around the subject at an approximately constant speed. No specific constraints were imposed during the data acquisition process, except that the region between the belly button and the sternal notch (both *included*) remained visible throughout the capture. On average, recording the videos took about 20 seconds per subject.

Baselines. We compare our model against the publicly available RBSM (Weiherer et al., 2023) and the recently proposed iRBSM (Weiherer et al., 2025). The mesh-based RBSM is fitted to point clouds X' using model-based non-rigid surface registration, optimizing

$$L(\alpha) = \|x(\alpha) - c_{x(\alpha)}(X')\|_2 + \lambda_1 \|x_J(\alpha) - l\|_2 + \lambda_2 \|\alpha\|_2 \tag{13}$$

for 2,000 iterations using Adam (Kingma and Ba, 2015), a learning rate of 10^{-2}, $\lambda_1 = 0.1$, and $\lambda_2 = 0.01$. The learning rate is decayed by a factor of 0.5 every 500 iterations. In Eq. (13), $x(\alpha) = \bar{x} + U\alpha$ denotes the classical PCA-based statistical shape model, where $\bar{x} \in \mathbb{R}^{3n}$ is the mean shape and $U \in \mathbb{R}^{3n \times q}$ holds the model's q principal components. Furthermore, $J \subset \mathbb{N}$ is an index set, selecting the m landmarks in the model that correspond to $l \in \mathbb{R}^{3m}$ on the target, X'. Due to the absence of the belly button and coracoid process in the RBSM, we use the following five landmarks to guide the registration process: sternal notch, left nipple, right nipple, left lower breast pole, and right lower breast pole; see, *e.g.*, (Hartmann et al., 2020) for a detailed explanation of these landmarks. We employ the augmented version of the RBSM that also includes mirrored scans, and use all $q = 219$ available principal components.

The iRBSM is fitted by solving Eq. (8) under the same setup used for our model, as detailed in Section 3.3.

Metrics. We report Chamfer distance (CD), F-Score with a threshold of 2.5 mm, and normal consistency (NC) to compare our surface reconstructions against ground truth 3D breast scans (we refer to Appendix A for an exact definition of these metrics). To ensure fair comparison, we compute metrics only within the breast region, defined by cropping meshes using axis-aligned planes passing through the upper breast poles, along axillary lines, and approximately 5 cm below the lower breast poles.

	CD ↓	F-Score ↑	NC ↑
RBSM	3.40 ± 1.20	52.0 ± 19.67	96.8 ± 1.73
iRBSM	1.13 ± 0.38	93.5 ± 5.55	99.0 ± 0.96
Ours	**0.77** ± 0.19	**98.6** ± 2.43	**99.6** ± 0.47

Table 3: **Intrinsic model evaluation**. We show quantitative results for surface reconstruction from clean point clouds consisting of 5,000 points. Our model outperforms the RBSM and iRBSM by a large margin.

	CD ↓	F-Score ↑	NC ↑
iRBSM $(- \mid -;\ 256)$	1.13 ± 0.38	93.5 ± 5.55	99.0 ± 0.96
$K = 4\ (96 \mid 32;\ 256)$	0.93 ± 0.24	96.6 ± 3.69	99.4 ± 0.59
$K = 6\ (32 \mid 32;\ 256)$	0.95 ± 0.28	96.3 ± 4.31	99.3 ± 0.73
$K = 6\ (64 \mid 32;\ 288)$	0.89 ± 0.24	97.5 ± 3.67	99.4 ± 0.70
$K = 6\ (96 \mid 32;\ 320)$	0.87 ± 0.21	97.7 ± 2.88	99.5 ± 0.54
Ours $K = 6\ (128 \mid 64;\ 576)$	**0.77** ± 0.19	**98.6** ± 2.43	**99.6** ± 0.47

Table 4: **Ablation on anchor layout and latent dimensions.** Numbers in parentheses represent global, local, and total latent dimension, *i.e.*, $(L_{\text{glob}} \mid L_{\text{loc}}; L)$. Proposed localized models consistently outperform the global iRBSM.

4.1 Intrinsic Model Evaluation

We begin by evaluating our model's ability to reconstruct *clean* point clouds, which are obtained through random sampling of 5,000 surface points from each of our test scans. This experiment is designed to measure the true representational capacity of our model in comparison to the RBSM and iRBSM, as surface reconstruction quality is assessed independently of any compromised (noisy) or incomplete data; hence, we refer to it as *intrinsic model evaluation*. We set $\lambda = 0.05$.

Results are summarized in Table 3. Our model consistently outperforms both baselines across all evaluation metrics. Compared to the mesh-based RBSM, it achieves over a four times reduction in CD, indicating a significantly more accurate reconstruction of the underlying surface geometry. Furthermore, our localized model surpasses the global iRBSM by reducing CD by more than 30%, demonstrating the effectiveness of incorporating spatial locality. We also investigated the impact of the number of input points during surface fitting and found that both the iRBSM and our model benefit from increased point density. Please see Appendix B for quantitative and qualitative results.

Finally, we study our model's shape space by inspecting random samples and linear interpolants in latent space. As seen from Figures 5 and 6, our model is highly expressive and its shape space is continuous and well-behaved, being able to generate a variety of plausible and realistic-looking breast shapes.

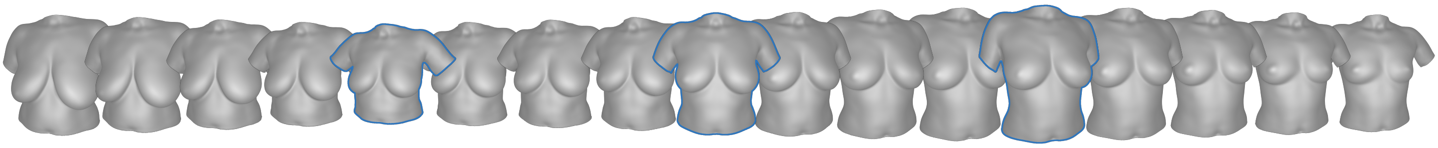

Figure 5: **Latent space interpolation.** We linearly interpolate between the latent codes corresponding to the shapes outlined in blue. Starting from the mean shape at the center, we interpolate in positive direction along the first and second principal components on the right, and in the negative direction on the left. Principal components are derived by applying PCA to the optimized latent codes from the training data. Our latent space is continuous and well-behaved.

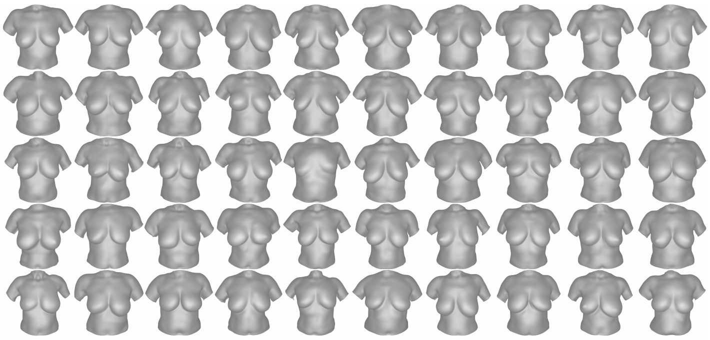

Figure 6: **Random samples from our model.** The proposed model is highly expressive and allows for a variety of plausible breast shapes.

4.2 Ablations

Next, we ablate key design choices, the number of anchor points, and the model's global and local latent dimensions.

Quantitative results can be found in Table 4, and we refer to Appendix C for qualitative comparisons. Interestingly, when the total latent dimensionality of our models is matched to that of the iRBSM ($L = 256$), the model with four anchors slightly outperforms the six-anchor setup (0.93 vs. 0.95 in CD). However, when both the global *and* local latent dimensionalities are matched across configurations ($L_{\text{loc}} = 96$, $L_{\text{glob}} = 32$) and the total latent dimension is disregarded, the six-anchor model significantly outperforms the four-anchor variant (0.87 vs. 0.93). In general, as the global latent dimensionality in the six-anchor configuration increases, the model begins to surpass the four-anchor variant, with reconstruction quality improving steadily with larger L_{glob}. Finally, all of our models consistently outperform the global iRBSM, regardless of anchor configuration or latent dimensionality.

	CD ↓	F-Score ↑	NC ↑
Non-metrical evaluation			
iRBSM	2.06 ± 1.01	78.1 ± 16.97	98.6 ± 1.46
Ours w\o $\mathcal{L}_{\text{anc}}$	1.74 ± 0.65	86.0 ± 11.17	99.2 ± 0.71
Ours	**1.73** ± 0.65	**86.1** ± 11.15	**99.2** ± 0.70
Metrical evaluation			
Exact metrical			
iRBSM	2.28 ± 0.65	70.1 ± 12.36	98.5 ± 1.34
Ours w\o $\mathcal{L}_{\text{anc}}$	1.97 ± 0.51	76.5 ± 10.86	99.0 ± 0.75
Ours	**1.96** ± 0.47	**76.7** ± 10.05	**99.0** ± 0.68
Approximate metrical			
iRBSM	2.30 ± 0.77	70.0 ± 13.60	98.5 ± 1.34
Ours w\o $\mathcal{L}_{\text{anc}}$	1.98 ± 0.47	76.2 ± 10.84	99.0 ± 0.75
Ours	**1.97** ± 0.47	**76.4** ± 10.00	**99.0** ± 0.68

Table 5: **Quantitative results for 3D surface reconstruction from monocular RGB videos.** The proposed pipeline in combination with our new model performs best.

4.3 Surface Reconstruction From Monocular RGB Videos

Finally, we evaluate our proposed 3D surface reconstruction pipeline. Since the reconstructed surfaces and ground truth 3D breast scans are not precisely aligned in this setting—unlike in the clean point cloud fitting scenario described in Section 4.1, where points are sampled directly from the ground truth surface—we follow common practice (Sanyal et al., 2019) and apply landmark-based Procrustes alignment to align the reconstructed surface to the ground truth 3D breast scan, followed by an iterative closest point (ICP) algorithm. Similar to (Zielonka et al., 2022), we evaluate performance under two settings, non-metrical and metrical. In the non-metrical setting, we disregard the fact that our method produces metrically correct 3D reconstructions and allow the Procrustes and ICP steps to estimate a global scaling factor in addition to rotation and translation. In the metrical setting, we respect the real-world scale of the method's output, restricting alignment to be purely rigid (only rotation and translation). Once aligned, evaluation metrics are computed as described above. Metrics obtained from the non-metrical evaluation measure the true discrepancy between reconstructed and ground truth geometry, whereas numbers obtained from the metrical evaluation better reflect the overall reconstruction quality when results are to be used within a real-world application that requires a metrical context (such as breast volume estimation, for example). We use $\lambda = 0.1$ during surface fitting and further employ the anchor loss in Eq. (7), weighted with 0.1, as

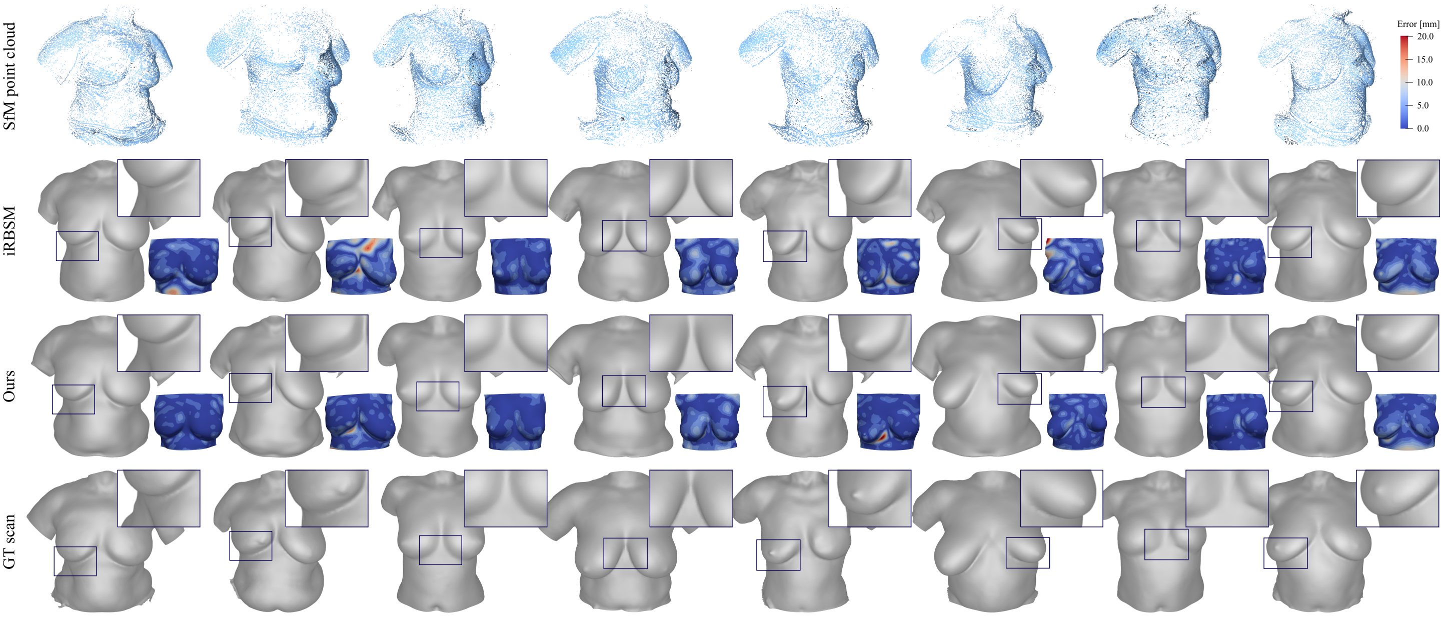

Figure 7: **Qualitative results for surface reconstruction from monocular RGB videos.** We show surface reconstructions obtained from fitting the iRBSM and our model to point clouds acquired by applying SfM to monocular RGB videos. In contrast to the iRBSM, our model is capable of recovering fine anatomical details such as skin folds (1st column) and nipples (5th and 8th columns), and is generally better at inferring the correct underlying shape, as seen in the remaining columns.

we found that this slightly improves our reconstructions.

Results are shown in Table 5, obtained by extracting 30 frames from the input video over time. The introduced pipeline in combination with the proposed model is able to recover accurate 3D breast surfaces, reaching a CD of less than 2 mm in the metrical setting compared to 2.3 mm when using the iRBSM for surface reconstruction. This improvement is also evident visually, as shown in Figure 7. Our model yields high-fidelity 3D surface reconstructions, successfully capturing even fine anatomical details such as skin folds and nipples. Moreover, the point clouds obtained via SfM appear remarkably accurate, despite the relatively low quality of the extracted video frames, which suffer from noise and suboptimal lighting during capture. We attribute this robustness to VGGSfM's learning-based approach, as we found that traditional methods such as COLMAP (Schönberger and Frahm, 2016) produce *significantly* noisier and sparser point clouds, even after moderate parameter tuning. Interestingly, when using our first strategy to obtain real-world scale reconstructions (which requires the measurement of an additional landmark distance on the real subject), CD decreases only by 0.01 mm, suggesting that approximate metrical reconstructions are sufficient in practice.

4.3.1 Ablations

Sensitivity to Landmark Selection. We test how sensitive our reconstruction pipeline is regarding the selection of the six 2D landmarks required for aligning the SfM-generated point cloud to our model's mean shape. To this end, we added varying amounts of noise to the ground truth landmark positions (simulated as random perturbations drawn from a discrete uniform distribution in range $\{-k, \ldots, -1, 0, 1, \ldots, k\}$ pixels, where $k \in \{5, 10, 20, 30\}$) and run our pipeline. We repeated this procedure ten times for each noise level and all subjects in our test dataset. The results in Figure 8 demonstrate that uncertainty in the landmark selection process of up to 10 pixels—corresponding to a spherical uncertainty region with ∼5 mm radius in real-world around each landmark—has a negligible impact on reconstruction quality, with CD rising only slightly from 1.97 mm (no noise) to 2.1 mm. Beyond this point, CD increases steadily, reaching about 2.3 mm at 30 pixels of noise, which corresponds to an uncertainty region with a radius of about 14 mm. The increasing standard deviation further suggests that individual reconstructions become increasingly susceptible to noise.

Number of Views vs. Reconstruction Quality and Runtime. We further evaluate how reconstruction accuracy and runtime of our pipeline is affected by the number of views extracted from the input video. Since our method recovers 3D surfaces by fitting onto point clouds generated by SfM, the number *and* quality of input views directly impact the final 3D reconstruction results. As shown in Figure 9, VGGSfM reliably estimates camera poses and produces sufficiently accurate point clouds even with as few as

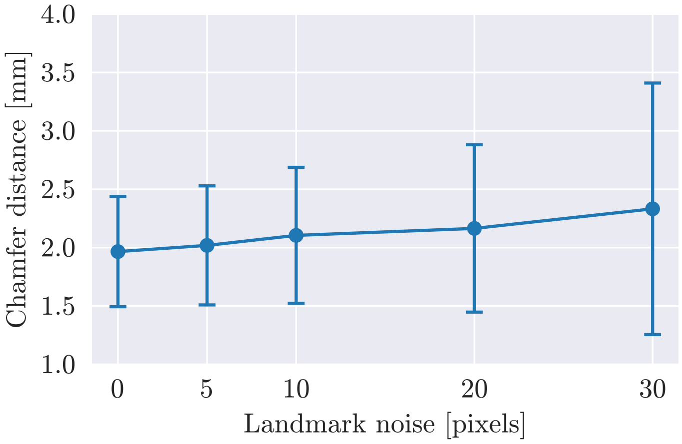

Figure 8: **Sensitivity to landmark selection.** We test how sensitive our reconstruction pipeline is against noise in the 2D landmark selection process. Our method is quite robust, leaving reconstruction quality nearly unchanged up to an uncertainty of 10 pixels.

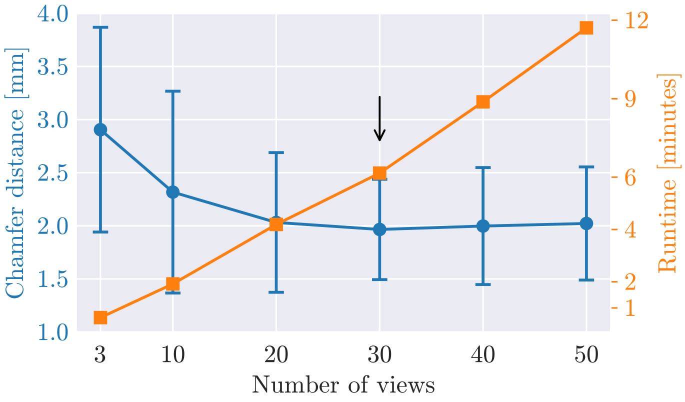

Figure 9: **Number of views vs. reconstruction quality and runtime.** We ablate the number of frames extracted from the input video and fed into our 3D reconstruction pipeline. We found that, in our setting, 30 images work best, being the ideal trade-off between speed and accuracy.

ten input images, achieving a CD of already 2.3 mm when evaluating the reconstructed surface. Adding more views consistently improves reconstruction quality, peaking at 30 images. Beyond this point, performance slightly declines, as the inclusion of additional (low-quality) frames into SfM *inevitably* introduces more noise, leading to stagnation or a mild decrease in reconstruction quality. We also tested our method in a few-shot setting, using only three input frames, matching the number of images required by the commercial Vectra H2 system. While the reconstruction quality is quantitatively reduced, the error remains below 3 mm, and the visual results are still compelling, see Appendix D. Finally, we observe that the runtime—measured only for the SfM step, as this is the primary bottleneck (the rest of the pipeline completes in under a minute)—increases linearly with the number of input images, remaining by about six minutes for 30 images. Please note that, while VGGSfM is generally reported to be faster (Wang et al., 2024a), our adapted parameter settings lead to slightly longer runtimes. We refer to Appendix E for a detailed discussion on how these parameters influence reconstruction quality and runtime.

5. Limitations and Future Work

While the introduced model and 3D reconstruction pipeline represent a significant step toward making 3D parametric breast models and reconstruction methods more expressive, accurate, and accessible, they still come with certain limitations and opportunities for further improvement.

First, due to the lack of automatic landmark detection algorithms specifically designed for the female breast, our surface reconstruction pipeline currently requires manual annotation of six anatomical landmarks in a single image. Although this process usually takes no more than five seconds, once such automatic landmarking methods become available, they can be seamlessly integrated into our pipeline, enabling fully automated surface reconstruction. Furthermore, such landmarking algorithms would not only make our surface reconstruction pipeline easier to use, but also more accurate, as back-projecting 2D landmarks from just one image is less precise than triangulating landmarks selected in *every* image. However, since no suitable automatic landmarking methods are currently available, manually annotating landmarks in every image would be very time-consuming and infeasible during daily clinical routine. Therefore, we opted for a practical trade-off between speed, ease of use, and accuracy by requiring manual landmark annotation only once in a single image. Moreover, it is important to note that our current approach of back-projecting landmarks via closest point searches yields accurate results only if the SfM-generated point cloud is sufficiently dense. We did not encounter issues in practice, as VGGSfM reliably reconstructs dense point clouds even from low-texture images, typically containing well over 5,000 points.

Second, our model *occasionally* struggles to precisely reconstruct the arms. This limitation arises from the fact that we did not model arms as articulated body parts, and the presented pose variability in the training data is simply not enough for the model to generalize across a wide range of arm positions. A potential solution would be to model the thorax as an articulated body as in (Palafox et al., 2021), for instance. However, this would require a dataset containing 3D breast scans in a canonical pose, along with additional scans of the same subjects captured in diverse arm positions. Such data that is currently unavailable and also not easy to acquire in routine clinical practice.

Finally, in future work, we plan to further evaluate the proposed surface reconstruction pipeline in a medical setting and a larger cohort, following common practice and assessing reconstruction quality in terms of breast volume and anthropometric measurements.

6. Conclusion

In this work, we have presented a state-of-the-art neural parametric 3D breast shape model, together with a low-cost and accessible surface reconstruction pipeline that accurately recovers metrically correct 3D breast geometry in minutes from a single monocular RGB video. Our approach requires neither specialized hardware nor proprietary software and can be used with any device capable of capturing RGB videos. At the core of our method lies an off-the-shelf differentiable SfM pipeline combined with robust model-based surface reconstruction. Unlike the existing iRBSM, which models breast geometry using a single *global* neural SDF, our model adopts a space partitioning strategy from state-of-the-art face models: the implicit breast domain is decomposed into multiple smaller regions, each represented by a *local* SDF anchored at anatomical landmark positions. When integrated into our reconstruction pipeline, the proposed model, dubbed *localized* iRBSM or liRBSM for short, achieves a significantly higher level of detail compared to the iRBSM, precisely recovering breast surfaces within a less than 2 mm error margin, and even capturing fine details such as skin folds and nipples. Both our model and the surface reconstruction pipeline are publicly available, and we hope this encourages clinicians and other researchers to further develop affordable, accessible, and open-source breast reconstruction methods. Openly sharing tools and data remains rare in this field, and we aim to help foster greater transparency and collaboration in breast shape modeling and reconstruction research.

Acknowledgments

We thank Marc Stamminger for providing valuable ideas and fruitful discussions. This work was partially funded by the German Federal Ministry of Education and Research (BMBF), FKZ: 01IS22082 (IRRW). The authors are responsible for the content of this publication. The authors gratefully acknowledge the scientific support and HPC resources provided by the Erlangen National High Performance Computing Center (NHR@FAU) of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) under the NHR project b112dc IRRW. NHR funding is provided by federal and Bavarian state authorities. NHR@FAU hardware is partially funded by the German Research Foundation (DFG) – 440719683.

Ethical Standards

The work follows appropriate ethical standards in conducting research and writing the manuscript, following all applicable laws and regulations regarding treatment of animals or human subjects.

Conflicts of Interest

We declare we don't have conflicts of interest.

Data availability

A core contribution of this publication is to share the proposed model, liRBSM, and surface reconstruction pipeline. As such, our trained model (in form of checkpoints containing the optimized weights), along with source code for training and inference as well as code for our 3D surface reconstruction method, is publicly available at `https://rbsm.re-mic.de/local-implicit`. Due to the sensitivity of the data and privacy constraints, we can not share the dataset used to train the model nor our test dataset.

References

Thiemo Alldieck, Hongyi Xu, and Cristian Sminchisescu. imGHUM: Implicit generative models of 3d human shape and articulated pose. In *ICCV*, 2021.

Lucy Bai, Ola Lundström, Hemming Johansson, Farid Meybodi, Brita Arver, Kerstin Sandelin, Marie Wickman, and Yvonne Brandberg. Clinical assessment of breast symmetry and aesthetic outcome: can 3d imaging be the gold standard? *J Plast Surg Hand Surg*, 57:145–152, 2023.

Annika S. Behrens, Hanna Huebner, Lothar Häberle, Marc Stamminger, Daniel Zint, Felix Heindl, Julius Emons, Carolin C. Hack, Naiba Nabieva, Michael Uder, Matthias Wetzl, Marius Wunderle, Matthias W. Beckmann, Peter A. Fasching, and Sabine Ohlmeyer. Comparative assessment of breast volume using a smartphone device versus mri. *Breast Cancer*, 32:166–176, 2024.

Vanessa Brébant, Maximilian Weiherer, Vivien Noisser, Stephan Seitz, Lukas Prantl, and Andreas Eigenberger. Implants versus lipograft: Analysis of long-term results following congenital breast asymmetry correction. *Aesthetic Plast Surg*, 46:2228–2236, 2022.

Xu Cao and Takafumi Taketomi. SuperNormal: Neural surface reconstruction via multi-view normal integration. In *CVPR*, 2024.

Pol Caselles, Eduard Ramon, Jaime Garcia, Gil Triginer, and Francesc Moreno-Noguer. Implicit shape and appearance priors for few-shot full head reconstruction. *IEEE Trans Pattern Anal Mach Intell*, 47:3691–3705, 2025.

Nikolas Chrobot, Michael Alfertshofer, Konstantin Frank, Wenko Smolka, Alexandra Anker, Juergen Taxis, Tobias Ettl, Lukas Prantl, Vanessa Brébant, and Robin Hartmann. Advancing digital anthropometry in plastic surgery: Comparing smartphone 3d surface imaging to vectra h2 in breast reconstruction. *JPRAS*, 104:398–406, 2025.

Pedro Costa, João P. Monteiro, Hooshiar Zolfagharnasab, and Hélder P. Oliveira. Tessellation-based coarse registration method for 3d reconstruction of the female torso. In *IEEE BIBM*, 2014.

Pablo de Heras Ciechomski, Mihai Constantinescu, Jaime Garcia, Radu Olariu, Irving Dindoyal, Serge Le Huu, and Mauricio Reyes. Development and implementation of a web-enabled 3d consultation tool for breast augmentation surgery based on 3d-image reconstruction of 2d pictures. *J Med Internet Res*, 14, 2012.

Boyang Deng, J.P. Lewis, Timothy Jeruzalski, Gerard Pons-Moll, Geoffrey Hinton, Mohammad Norouzi, and Andrea Tagliasacchi. NASA: Neural articulated shape approximation. In *ECCV*, 2020.

Benthe A. M. Dijkman, Niels P. T. J. Liberton, Sjoerd te Slaa, Jan Maerten Smit, Chantal M. Wiepjes, Koen M. A. Dreijerink, Martin den Heijer, Rudolf M. Verdaasdonk, and Christel J. M. de Blok. A comparative study of 3d measuring methods for monitoring breast volume changes. *PLoS One*, 19, 2024.

Bardienus Pieter Duisterhof, Lojze Zust, Philippe Weinzaepfel, Vincent Leroy, Yohann Cabon, and Jerome Revaud. MASt3R-SfM: A fully-integrated solution for unconstrained structure-from-motion. In *3DV*, 2025.

Maximilian Eder, v. Fee Waldenfels, Alexandra Swobodnik, Markus Klöppel, Ann-Kathrin Pape, Tibor Schuster, Stefan Raith, Elena Kitzler, Nikolaos A. Papadopulos, Hans-Günther Machens, and Laszlo Kovacs. Objective breast symmetry evaluation using 3-d surface imaging. *The Breast*, 21:152–158, 2012.

Bernhard Egger, William A.P. Smith, Ayush Tewari, Stefanie Wuhrer, Michael Zollhoefer, Thabo Beeler, Florian Bernard, Timo Bolkart, Adam Kortylewski, Sami Romdhani, Christian Theobalt, Volker Blanz, and Thomas Vetter. 3d morphable face models - past, present and future. *ACM TOG*, 39:1–38, 2020.

Elise Fu, Haoyin Zhou, Ángela Alarcón de la Lastra, Ruisi Zhang, Haripriya Ayyala, Justin Broyles, Bohdan Pomahac, and Jayender Jagadeesan. Slam-based breast reconstruction system for surgical guidance using a low-cost camera. *IEEE Trans Med Robot Bionics*, 6:1345–1353, 2024.

Cristina Gallego and Anne L. Martel. Automatic model-based 3d segmentation of the breast in mri. In *Proc. SPIE 7962, Medical Imaging 2011: Image Processing*, 2011.

Joachim Georgii, Maximilian Eder, Kai Bürger, Sebastian Klotz, Florian Ferstl, Laszlo Kovacs, and Rüdiger Westermann. A computational tool for preoperative breast augmentation planning in aesthetic plastic surgery. *IEEE J Biomed Health Inform*, 18:907–919, 2014.

Simon Giebenhain, Tobias Kirschstein, Markos Georgopoulos, Martin Rünz, Lourdes Agapito, and Matthias Nießner. Learning neural parametric head models. In *CVPR*, 2023.

Simon Giebenhain, Tobias Kirschstein, Markos Georgopoulos, Martin Rünz, Lourdes Agapito, and Matthias Nießner. MonoNPHM: Dynamic head reconstruction from monocular videos. In *CVPR*, 2024.

Pedro F. Gouveia, Hélder P. Oliveira, João P. Monteiro, João F. Teixeira, Nuno L. Silva, David Pinto, Carlos Mavioso, João Anacleto, Marta Martinho, Inês Duarte, Jaime S. Cardoso, Fátima Cardosoa, and Maria João Cardoso. 3d breast volume estimation. *Eur Surg Res*, 63:3–8, 2021.

Philip-William Grassal, Malte Prinzler, Titus Leistner, Carsten Rother, Matthias Nießner, and Justus Thies. Neural head avatars from monocular rgb videos. In *CVPR*, 2022.

Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. In *ICML*, 2020.

Michael W. Göpper, Jakob Neubauer, Ziad Kalash, G. Björn Stark, and Filip Simunovic. Improved accuracy of breast volume calculation from 3d surface imaging data using statistical shape models. *PLoS One*, 15, 2020.

Woo Yeon Han, Seok Joon Lee, Hsien Pin Chang, Jin Heo, Jin Sup Eom, Eun Key Kim, and Hyun Ho Han. Development of three-dimensional breast scan and measurement application using laser imaging detection and ranging sensor on iphone. *Plast Reconstr Surg*, 152, 2023.

Robin Hartmann, Maximilian Weiherer, Daniel Schiltz, Stephan Seitz, Luisa Lotter, Alexandra Anker, Christoph

Palm, Lukas Prantl, and Vanessa Brébant. A novel method of outcome assessment in breast reconstruction surgery: comparison of autologous and alloplastic techniques using three-dimensional surface imaging. *Aesthetic Plast Surg*, 44:1980–1987, 2020.

Robin Hartmann, Maximilian Weiherer, Daniel Schiltz, Magnus Baringer, Vivien Noisser, Vanessa Hösl, Andreas Eigenberger, Stephan Seitz, Christoph Palm, Lukas Prantl, and Vanessa Brébant. New aspects in digital breast assessment: further refinement of a method for automated digital anthropometry. *Arch Gynecol Obstet*, 303:721–728, 2021.

Helga Henseler, Xiangyang Ju, Ashraf Ayoub, and Arup K. Ray. The importance of the pose in three-dimensional imaging of the ptotic breast. *JPRAS*, 66:1551–1556, 2013.

Helga Henseler, Alina Kuznetsova, Peter Vogt, and Bodo Rosenhahn. Validation of the kinect device as a new portable imaging system for three-dimensional breast assessment. *JPRAS*, 67:483–488, 2014.

Helga Henseler, Sarah Kim Bonkat, Peter Maria Vogt, and Bodo Rosenhahn. The kinect recording system for objective three- and four-dimensional breast assessment with image overlays. *JPRAS*, 69:E27–E34, 2016.

Marilyn Keller, Keenon Werling, Soyong Shin, Scott Delp, Sergi Pujades, C. Karen Liu, and Michael J. Black. From skin to skeleton: Towards biomechanically accurate 3d digital humans. *ACM TOG*, 42:1–12, 2023.

Youngjun Kim, Kunwoo Lee, and Wontae Kim. 3d virtual simulator for breast plastic surgery. *Comp Anim Virtual Worlds*, 19:515–526, 2008.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015.

Konstantin C. Koban, Felix Härtnagl, Virginia Titze, Thilo L. Schenck, and Riccardo E. Giunta. Chances and limitations of a low-cost mobile 3d scanner for breast imaging in comparison to an established 3d photogrammetric system. *JPRAS*, 71:1417–1423, 2018.

Laszlo Kovacs, Maximilian Eder, Regina Hollweck, Alexander Zimmermann, Markus Settles, Armin Schneider, Matthias Endlich, Andreas Mueller, Katja Schwenzer-Zimmerer, Nikolaos A. Papadopulos, and Edgar Biemer. Comparison between breast volume measurement using 3d surface imaging and classical techniques. *The Breast*, 16:137–145, 2007.

Ioannis Kyriazidis, Juan Enrique Berner, Karl Waked, and Moustapha Hamdi. 3d breast scanning in plastic surgery utilizing free iphone lidar application: Evaluation, potential, and limitations. *Aesthet Surg J*, 45:NP99–NP104, 2025.

René M. Lacher, John H. Hipwell, Norman R. Williams, Mohammed R. S. Keshtgar, David J. Hawkes, and Danail Stoyanov. Low-cost surface reconstruction for aesthetic results assessment and prediction in breast cancer surgery. In *EMBC*, 2015.

René M. Lacher, Francisco Vasconcelos, David C. Bishop, Norman R. Williams, Mohammed Keshtgar, David J. Hawkes, John H. Hipwell, and Danail Stoyanov. A comparative study of breast surface reconstruction for aesthetic outcome assessment. In *MICCAI*, 2017.

René M. Lacher, Francisco Vasconcelos, Norman R. Williams, Gerrit Rindermann, John H. Hipwell, David J. Hawkes, and Danail Stoyanov. Nonrigid reconstruction of 3d breast surfaces with a low-cost rgbd camera for surgical planning and aesthetic evaluation. *Med Image Anal*, 53:11–25, 2019.

Woo Yeon Lee, Min Jung Kim, Dae Hyun Lew, Seung Yong Song, and Dong Won Lee. Three-dimensional surface imaging is an effective tool for measuring breast volume: A validation study. *Arch of Plast Surg*, 43:430–437, 2016.

Astrid Leusink, Rachel L. O'Connell, Stephanie L. Dean, Rosa Di Micco, Naser Alotaibi, Peter A. Barry, Anna M. Kirby, and Jennifer E. Rusby. A comparison of volume and anthropometric breast measurements using the crisalix and vectra xt 3-dimensional surface imaging systems in women who have undergone breast-conserving surgery. *Med Res Arch*, 9, 2021.

Zhaoshuo Li, Thomas Müller, Alex Evans, Russell H. Taylor, Mathias Unberath, Ming-Yu Liu, and Chen-Hsuan Lin. Neuralangelo: High-fidelity neural surface reconstruction. In *CVPR*, 2023.

Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM TOG*, 34:1–16, 2015.

William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. *SIGGRAPH*, 21:163–169, 1987.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *ICLR*, 2019.

Thi Hong Nhung Luu, Ngoc Tu Nguyen, Nhat Trinh Nguyen, Minh Hieu Pham, Thi Dieu Linh Nguyen, and Thi Le Nguyen. Development of 3d breast measurement system using structured light for bra design. In *ICISN*, 2021.

Arnaud Mazier, Sophie Ribes, Benjamin Gilles, and Stéphane Bordas. A rigged model of the breast for preoperative surgical planning. *J Biomech*, 128, 2021.

Marko Mihajlovic, Yan Zhang, Michael J. Black, and Siyu Tang. LEAP: Learning articulated occupancy of people. In *CVPR*, 2021.

Marko Mihajlovic, Shunsuke Saito, Aayush Bansal, Michael Zollhoefer, and Siyu Tang. COAP: Compositional articulated occupancy of people. In *CVPR*, 2022.

Marko Mihajlovic, Siwei Zhang, Gen Li, Kaifeng Zhao, Lea Müller, and Siyu Tang. VolumetricSMPL: A neural volumetric body model for efficient interactions, contacts, and collisions. In *ICCV*, 2025.

Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *CVPR*, 2020.

Vivien Noisser, Andreas Eigenberger, Maximilian Weiherer, Stephan Seitz, Lukas Prantl, and Vanessa Brébant. Surgery of congenital breast asymmetry—which objective parameter influences the subjective satisfaction with long-term results. *Arch Gynecol Obstet*, 305:95–102, 2022.

Michael Oechsle, Songyou Peng, and Andreas Geiger. UNISURF: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In *ICCV*, 2021.

Carlo M. Oranges, Srinivas Madduri, Philipp Brantner, Bilal Msallem, Salvatore Giordano, Benito Benitez, Daniel F. Kalbermatten, Dirk J. Schaefer, and Florian M. Thieringer. Three-dimensional assessment of the breast: Validation of a novel, simple and inexpensive scanning process. *In Vivo*, 33:839–842, 2019.

Ahmed A. A. Osman, Timo Bolkart, and Michael J. Black. STAR: Sparse trained articulated human body regressor. In *ECCV*, 2020.

Ahmed A. A. Osman, Timo Bolkart, Dimitrios Tzionas, and Michael J. Black. SUPR: A sparse unified part-based human representation. In *ECCV*, 2022.

Pablo Palafox, Aljaz Bozic, Justus Thies, Matthias Nießner, and Angela Dai. Neural parametric models for 3d deformable shapes. In *ICCV*, 2021.

Pablo Palafox, Nikos Sarafianos, Tony Tung, and Angela Dai. SPAMs: Structured implicit parametric models. In *CVPR*, 2022.

Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Lovegrove Steven. DeepSDF: Learning continuous signed distance functions for shape representation. In *CVPR*, 2019.

Marta Pinto, Joana Boita, Koen Michielsen, and Ioannis Sechopoulos. iphone truedepth cameras performance compared to optical 3d scanner for imaging the compressed breast shape. In *IWBI*, 2022.

Rolandos Alexandros Potamias, Stathis Galanakis, Jiankang Deng, Athanasios Papaioannou, and Stefanos Zafeiriou. ImHead: A large-scale implicit morphable model for localized head modeling. In *ICCV*, 2025.

Stefanie T. L. Pöhlmann, Jeremy Hewes, Andrew I. Williamson, Jamie C. Sergeant, Alan Hufton, Ashu Gandhi, Christopher J. Taylor, and Susan M. Astley. Breast volume measurement using a games console input device. In *IWBI*, 2014.

Stefanie T. L. Pöhlmann, Elaine Harkness, Christopher J. Taylor, Ashu Gandhi, and Susan M. Astley. Preoperative implant selection for unilateral breast reconstruction using 3d imaging with the microsoft kinect sensor. *JPRAS*, 70: 1059–1067, 2017.

Eduard Ramon, Gil Triginer, Janna Escur, Albert Pumarola, Jaime Garcia, Xavier Giro-i-Nieto, and Francesc Moreno-Noguer. H3D-Net: Few-shot high-fidelity 3d head reconstruction. In *ICCV*, 2021.

Kathleen M. Robinette, Sherri Blackwell, Hein Daanen, Mark Boehmer, Scott Fleming, Tina Brill, and Dennis Burnsides. Civilian american and european surface anthropometry resource (caesar) final report. Technical report, US Air Force Research Laboratory, 2002.

Hayeem L. Rudy, Yi-Hsueh Lu, Evan Rothchild, Oren M. Tepper, and Katie Weichman. Expanding access to 3d technology in plastic surgery of the breast: Validation of the iphone against the vectra h2. *Aesthet Surg J*, 44: 1350–1358, 2024.

Guillermo Ruiz, Eduard Ramon, Jaime Garcia, Federico M. Sukno, and Miguel A. González Ballester. Weighted regularized statistical shape space projection for breast 3d model reconstruction. *Med Image Anal*, 47:164–179, 2018.

Soubhik Sanyal, Timo Bolkart, Haiwen Feng, and Michael J. Black. Learning to regress 3d face shape and expression from an image without 3d supervision. In *CVPR*, 2019.

Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *CVPR*, 2016.

Hyewon Seo, Frederic Cordier, and Kyunghi Hong. A breast modeler based on analysis of breast scans. *Comp Anim Virtual Worlds*, 18:141–151, 2007.

Lama Seoud, Joyce Ramsay, Stefan Parent, and Farida Cheriet. A novel fully automatic measurement of apparent breast volume from trunk surface mesh. *Med Eng Phys*, 41:46–54, 2017.

Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *NeurIPS*, 2020.

Yating Tian, Hongwen Zhang, Yebin Liu, and Limin Wang. Recovering 3d human mesh from monocular images: A survey. *IEEE Trans Pattern Anal Mach Intell*, 45:15406–15425, 2023.

Olivia L. H. Tong, Astrid Chamson-Reig, Lawrence C. M. Yip, Muriel Brackstone, Mamadou Diop, and Jeffrey J. L. Carson. Structured-light surface scanning system to evaluate breast morphology in standing and supine positions. *Sci Rep*, 10, 2020.

Shinji Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Trans Pattern Anal Mach Intell*, 13:376–380, 1991.

Jianyuan Wang, Nikita Karaev, Christian Rupprecht, and David Novotny. VGGSfM: Visual geometry grounded deep structure from motion. In *CVPR*, 2024a.

Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David Novotny. VGGT: Visual geometry grounded transformer. In *CVPR*, 2025a.

Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In *NeurIPS*, 2021.

Shu Wang, Lei Ren, Jihua Liu, Cong Su, Shanshan He, Bo Ren, and Jian Yin. Three-dimensional breast imaging using artificial-intelligence-based automatic measurement system. *JPRAS Open*, 44:107–118, 2025b.

Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jerome Revaud. DUSt3R: Geometric 3d vision made easy. In *CVPR*, 2024b.

Yiming Wang, Qin Han, Marc Habermann, Kostas Daniilidis, Christian Theobalt, and Lingjie Liu. NeuS2: Fast learning of neural implicit surfaces for multi-view reconstruction. In *ICCV*, 2023.

Maximilian Weiherer, Andreas Eigenberger, Bernhard Egger, Vanessa Brébant, Lukas Prantl, and Christoph Palm. Learning the shape of female breasts: an open-access 3d statistical shape model of the female breast built from 110 breast scans. *Vis Comput*, 39:1597–1616, 2023.

Maximilian Weiherer, Antonia von Riedheim, Vanessa Brébant, Bernhard Egger, and Christoph Palm. iRBSM: A deep implicit 3d breast shape model. In *Bildverarbeitung für die Medizin*, 2025.

Jonathan S. Wheat, Simon Choppin, and Amit Goyal. Development and assessment of a microsoft kinect based system for imaging the breast in three dimensions. *Med Eng Phys*, 36:732–738, 2014.

Hongyi Xu, Gabriel Eduard Bazavan, Andrei Zanfir, William T. Freeman, Rahul Sukthankar, and Cristian Sminchisescu. GHUM & GHUML: Generative 3d human shape and articulated pose models. In *CVPR*, 2020.

Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. In *NeurIPS*, 2021.

Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, and Andreas Geiger. MonoSDF: Exploring monocular geometric cues for neural implicit surface reconstruction. In *NeurIPS*, 2022.

Mingwu Zheng, Hongyu Yang, Di Huang, and Liming Chen. ImFace: A nonlinear 3d morphable face model with implicit neural representations. In *CVPR*, 2022.

Wojciech Zielonka, Timo Bolkart, and Justus Thies. Towards metrical reconstruction of human faces. In *ECCV*, 2022.

Appendix A. Definition of Metrics

We compute metrics between the ground truth 3D breast scan and reconstruction by sampling 100k points and corresponding surface normals from the respective meshes, denoted as $X_{\text{gt}}, N_{\text{gt}} \subset \mathbb{R}^3$ and $X_{\text{rec}}, N_{\text{rec}} \subset \mathbb{R}^3$.

Chamfer Distance. The Chamfer distance (CD) is defined as

$$\frac{1}{2}(\text{Comp.} + \text{Acc.}), \tag{14}$$

where

$$\text{Comp.} = \frac{1}{|X_{\text{gt}}|} \sum_{x_{\text{gt}} \in X_{\text{gt}}} \min_{x_{\text{rec}} \in X_{\text{rec}}} \|x_{\text{gt}} - x_{\text{rec}}\|, \tag{15}$$

$$\text{Acc.} = \frac{1}{|X_{\text{rec}}|} \sum_{x_{\text{rec}} \in X_{\text{rec}}} \min_{x_{\text{gt}} \in X_{\text{gt}}} \|x_{\text{rec}} - x_{\text{gt}}\|. \tag{16}$$

F-Score. The F-Score is given by

$$\frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{17}$$

where

$$\text{Precision} = \frac{|\{x_{\text{rec}} \in X_{\text{rec}} : \min_{x_{\text{gt}} \in X_{\text{gt}}} \|x_{\text{gt}} - x_{\text{rec}}\| < \xi\}|}{|X_{\text{rec}}|}, \tag{18}$$

$$\text{Recall} = \frac{|\{x_{\text{gt}} \in X_{\text{gt}} : \min_{x_{\text{rec}} \in X_{\text{rec}}} \|x_{\text{rec}} - x_{\text{gt}}\| < \xi\}|}{|X_{\text{gt}}|}. \tag{19}$$

We use $\xi = 2.5$ mm for all of our experiments.

Normal Consistency. The normal consistency (NC) is computed as

$$\frac{1}{2}\left(\sum_{x_{\text{gt}} \in X_{\text{gt}}} |\langle n_{\text{gt}}, n_{\text{rec}}(x_{\text{gt}})\rangle| + \sum_{x_{\text{rec}} \in X_{\text{rec}}} |\langle n_{\text{rec}}, n_{\text{gt}}(x_{\text{rec}})\rangle|\right), \tag{20}$$

where $n_{\text{rec}}(x_{\text{gt}})$ denotes the surface normal at the closest point of x_{gt} in X_{rec}. Similarly, $n_{\text{gt}}(x_{\text{rec}})$ denotes the surface normal at the closest point of x_{rec} in X_{gt}.

Appendix B. Additional Results for Intrinsic Model Evaluation

Here, we provide results for the experiment described in Section 4.1 of the main paper, which investigates the influence of the number of input points during surface fitting on clean point clouds.

Figures 10 and 11 show quantitative and qualitative results, respectively.

Appendix C. Qualitative Results for Ablations

We provide accompanying qualitative results in Figure 12 corresponding to the ablation study presented in Section 4.2 of the main paper.

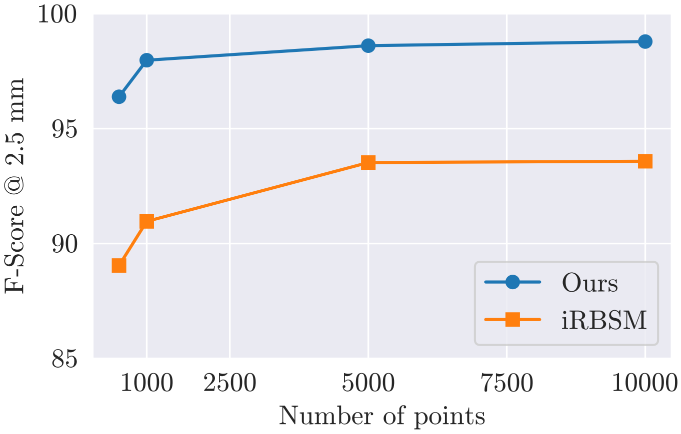

Figure 10: **Surface fitting under varying point densities.** We investigate the influence of the number of input points used during surface fitting. Our model outperforms the iRBSM across all input densities, with reconstruction quality improving as more points are provided. Higher is better.

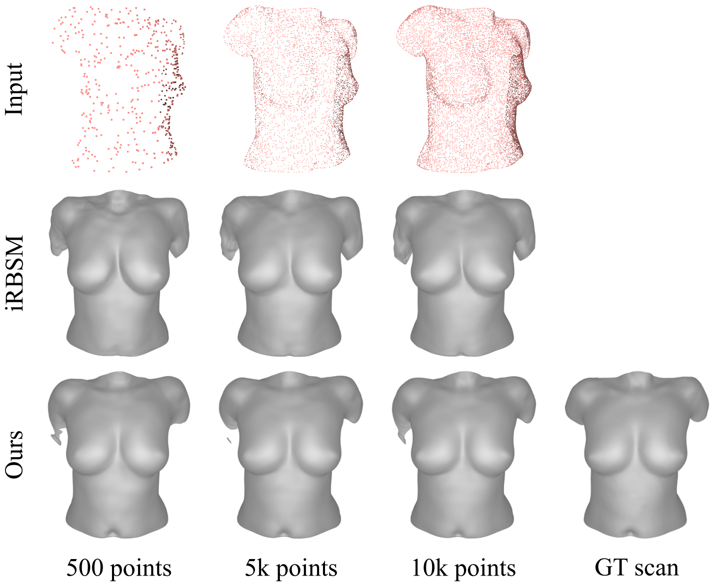

Figure 11: **Qualitative results for surface fitting under varying point densities.** Comparison of surface reconstructions obtained with the iRBSM and our model when fitted to point clouds of varying densities.

Appendix D. Qualitative Results for Varying Number of Views

We report qualitative results for the ablation study on the number of input views to our 3D surface reconstruction pipeline (see Section 4.3.1 of the main paper) in Figure 13.

Appendix E. Discussion On VGGSfM Parameters

As briefly mentioned in Section 4.3.1 of the main paper, VGGSfM's runtime is actually reported to be faster (Wang et al., 2024a) compared to what we state in our paper,

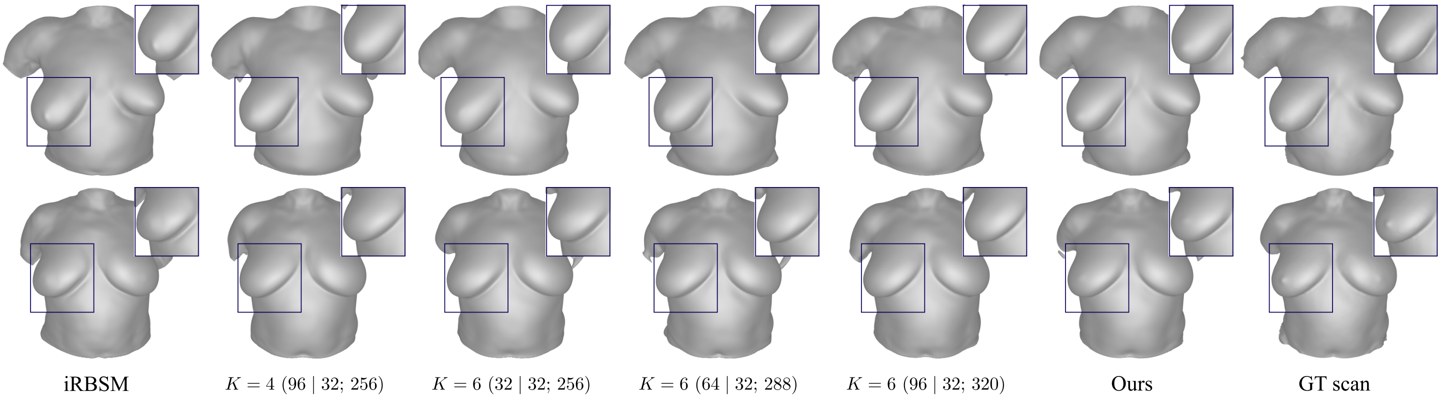

Figure 12: **Qualitative results for the ablation on anchor layout and latent dimensions.** Again, numbers in parentheses represent global, local, and total latent dimension, *i.e.*, $(L_{\text{glob}} \mid L_{\text{loc}}; L)$.

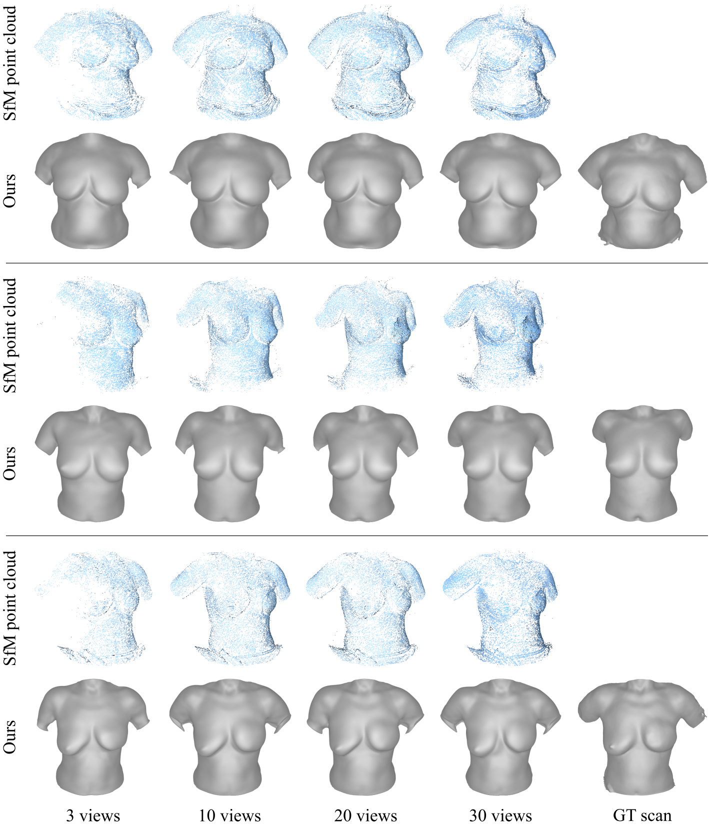

Figure 13: **Surface reconstruction from monocular RGB videos under varying number of views.** We show surface reconstructions obtained by varying the number of extracted input views to our pipeline.

which is due to an adapted parameter. VGGSfM essentially has two parameters that mainly influence the reconstruction result: `query_frame_num` and `max_query_pts` (per default set to 3 and 2,048). We found that increasing `max_query_pts` to 8,196 significantly improves surface reconstruction quality. In Table 6, we provide a small ablation on the two parameters to justify our adaptation.

Generally, we observe that both parameters lead to improved surface reconstructions but longer runtimes when increased up to a certain point. Specifically, although a `query_frame_num` of 7 and setting `max_query_pts` to 8,192 leads to the overall best result, due to the long runtime (almost 15 minutes), we decided against using this configuration and chose instead the best trade-off between speed and accuracy, which is the default `query_frame_num` of 3 and `max_query_pts` of 8,192.

	CD ↓	Runtime ↓
3 | 2,048	3.06	1.56
3 | 4,096	2.08	3.05
3 | 8,192	1.97	6.11
3 | 16,384	2.39	11.70
5 | 4,096	2.07	4.96
5 | 8,192	1.97	10.24
7 | 4,096	2.04	6.95
7 | 8,192	1.93	14.47

Table 6: **Ablation on VGGSfM parameters.** Numbers represent `query_frame_num` and `max_query_pts`. Runtime given in minutes. We underline the configuration we have used for our experiments.