Supplementary Materials: SALVE: A 3D Reconstruction Benchmark of Wounds from Consumer-grade Videos

Remi Chierchia[1,2], Leo Lebrat[1,2], David Ahmedt-Aristizabal[1,2], Olivier Salvado[1],
Clinton Fookes[1], Rodrigo Santa Cruz[1,2]
Queensland University of Technology[1], CSIRO Data61[2]

In this supplementary material, we present further details on the acquisition setup and additional qualitative results, including error color-coded meshes and novel-view renderings. Additionally, we present the non-suitability of the surfaces from DUSt3R [2], Instant-NGP [1], and NeuS2 [3] for geometric assessment.

In Section A we present more details relative to the **SALVE**'s dataset challenges. Next, in Section B we describe more precisely the evaluation protocol, namely a qualitative evaluation of the ground-truth point clouds and specifications about sampling of the meshes adopted for the experiments of Section 3.3 of the main paper. Following, with respect to Figure 4 and Figure 6 of the main paper, we present the additional error color-coded meshes (Figure 3) and renderings (Figure 4 for PIS3 and Figure 5 for PIS4) for the remaining devices and wound types we omitted in the main paper.

In addition, we present the surfaces extracted from DUSt3R, Instant-NGP, and NeuS2 in Figures 6, 7, and 8 respectively, and justify why those methods are not included in 3D reconstruction benchmark.

Finally, in Figure 9 we present some cases of failed reconstruction for the photogrammetric approaches when we input a greater amount of images, specifically 100 and 150. Compared to a set of 50 images, 100 and 150 present more blurring artefacts as mentioned in Section 3.1 of the main paper. As a result, we chose sets of 50 images for our **SALVE** dataset as they provided more consistent results and better wound representations when compared to fewer images.

A. Dataset additional challenges

As depicted in Figure 3 of the main paper, we use a Logitech 4K webcam and an iPhone 14 Pro Max for video recordings, and a Revopoint POP 3D scanner to acquire the ground-truth point clouds. Both recording devices utilise default acquisition settings (e.g. exposure time, frame rate, aperture, white balance, etc.) to simulate acquisition scenarios as close as possible to a telehealth application, Figure 1 shows an example of dynamic lighting conditions comprised in our dataset. Furthermore, when extracting frames from the recordings, we select the sharpest frames as outlined in Section 3.1 of the main paper, however, we might still observe blurred images in the datasets. An example regarding the Logitech recordings is partially observable from the image in the second row of Figure 1. Additionally, **SALVE** presents moving shadows cast by the operator during the acquisition, which is a common issue in multiple illumination sources.

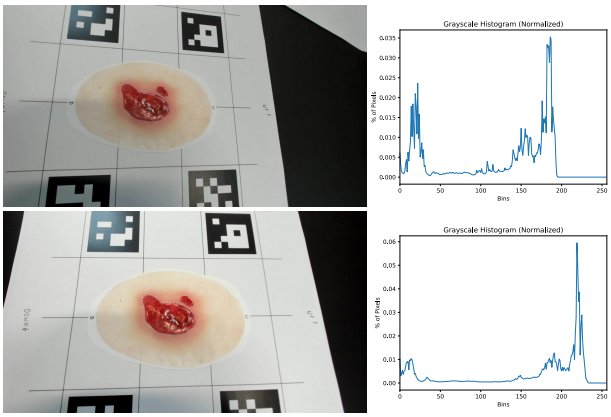

Figure 1. Two images of the PIS3 wound type captured from slightly different angles using the Logitech camera show different luminosity levels, as indicated by their normalized grayscale histograms (on the right).

B. Evaluation details

In Figure 2, we illustrate qualitatively the three ground-truth point clouds acquired with the Revopoint POP 3D scanner.

For the metrics AD, HD, HD_{90}, and NC we uniformly sampled 2 million points from the reconstructed meshes, while for $\mathcal{W}_2$ and $\mathcal{W}_2$-NC we sampled around 30 thousand points from the reconstructed meshes due to the computational complexity of optimal transport metrics.

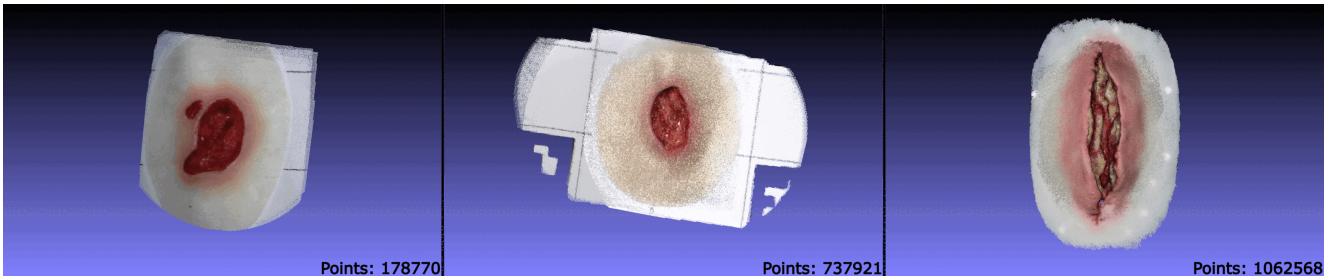

Figure 2. Ground-truth point clouds acquired with the Revopoint POP 3D scanner and their respective total number of points. From left to right: PIS3, PIS4, and SD wound types.

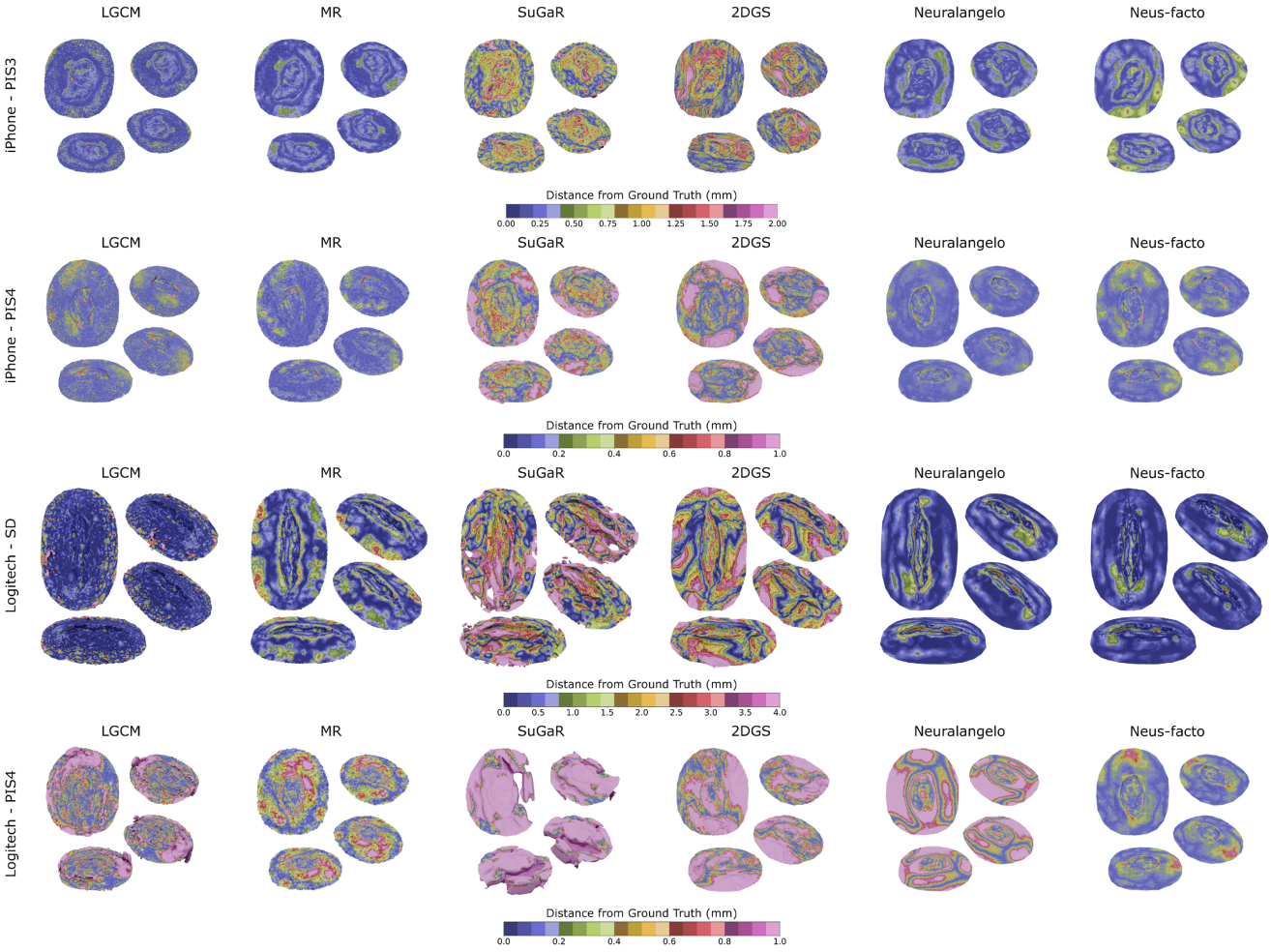

Figure 3. Reconstructed surface results for different methods, wound types and recording devices. The color represents the distance to the closest point in the ground-truth point cloud.

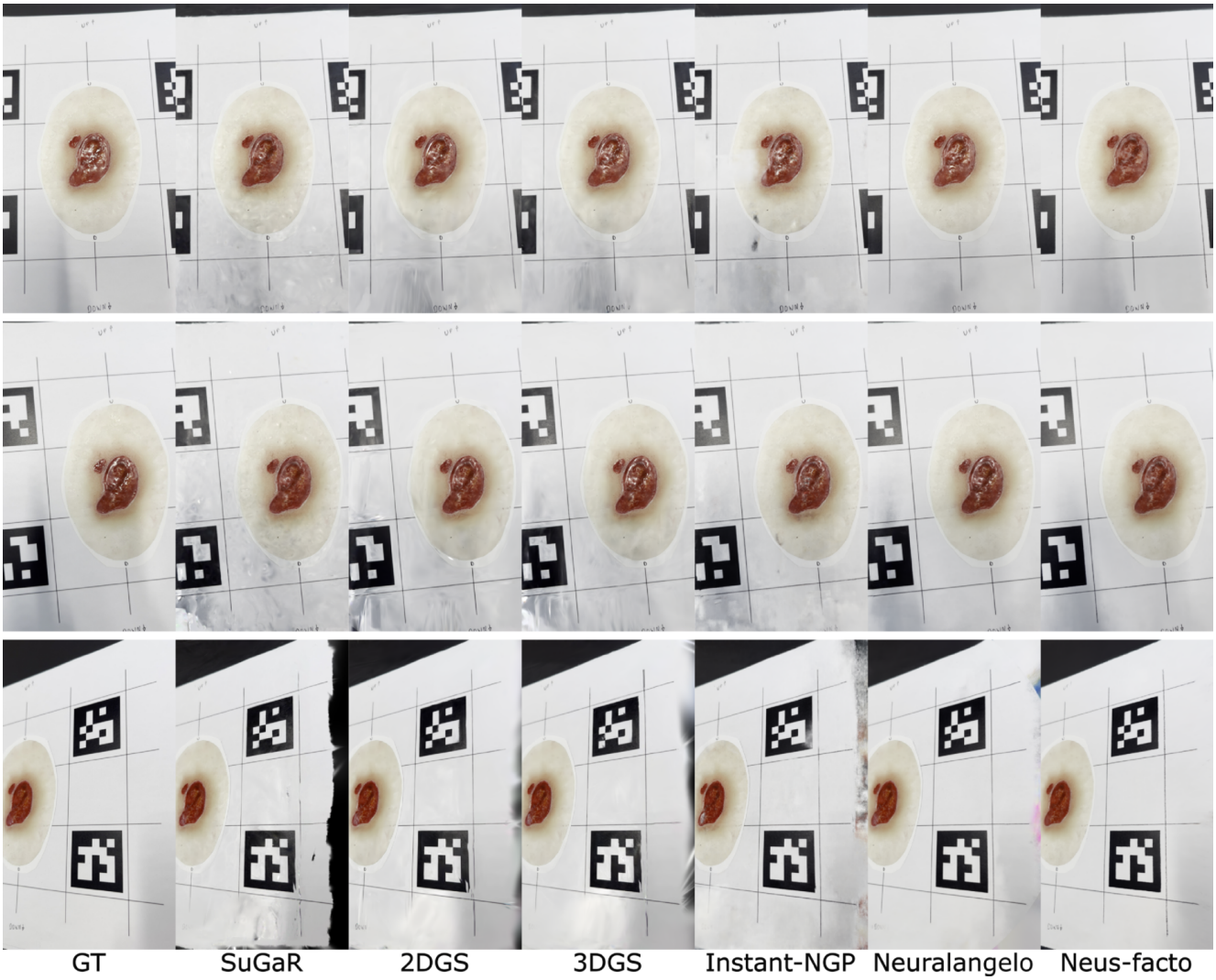

Figure 4. Qualitative evaluation of rendering methods for the PIS3 wound model. The first row presents the renderings for a view perspective in the test set that is well-represented in the training set. The second row shows a rendering of the wound from an oblique view, where we can observe the presence of floaters in non-SDF-based methods. The last row displays a view perspective under-observed in the training set. Notably, the PIS3 wound type presents hard reflections that appear especially in the first two rows.

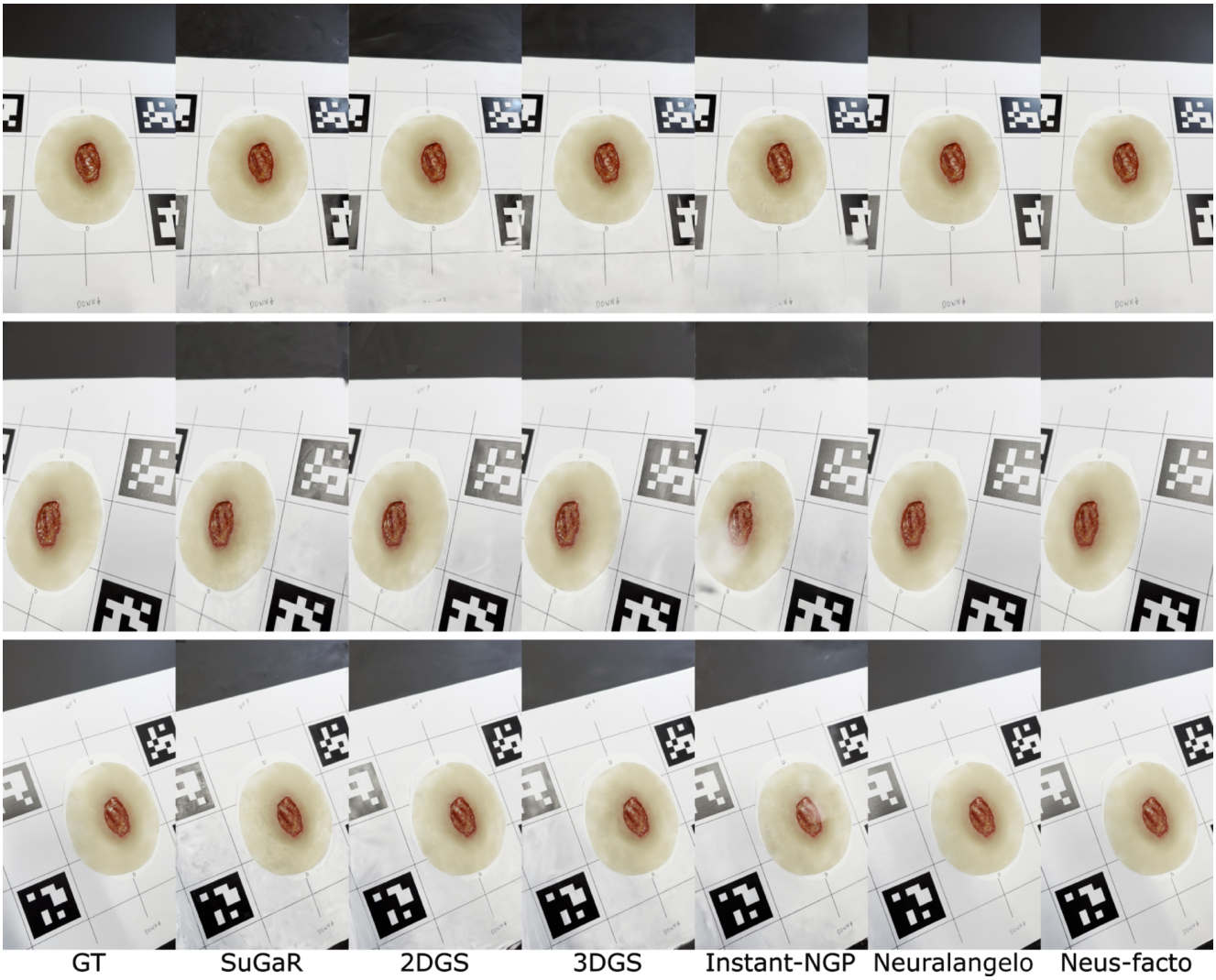

Figure 5. Qualitative evaluation of rendering methods for the PIS4 wound model. The first row presents the renderings for a view perspective in the test set that is well-represented in training set, however, floaters are already visible for non-SDF-based methods. The second row shows a rendering of the wound from an oblique view, where more floaters are present. The last row displays an image from another view angle where artifacts similar to those in the second row can be observed.

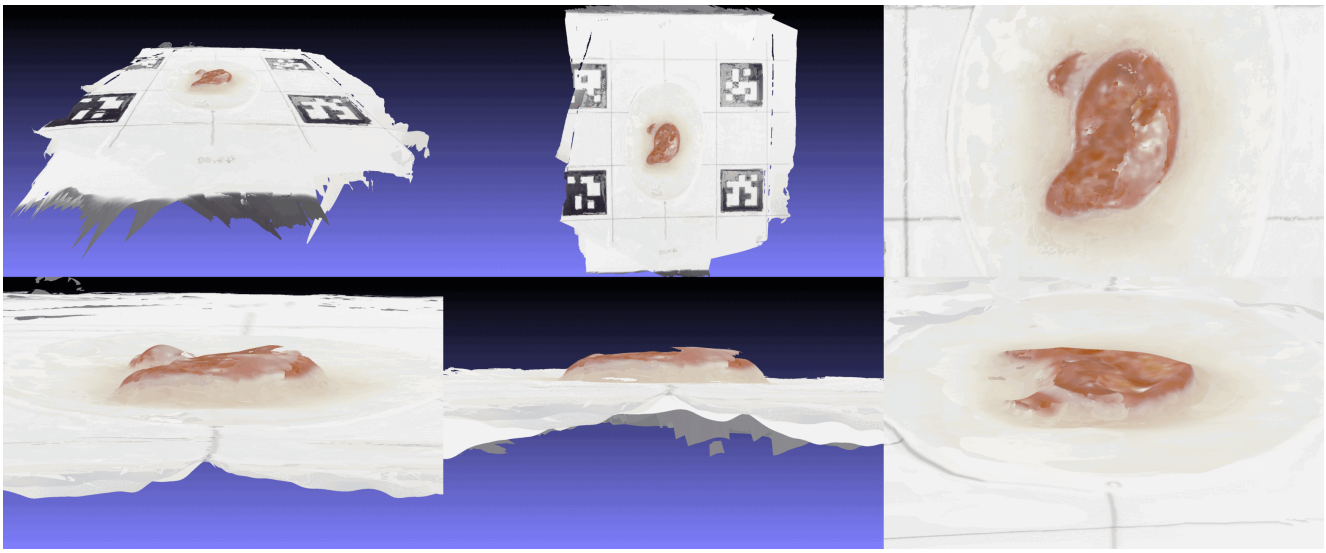

Figure 6. Six views of the mesh generated by DUSt3R for the PIS3 wound type. From the middle figure of the second row, we can observe how DUSt3R does not retrieve accurate geometry. The perimeter of the wound is much higher than the surrounding regions, not reflecting the real geometry of the scene. DUSt3R, unlike traditional multi-view stereo approaches, do not follow epipolar constraints to generate the 3D structure but relies on transformers tailored to solve reconstructions from a few views in the wild.

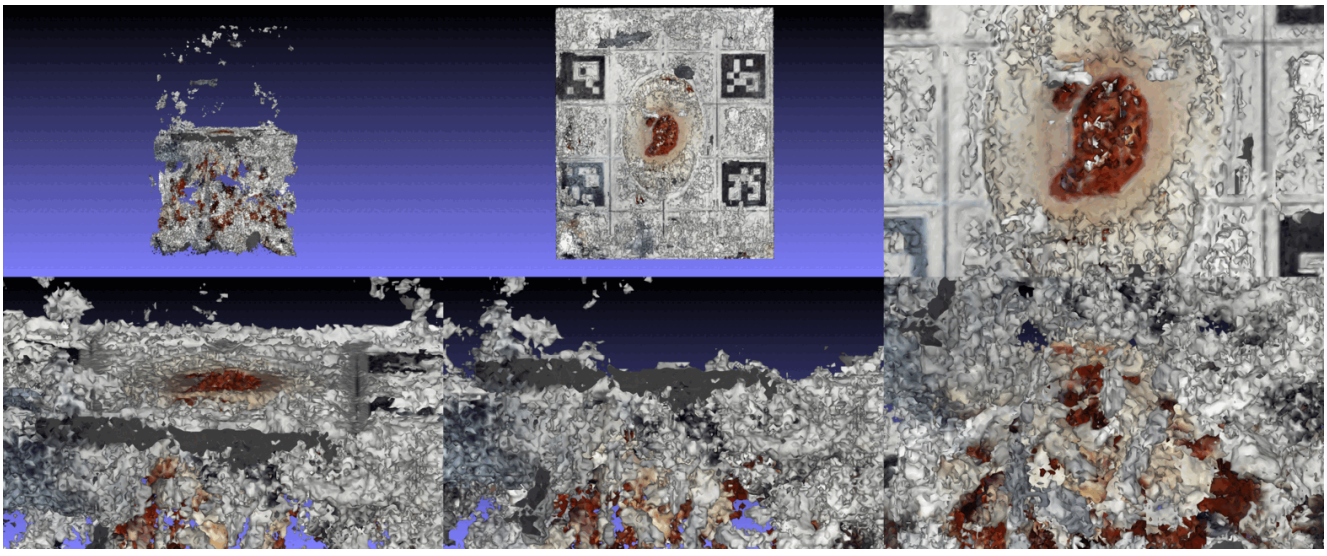

Figure 7. Six views of the mesh generated by Instant-NGP for the PIS3 wound type. As Instant-NGP is not a method developed for 3D surface reconstruction, its density is not regularised. As a result, when extracting a mesh using Marching Cubes algorithms, it presents a common structure to the one reported in the figure. Given the absence of surface regularizer terms, Instant-NGP solves photogrammetric consistency by placing density values under the wound's real surface, creating a scattered reconstruction.

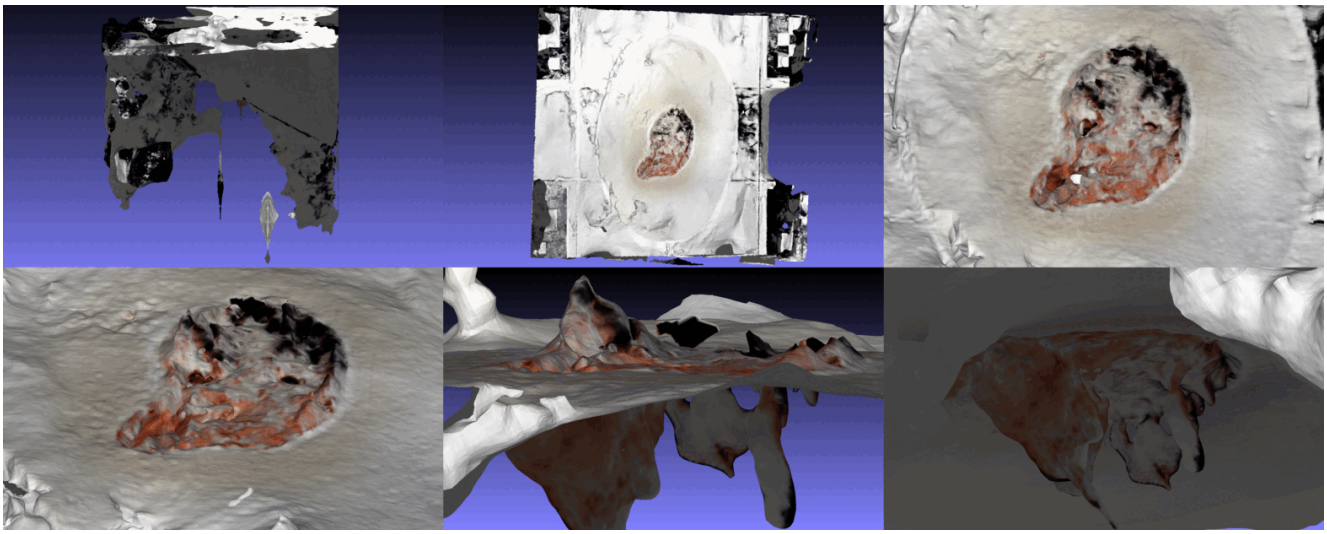

Figure 8. Six views of the mesh generated by NeuS2 for the PIS3 wound type. NeuS2 is a fast method for 3D surface reconstruction, however, we excluded it from our benchmark because it did not perform consistently well in our **SALVE** dataset. Although more fine-tuning might be required for **SALVE**, we considered Neuralangelo and Neus-facto as better representing SDF methods in terms of robustness against photogrammetric complexities. For example, PIS3 presents hard reflections, as analysed in the renderings evaluation above, and we can observe in both figures on the right how NeuS2 attempts to satisfy photometric consistency and surface regularization by "pushing" the surface under the wound level. This behaviour can be attributed to its architecture similar to Instant-NGP.

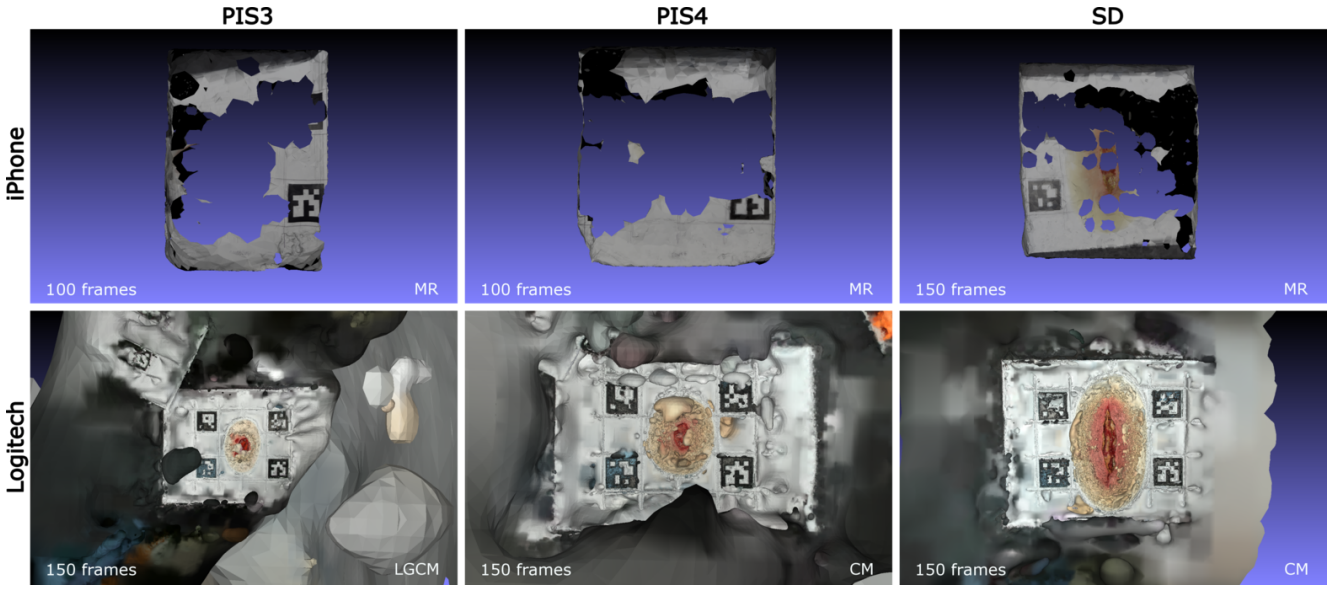

Figure 9. Examples of failed reconstruction while exploring how larger samples of images impact photogrammetric methods. In the first row, we display examples of 100 and 150 frames sampled from the iPhone sequence of each wound type. Meshroom (MR) was not able to reconstruct any detail in the wound and surrounding regions. In the second row, we display both COLMAP (CM) and COLMAP equipped with LightGlue feature matching (LGCM) at 150 frames. While both methods manage to reconstruct the wound area, the quality of the estimation drastically decreases compared to samples of 50 images.

References

[1] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, 41(4):1–15, 2022. 1

[2] Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jerome Revaud. Dust3r: Geometric 3d vision made easy, 2023. 1

[3] 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 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3295–3306, 2023. 1