A Novel Depth from Defocus Framework Based on a Thick Lens Camera Model

SUPPLEMENTARY MATERIAL

Matthew Bailey Jean-Yves Guillemaut
Centre for Vision, Speech and Signal Processing
University of Surrey, UK
{m.j.bailey, j.guillemaut}@surrey.ac.uk

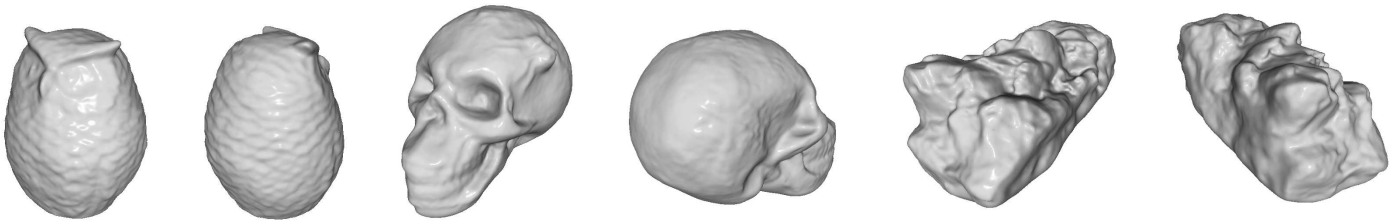

Figure 1: Our framework generates complete 3D models of macro-scale scenes using only defocus cues. To do so, we adopt a thick lens model instead of the traditional thin lens model to better estimate the blurring function of the camera.

1 Datasets

We present 3 Depth from Defocus (DFD) multi-view datasets: Owl, Skull and Quartz. These were captured using a Canon EOS 5D camera with a 100mm macro lens, and downsampled to a resolution of 2184 x 1464 pixels. Each dataset is summarised in Table 1.

Dataset	Views	Focus Settings	Aperture	Focal Length (mm)	Resolution (px)	Object Size (cm)
Owl	29	5	f/5.6	100	2184 x 1464	2.5 x 3.5 x 2.5
Skull	16	7	f/7.1	100	2184 x 1464	3.0 x 4.5 x 5.5
Quartz	7	7	f/7.1	100	2184 x 1464	3.0 x 3.0 x 8.0

Table 1: Overview of multi-view DFD datasets

Using our novel calibration, we successfully model the camera as a thick lens. Tables 2 and 3 show the keys parameters derived, with Skull and Quartz sharing the same calibration.

Setting	f (mm)	a (mm)	v (mm)	d (mm)	s	t (px)
1	98.13	8.76	143.43	310.68	1.0000	[0.00, 0.00]
2	97.73	8.73	141.50	315.90	1.0096	[-10.47, -6.97]
3	97.67	8.72	140.43	320.78	1.0193	[-21.01, 14.04]
4	97.42	8.70	138.93	326.08	1.0290	[-31.40, -21.01]
5	97.16	8.68	137.43	331.57	1.0385	[-41.54, -28.23]

Table 2: Defocus and registration parameters for the Owl dataset.

Here, f is focal length, a is aperture radius, v is image distance, d is focusing distance, s is scale and t is translation. All parameters are derived experiementally - see paper for further details. Table 4 shows the offsets independently optimised for each

Setting	f (mm)	a (mm)	v (mm)	d (mm)	s	t (px)
1	99.16	6.98	138.59	348.55	1.0000	[0.00, 0.00]
2	98.83	6.96	136.76	356.31	1.0108	[-11.85, -7.82]
3	99.07	6.98	136.24	363.18	1.0218	[-23.83, -15.89]
4	98.41	6.93	133.60	373.60	1.0326	[-35.62, -23.73]
5	98.54	6.94	132.85	381.54	1.0436	[-47.57, -31.66]
6	98.26	6.92	131.16	391.72	1.0545	[-59.50, -39.72]
7	98.20	6.92	129.95	401.83	1.0655	[-71.22, -48.31]

Table 3: Defocus and registration parameters for the Quartz and Skull datasets.

view. This is necessary to account for variations in the pupil ratio. Figures 2, 3 and 4 show example focal stacks from each of the datasets, which serve as input to our reconstruction framework. These images have lens distortion removed, as well as differences in scale and translation introduced by refocusing through the scene.

View	Owl (mm)	Skull (mm)	Quartz (mm)
	53.62	46.40	46.40
1	53.90	44.03	43.84
2	53.31	45.53	43.65
3	53.15	45.25	43.06
4	53.70	43.77	43.03
5	54.19	42.68	44.41
6	53.87	42.40	44.62
7	55.02	44.20	44.96
8	55.27	44.65	N/A
9	55.35	45.52	N/A
10	55.22	43.90	N/A
11	55.28	43.70	N/A
12	54.68	43.75	N/A
13	54.60	43.83	N/A
14	53.73	43.42	N/A
15	53.46	43.23	N/A
16	52.95	44.03	N/A
17	53.49	N/A	N/A
18	53.35	N/A	N/A
19	53.15	N/A	N/A
20	53.55	N/A	N/A
21	53.83	N/A	N/A
22	54.10	N/A	N/A
23	53.83	N/A	N/A
24	54.18	N/A	N/A
25	55.01	N/A	N/A
26	55.57	N/A	N/A
27	55.66	N/A	N/A
28	55.51	N/A	N/A
29	55.93	N/A	N/A

Table 4: Per-view pinhole offsets w derived for each view.

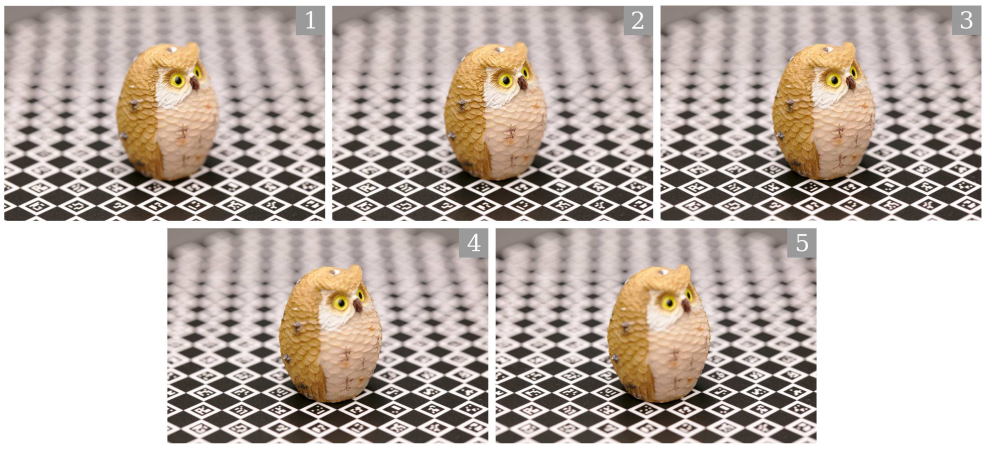

Figure 2: Focal stack for View 2 of the Owl dataset

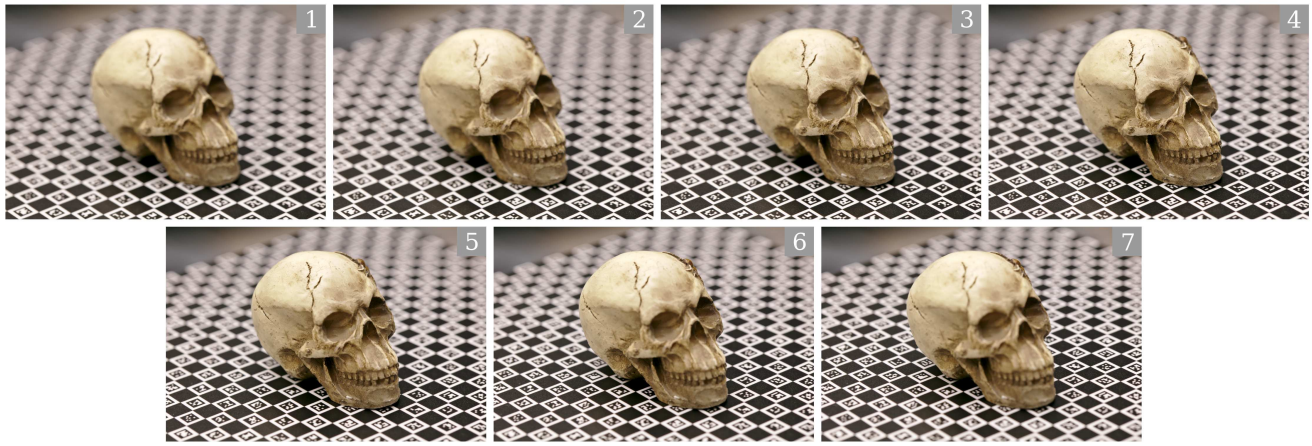

Figure 3: Focal stack for View 2 of the Skull dataset

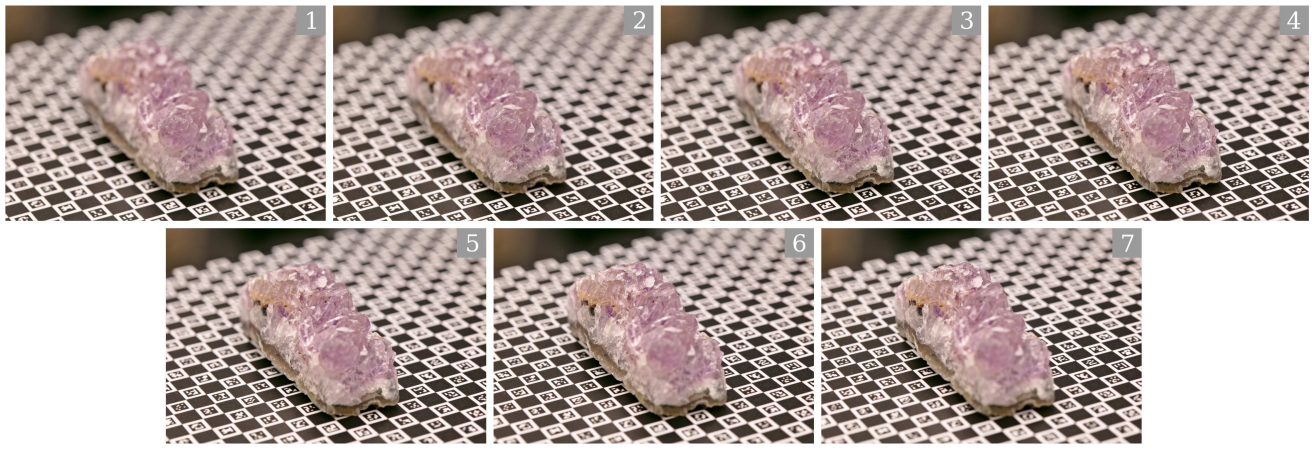

Figure 4: Focal stack for View 5 of the Quartz dataset

2 Image Registration

As described in the main paper, our image registration applies an affine transform that aligns all images in a given focal stack to a reference focus setting. Here, we evaluate this approach using the world reference calibration board. Specifically, features are identified in pseudo-pinhole images of the focal stack, and the distance between common features across different focus settings calculated. This is performed both before and after registration, as shown in Table 5. Note that the features used here are not part of the initial calibration, but could be used to further improve registration accuracy. It is clear that, even without this refinement, our registration is very effective - achieving subpixel alignment in all datasets. The effect this process has is illustrated in Figure 5.

Dataset		Registration Error (px) Min	Max	RMS
Owl	Before	4.474	39.95	16.92
	After	**0.0162**	**0.7602**	**0.2862**
Skull	Before	4.609	67.07	22.51
	After	**0.0079**	**0.6800**	**0.2833**
Quartz	Before	4.869	62.27	22.99
	After	**0.0075**	**0.7193**	**0.2889**

Table 5: Evaluation of image registration on one focal stack for all datasets. Here, we present the minimum, maximum and root mean square (RMS) pixel errors between features common across focus settings before and after registration. Our approach achieves subpixel registration, reducing the initial alignment error by several orders of magnitude.

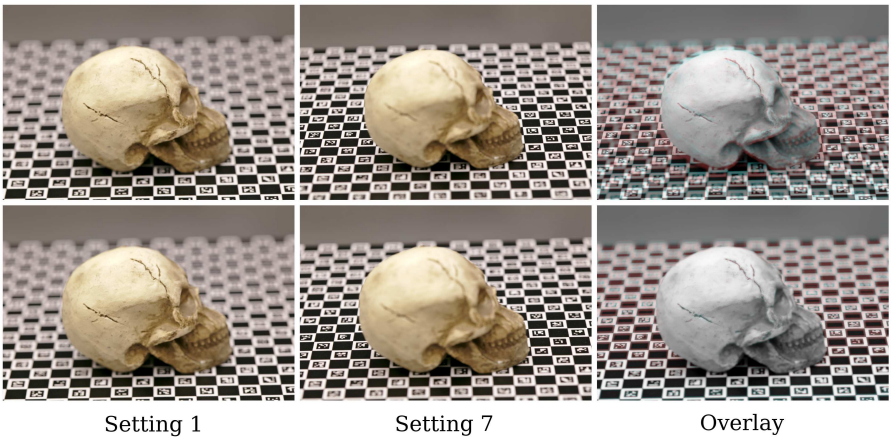

Figure 5: View 4 of the Skull dataset before (top row) and after (bottom row) registration.

3 Defocus Modelling

To evaluate the effectiveness our of model without knowing ground truth geometry, we synthetically generate defocused images to match those present in the captured focal stack using our DFD reconstructions. To do so, we estimate the radiance (i.e. the fully-focused appearance of the scene) by applying a focus measure to the focal stack. However, since our focal stacks contain relatively few images, some regions may not appear sharp. As a result, the absolute performance of the model cannnot be properly evaluated in this way, but its relative performance can. Here, we compare against the traditional thin lens model using the first view of each dataset. Figures 6, 7 and 8 illustrate the improvement our model has over the thin lens model, and Table 6 quantitively verifies this using the peak signal to noise ratio (PSNR) compared to the ground truth.

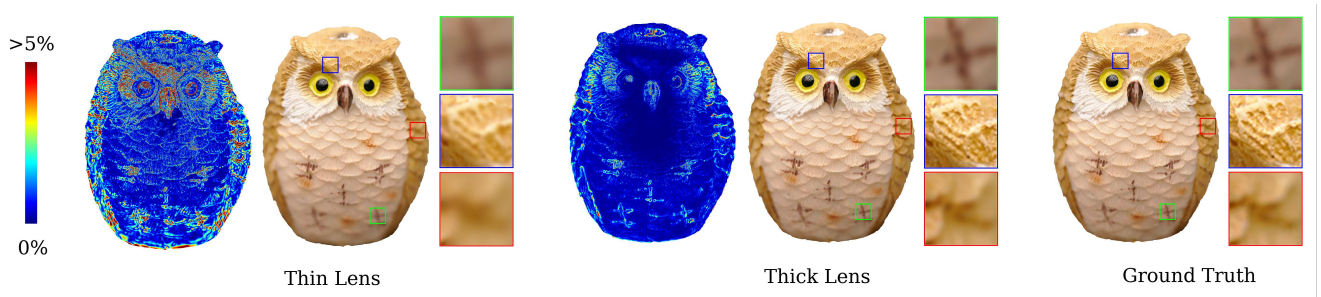

Figure 6: Reblurring results on the Owl dataset (setting 1). Error maps show absolute difference compared to the ground truth, clipped at 5% of the maximum pixel value.

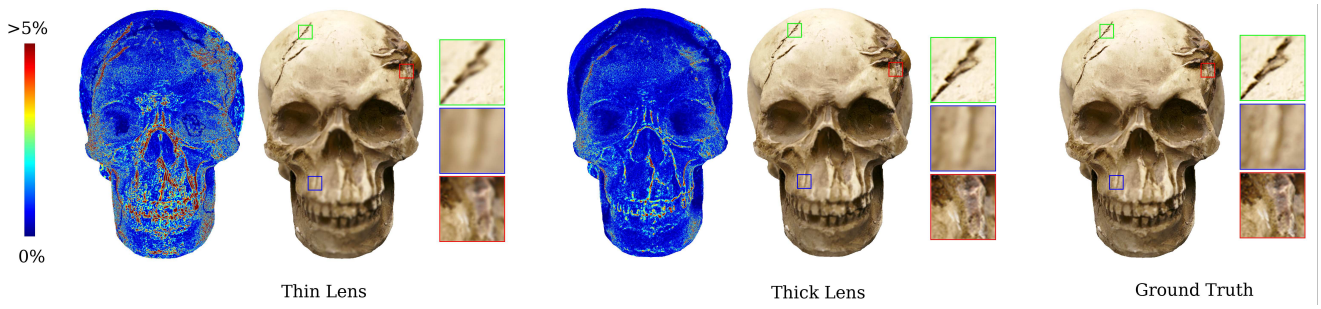

Figure 7: Reblurring results on the Skull dataset (setting 3). Error maps show absolute difference compared to the ground truth, clipped at 5% of the maximum pixel value.

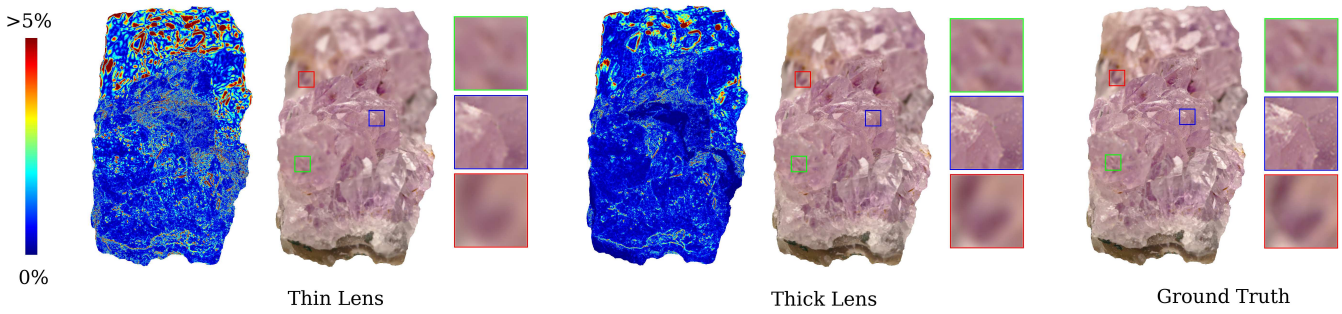

Figure 8: Reblurring results on the Quartz dataset (setting 2). Error maps show absolute difference compared to the ground truth, clipped at 5% of the maximum pixel value.

Dataset	Model	Focus Setting PSNR (dB)						
		0	1	2	3	4	5	6
Owl	Thin	30.80	32.62	32.31	29.32	28.75	N/A	N/A
	Thick	**35.83**	**39.28**	**38.90**	**35.07**	**32.65**	N/A	N/A
Skull	Thin	31.78	32.81	32.27	32.59	31.49	30.42	29.64
	Thick	**35.51**	**37.10**	**37.88**	**37.57**	**35.69**	**34.31**	**32.81**
Quartz	Thin	33.07	32.95	33.76	33.49	32.49	32.73	31.66
	Thick	**37.30**	**38.40**	**37.63**	**37.56**	**35.88**	**37.93**	**35.72**

Table 6: Results of the comparison between the synthetically generated and the captured images. These results were generated from a single view and across all focus settings for each dataset. In all cases, the thick lens model outperforms the thin lens model.

4 Reconstructions

We demonstrate the effectiveness of our model using our novel reconstruction framework. For the results shown, we set $\lambda = 10^4$, $\Psi_{max} = 10^3$ and run for 5 iterations. Skull and Owl use 100 labels, while Quartz uses 50 labels. See paper for details. Note that we exclude views 9 and 16 of the Owl dataset from the final mesh.

In Figures 9, 10 and 11, we present the reconstructions for each dataset. The reconstructions produced by the thin lens model do not combine to form a cohesive mesh because of invalid parameterisation, and due to inherent limitations of the model. The single view depth maps exhibit significant errors. While some surface detail is recovered, this is overshadowed by the incorrect overall structure. In contrast, our model successfully recovers highly detailed surfaces which are geometrically consistent. This allows multiple views to be combined to form a coherent model free from occlusions.

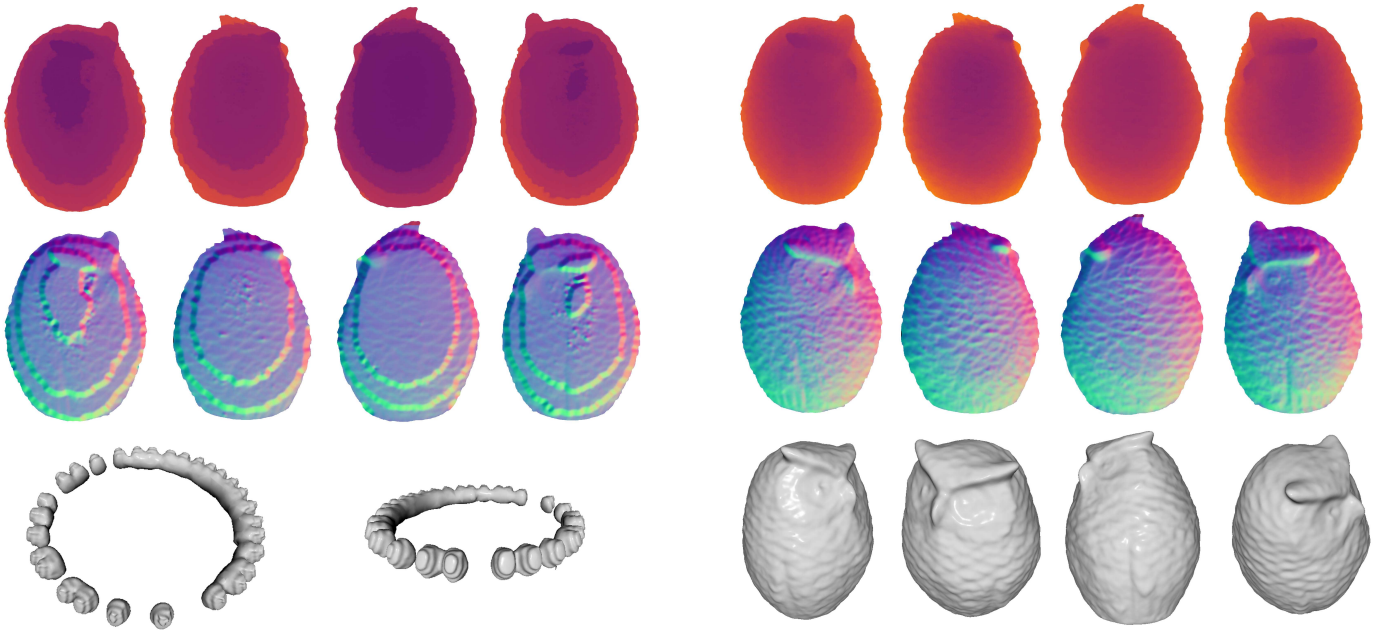

Figure 9: Owl reconstruction results using the thin lens (left) and our thick lens model (right). Here, we show a number of single view reconstructions (top row) with corresponding normals (middle row), and novel viewpoints of the final mesh (bottom row).

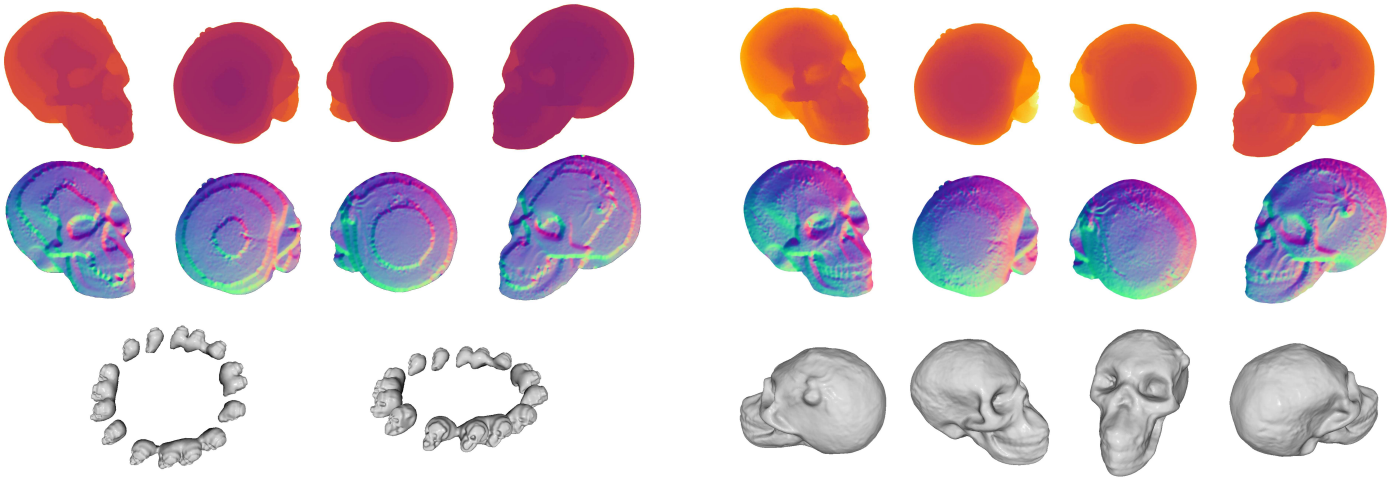

Figure 10: Skull reconstruction results using the thin lens (left) and our thick lens model (right). Here, we show a number of single view reconstructions (top row) with corresponding normals (middle row), and novel viewpoints of the final mesh (bottom row).

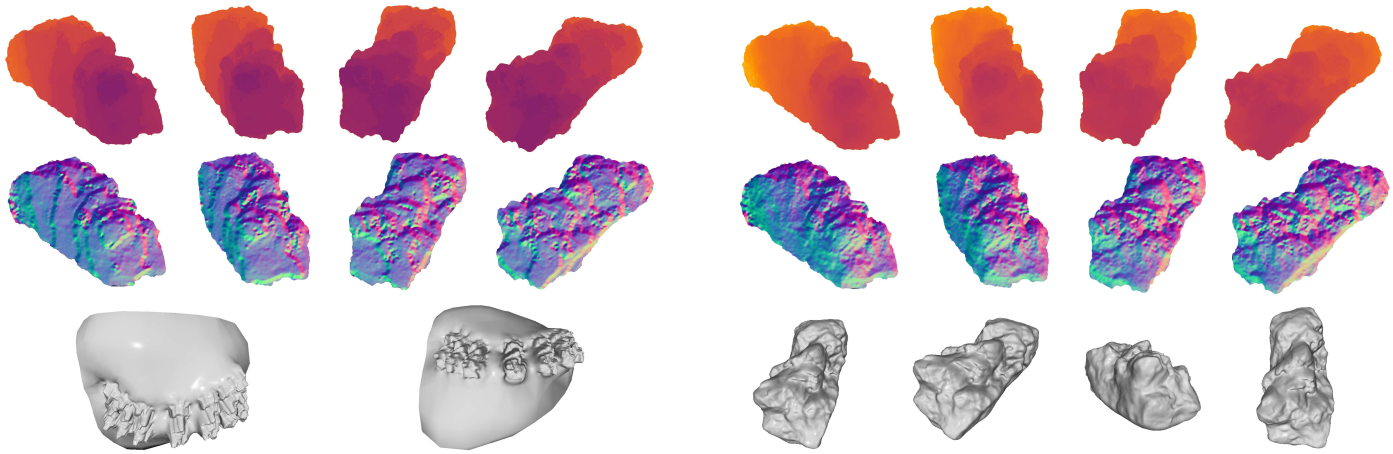

Figure 11: Quartz reconstruction results using the thin lens (left) and our thick lens model (right). Here, we show a number of single view reconstructions (top row) with corresponding normals (middle row), and novel viewpoints of the final mesh (bottom row).