

LUCES: A Dataset for Near-Field Point Light Source Photometric Stereo

Roberto Mecca¹ Fotios Logothetis¹
¹ Cambridge Research Laboratory,
Toshiba Europe, Cambridge, UK
{rmecca, flogothetis}@crl.toshiba.co.uk

Ignas Budvytis² Roberto Cipolla^{1,2}
² University of Cambridge
Cambridge, UK
{ib255, rc10001}@cam.ac.uk

Abstract

Three-dimensional reconstruction of objects from shading information is a challenging task in computer vision. As most of the approaches facing the Photometric Stereo problem use simplified far-field assumptions, real-world scenarios have essentially more complex physical effects that need to be handled for accurately reconstructing the 3D shape. An increasing number of methods have been proposed to address the problem when point light sources are assumed to be nearby the target object. The proximity of the light sources complicates the modeling of the image formation as the light behaviour requires non-linear parameterisation to describe its propagation and attenuation.

To understand the capability of the approaches dealing with this near-field scenario, the literature till now has used synthetically rendered photometric images or minimal and very customised real-world data. In order to fill the gap in evaluating near-field photometric stereo methods, we introduce LUCES the first real-world 'dataset for near-field point light soUrCe photomEtric Stereo' of 14 objects of a varying of materials. A device counting 52 LEDs has been designed to lit each object positioned 10 to 30 centimeters away from the camera. Together with the raw images, in order to evaluate the 3D reconstructions, the dataset includes both normal and depth maps for comparing different features of the retrieved 3D geometry. Furthermore, we evaluate the performance of the latest near-field Photometric Stereo algorithms on the proposed dataset to assess the SOTA method with respect to actual close range effects and object materials.

1. Introduction

Since the introduction of the Photometric Stereo problem (PS) by Woodham in the early '80s [37], a wide variety of approaches tackled the very same problem of reconstructing 3D geometry of an object under varying illumination from the same view point. Despite the very simplified assumption in [37] to make the PS problem solvable as an



Figure 1. Our Photometric Stereo setup consists of 52 LEDs placed within the same plane of the camera image. 14 objects have been acquired at a distances 10-30cm in order to capture full near-field effects. The materials of the objects provide a variety of diffuse and specular reflections, including wood, aluminium, plastic, etc.

(over-determined) linear system, similar simplifications are often still considered nowadays to make the problem applicable to real-world scenarios. Nonetheless, diffuse material assumption was relaxed in [15, 36, 34], camera perspective viewing was modelled in [35, 23], and robust optimisation methods were employed by [10, 14] to increase robustness to outliers. Light calibration assumption was also relaxed by [30, 27].

One of the most challenging aims of more recent PS methods is realistic illumination modelling, as uniform directional lighting is hard to achieve in practice. For this purpose, several methods have proposed using point light sources instead of directional ones [16, 8, 24, 26, 28, 19, 17]. Point light sources require non-linear modeling of light propagation and attenuation but they are a more realistic assumption than directional lights for near-field photometric imaging acquisitions. Note that proximity of the camera and lights to the object are very favorable in order to capture detailed geometry and minimise the ambient light interference. For example, near-field photometric stereo has been used in practice with handheld acquisition devices [12] and in endoscope-like inspections [9]. Whereas, the far-field assumptions do not allow to combine PS with multi-view for



Figure 2. From left to right: (1) the stage of our Photometric Stereo setup (2) a top view of a sample object (Squirrel), (3) acquisition with the GOM scanner (4) the 3D scanned mesh.

volumetric reconstruction [20].

However, despite the increased contribution from the computer vision community to tackle the near-field PS problem, the evaluation of such methods has relied on synthetic ([17]) or very minimal real-world datasets [28, 33]. The lack of shared data has prevented detailed and fair comparisons across the different methods. The aim of this work is to provide a comprehensive near-field PS benchmark with a variety of objects having different materials in order to evaluate several algorithms and understand their strengths and weaknesses. For this purpose, ground truth normal map and depth are provided for each object.

Our contribution is as follows:

- introducing the first near-field PS dataset of 14 real objects having a wide variety of materials;
- evaluating most relevant algorithms for the near-field PS problem and establish the actual SOTA method.

The dataset (including all images, light and camera calibration parameters and ground truth meshes) and the evaluation of the methods will be made available for download at <https://www.toshiba.eu/pages/eu/Cambridge-Research-Laboratory/download-luces>.

2. Related Work

A number of approaches for the PS problem has been proposed since it was first introduced [37]. We refer to some fairly recent surveys [3, 11] to cover the initial evolution of the PS methodologies. Here we discuss more contemporary algorithms and the datasets used to evaluate their performances.

2.1. PS datasets

Across the years, a number of custom real-world PS datasets have been created to suit the purposes of the proposed approaches. Alldrin *et al.* [4] proposed a dataset consisting of 3 objects lit by roughly a hundred distant light directions. The light calibration on terms of positioning and

intensity has been performed by using respectively a mirror sphere and a diffuse sphere. Xiong *et al.* [38] have proposed a dataset of 7 objects using 20 directional lights calibrated with two chrome spheres. As the approach was mostly modeling PS images with Lambertian irradiance equations, the material of the objects was quite diffuse. A limited number of PS data has been released by Quéau *et al.* to prove the working principle of an edge preserving method [29] and a multi-spectral PS approach [31].

Although initially designed for evaluating multi-view approaches, the datasets released by Aanæs *et al.* [1, 2] are useful for works on Photometric Stereo problem as they also contain images under varying illumination.

As most of the methods aimed at tackling the PS problem deal with the far-field setting, recently Shi *et al.* [34] introduced the first dataset in this category, namely DiLiGenT aimed at evaluating reconstruction methods over a wide variety of materials for 10 different objects. This work also contains a well discussed taxonomy for non-Lambertian and uncalibrated PS approaches. 96 LEDs were placed several meters away from the objects to approximate directional illumination and the camera (with a 50mm lens) was placed at 1.5m from the object. Such distance between the object and the camera/lights system does not provide to this dataset the near-field light variation studied in many recent approaches.

Near-field datasets: There are very limited, proper near-field labeled data including a single object from [28] and 3 simple objects from [33].

2.2. Near-field PS

The near-field setting is intrinsically more complicated to model than the far-field on as it requires handling not only different type of BRDFs [5, 25] but also anisotropic light propagation [26], inconsistent light intensity among the set of LEDs [32, 19] and finally the uncalibrated case [27]. Given this wide variety of difficulties, most of the proposed near-field PS methods have proposed custom PS data.



Figure 3. Top view of the objects captured for this dataset. Below every object the acquisition distance between the object and the camera, and the material of the object are reported.

With the aim to tackle all these issues simultaneously for the near-field PS, latest approaches have been exploiting deep learning capability training their networks with synthetically rendered data and data driven rendered data [21]. In particular, Logothetis *et al.* [17] used a per-pixel training strategy that allows to render unlimited data without carrying any training dataset. The sim-to-real gap is then filled by augmenting the data with physical effects such as noise, ambient light, interreflections, etc. Finally, the far to near field compensation is performed by integrating the normal field to compute the depth. By doing so iteratively, the method converges to an estimate of the 3D geometry. Santo *et al.* [33] have recently introduced a near-field PS method where the near-field compensation is computed after computing the far-field normals map from PSFCN [6]. The surface optimisation is performed through a differentiable renderer which fuses the normal predictions and the lighting model to re-project to the original images. This step limits the evaluation of the method to small images due to very high requirements of GPU RAM (around 20GB for 0.5Mpx images). Furthermore, despite the near-field setting, the camera viewing is assumed orthographic.

3. Data Capture

This section gives an overview of the data capture and calibration procedure.

3.1. Photometric Stereo Data Capture

The Photometric Stereo setup. Our setup (see Figure 2) consists of the following main components:

- RGB camera FLIR bfs-u3-32s4c-c with 8mm lens
- 52 LED Golden Dragon OSRAM

- variable voltage for adjustable LED power
- Arduino Mega 2560

A custom printed circuit board (PCB) has been designed to host 52 bright LED controlled with by an Arduino Mega. The configuration of the LEDs was planar around the camera, as shown in Figure 1. A single set of images was captured per object and the camera parameters (acquisition and shutter speed) and LED voltage were adjusted to achieve the best object exposure, which is very critical for specular objects. All camera preprocessing was turned off during the acquisition, including white-balance and analog gain.

Several optomechanical tools have been used for holding the camera and the PCB jointly. A manual XYZ translation stage with differential adjusters has been used to positioning the PS camera accurately through the printed circuit board.

In order to limit interreflections and ambient light, the walls surrounding the setup have been covered with black, polyurethane-coated nylon fabric.

Camera Intrinsics. This is performed using 100 checkerboard images and the OpenCV calibration toolbox. Fourth degree radial distortion is estimated and this is used to rectify all the images. The calibration re-projection error was 0.42px. The RAW data (before demosaicing and rectification) will also be made available.

Near Lighting Model. The lighting model is the anisotropic point light sources [24] which is used for all SOTA methods ([19], [28], [17], [33] evaluated in Section 4). This model assumes that a light source m , has a position $\mathbf{P}_m \in \mathbb{R}^3$, principal direction $\mathbf{D}_m \in \mathbb{R}^3$, RGB brightness $\Phi_m \in \mathbb{R}^3$ and angular dissipation factor

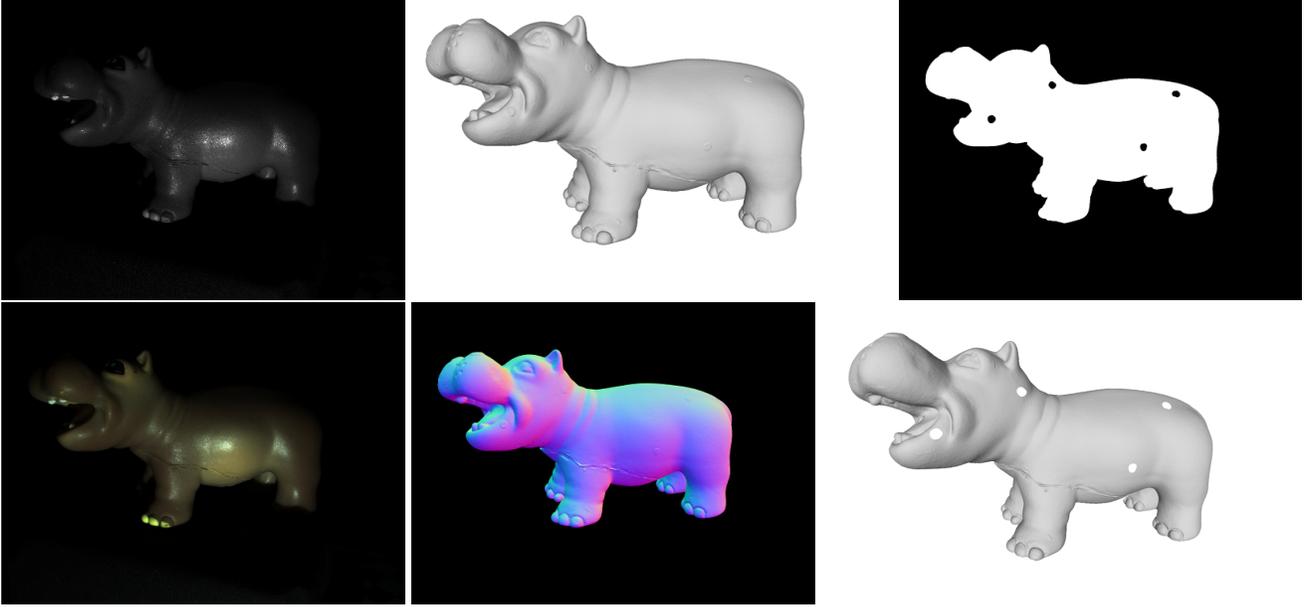


Figure 4. Demonstration of the steps performed per object: Firstly, compensation for radial distortion and demosaicing is performed on raw images to get RGB ones (left). Laser-scanned ground truth meshes are aligned with RGB images and ground truth normal maps are rendered (middle). Segmentation mask are generated (removing the pixels corresponding to markers) and ground truth normals are integrated [29] to obtain pseudo-depth (right) (see Section 4).

$\mu_m \in \mathbb{R}$. Therefore, a point $\mathbf{X} \in \mathbb{R}^3$ has a lighting vector $\mathbf{L}_m(\mathbf{X}) = \mathbf{P}_m - \mathbf{X}$ and light attenuation

$$a_m(\mathbf{X}) = \frac{(\hat{\mathbf{L}}_m(\mathbf{X}) \cdot \hat{\mathbf{D}}_m)^{\mu_m}}{\|\mathbf{L}_m(\mathbf{X})\|^2}, \quad (1)$$

where $\hat{\mathbf{L}}_m = \frac{\mathbf{L}_m}{\|\mathbf{L}_m\|}$ is the normalised light direction.

Calibration. In order to estimate the LED parameters (\mathbf{P}_m , \mathbf{D}_m , Φ_m and μ_m), a supplementary set of images have been acquired for LED light calibration. This includes measuring light intensity using a LuxMeter and images of a white diffuse plane; we used a diffuse reflectance plane with 99% nominal reflectance in UV-VIS-NIR wavelength range (350 - 1600nm). For every object, the calibration plane was captured twice, at different distances, in order to get data redundancy and produce a more accurate calibration. Thus, the Lambertian calibration object with albedo ρ and surface normal \mathbf{N} , should satisfy the resulting image irradiance equation:

$$\mathbf{I}_m = \Phi_m a_m \rho \mathbf{L}_m \cdot \mathbf{N}. \quad (2)$$

The irradiance Equation 2 was implemented into a differentiable renderer (using Keras of Tensorflow v2.0) with the LED parameters being the the model weights thus allowing refinement from a reasonable initial estimate. The parameters were initialised as follows: Φ_m for the LUX meter, $\mathbf{D}_m = [0, 0, 1]$, $\mu_m = 0.5$, \mathbf{P}_m from the schematic

of the LED circuit and $\rho = 1$. Using L_1 loss function for 30 epochs and converged to around 0.005 error i.e 0.5% of the maximum image intensity. The complete calibration parameters are included in the dataset.

3.2. 3D Ground Truth Capture

3D capturing device. 3D ground-truth has been acquired with the optical 3D scanner GOM ATOS Core 80/135 with a reported accuracy of 0.03mm (see Figure 2). The GOM scanner uses a stereo camera set-up and more than a dozen scans were performed and fused per object. In order to keep the geometry of the object consistent with the PS data, no spray coating has been used to ease the acquisition. Instead, markers were used for some objects.

Alignment. The laser scans of the objects were aligned and merged using MeshLab [7]. Some manual removal of noisy regions was performed and finally screened Poisson reconstruction was used in order to obtain full continuous surfaces (which are both useful for rendering normal maps and for mutual information alignment). As expected, not all parts of the surfaces of all objects have the same amount of noise, especially the metallic objects (Bell, Cup). Meshes were aligned with the photometric stereo images using the mutual information registration filter of MeshLab. This was initialised manually and pixel perfect accuracy was achieved. Using the aligned meshes, ground truth normal maps were rendered (using Blender). In addition, man-

ual segmentation was performed to remove regions where the GT was unreliable (markers on the objects, holes etc). Finally numerical integrations of the GT normals was performed (using [29]) to obtain *pseudo-depth* which is used to perform fair evaluation (see Section 4). The steps per object are summarised in Figure 4.

3.3. Dataset Overview

For each of the 14 object, 52 PS images have been acquired using the BayerRG16 RAW format. The total amount of PS images amounts then to 728. For all objects, rectified RGB PS images will be released (by compensating for the radial distortion). We note that color balancing was not performed on the images as this will distort the saturated pixels (which is an important feature for CNN-based PS methods [18, 13]). Instead, RGB light source brightness are provided along with the rest of point light source parameters. Both normal map and depth ground truth will be provided in order to evaluate the accuracy of near-field PS methods with either cases.

4. Experiments

In this section, we evaluate four competing near-field methods namely [19, 28, 17, 33]. In addition, we also evaluate with [13], the best performing far-field method (on the far-field benchmark [34]) to demonstrate the need for a near-field method.

Evaluation hyper-parameters. [19, 28] and [33] have publicly available code whereas for [17] we contacted the authors. Indeed, [17] has the disadvantage that the light configurations has to be known at train time therefore specific light positions had to be assigned to the networks to be trained for the dataset. [19] performs best with a manual initialisation of the specularity parameter c (0 is fully specular, 1 fully Lambertian) we used 0.1 for the Cup, 0.2 for the Bell, 0.25 for the Bawl and Tool, 0.5 for the Ball, Die, Hippo, Jar and Squirrel, 0.75 for the Bunny and 1 for the rest. For [28], we used the Cauchy estimator with 0.5 on the respective hyper-parameter. For both [19] and [28] we disabled the lighting calibration parameter. For all methods, we evaluated on full resolution images (2048x1536) except for [33], which is severely limited by GPU RAM so we had to subsample to (512x384) which was the maximum we could fit on 24GB Nvidia Titan RT. All other approaches are CPU RAM limited but ‘only’ require around 120GB. The computation time was varied from around 15 minutes (the fastest was [19] on the Bowl) to around two hours (the slowest was [17] on the Jar). For all of the methods, the initialisation was a flat plane at the mean depth computed exactly using the GT depth map.

Finally, we also evaluated the far-field method [13]. The assumed lighting direction was set the overage one for each

light and numerical integration was used on the output normal map to be able to compare surfaces.

Evaluation metrics. As it is the standard in PS literature, we first evaluate the competing approaches using the angular error on normal maps. We note that [19] and [28] output surfaces as dense depth maps, therefore the normals have to be estimated using finite differences. The other 3 methods output both surfaces and normals. It is very important to mention, that normal evaluation has two major limitations. Firstly, for real data, there can be regions where the ground truth normal uncertainty is non-negligible. This is inevitable due to capturing surfaces with a laser scanner that only provides very dense point clouds. Even micro-meter accuracy on the surface can generate a few degrees uncertainty of normals in regions of complicated geometry. The second important issue with evaluating on normals is that even on synthetic data, ground truth normals are not fully consistent with the ground truth depth. This is inevitable due to the fact that for any non-trivial object, the projection operation generates a depth map that is discontinuous and non-differentiable for a significant portion of the pixels. In fact, to quantify this discontinuity measure, we compared the ground truth normals with the normals that are obtained with differentiation of the ground truth depth and indeed observed a 3.3° error on average over the whole dataset (varying from 1.8° on the Jar to 9.2° on the House, as shown on the penultimate row of Table 1).

In addition, we also provide evaluation of the generated depth maps with respect to the ground truth. However, as all the evaluated approaches implicitly assume continuous and differential surfaces, the depth error is inflated at occlusion boundaries making the comparison between methods on the depth domain less informative. To overcome this limitation, we introduce the pseudo-depth, \tilde{Z} , generated by numerical integration of the ground truth normals. We note that this corresponds to a continuous and differentiable surface that is locally identical to the ground truth in all of the regions where the ground truth is continuous and differentiable. Therefore, evaluating the mean error with respect to the pseudo-depth allows for a more meaningful way of evaluating the ability of methods to generate surfaces. The difference between depth and pseudo-depth is quantified in the last row of Table 1 and it varies from 0.05mm for the mostly continuous Cup 7.8mm for the complicated geometry house. Complete error maps can be found in the supplementary.

5. Results

In this section, we analyse the performance of [13], [19, 28] [17], [33] which is shown quantitatively at Table 1 and qualitatively at Figures 5 and 6.

Method	Error	Bell	Ball	Buddha	Bunny	Die	Hippo	House	Cup	Owl	Jar	Queen	Squirrel	Bowl	Tool	Average
L17-[19]	MAE	28.25	9.77	11.5	20.15	11.95	15.42	29.69	30.76	13.77	10.56	13.05	15.93	12.5	15.1	17.03
	MZE	4.45	0.81	4.67	7.51	4.58	3.19	6.99	2.67	3.64	6.56	1.89	1.82	4.37	3.25	4.02
	MZE	4.33	0.92	6.56	8.29	4.13	2.33	3.16	2.66	1.68	6.48	2.1	2.25	4.02	2.92	3.7
Q18-[28]	MAE	25.8	12.12	14.07	13.73	13.77	18.51	30.63	37.63	14.74	15.66	13.16	14.06	11.19	16.12	17.94
	MZE	12.03	2.5	9.28	7.06	5.91	6.8	8.02	4.83	5.83	16.87	6.92	2.55	6.48	6.69	7.27
	MZE	11.77	2.36	6.65	5.21	5.85	6.12	2.94	4.82	3.28	16.37	3.89	1.99	5.72	6.11	5.93
S20-[33]	MAE	9.5	25.42	19.17	12.5	5.23	23.12	28.02	14.22	13.08	9.27	16.62	14.07	12.44	17.42	15.72
	MZE	1.9	5.5	5.53	6.02	2.76	7.04	6.15	1.62	3.75	6.09	3.91	2.81	5.22	4.68	4.5
	MZE	1.89	5.54	5.06	3.99	2.18	5.61	3.07	1.62	2.81	5.65	4.54	2.36	5.23	4.22	3.84
L20-[17]	MAE	14.74	12.43	10.73	8.15	6.55	7.75	30.03	23.35	12.39	8.6	10.96	15.12	8.78	17.05	13.33
	MZE	1.84	0.7	3.32	3.27	4.61	2.96	9.47	2.04	3.78	17.98	2.11	1.28	2.94	6.16	4.46
	MZE	1.63	0.69	4.17	1.83	5.11	1.26	2.36	2.04	1.53	18.31	1.62	1.63	2.68	5.97	3.63
I18-[13]	MAE	23.55	44.29	35.29	36	41.52	44.9	49.05	35.78	40.27	40.66	32.89	41.09	28.04	31.71	37.5
	MZE	5.93	6.59	10.92	6.88	7.83	7.59	8.98	3.17	8.67	15.54	8.08	5.8	6.69	12.45	8.22
	MZE	5.84	6.4	7.94	5.35	7.29	6.75	2.65	3.17	5.1	15.29	4.85	4.93	6.18	11.85	6.68
GT	N-MAE	2.5	2.69	2.69	2.93	2.49	3.2	9.19	2.85	4.3	1.79	4.22	3.26	2.27	2.34	3.34
	Z-MZE	0.29	0.29	3.3	3.18	0.96	2.92	7.77	0.05	3.9	0.95	3.34	1.57	0.82	0.93	2.16

Table 1. Complete evaluation of five methods on all objects. Mean angular error MAE (degrees), mean depth error MZE (mm) and mean pseudo-depth error MZE (mm) are reported. In addition, we report the angular error of normals generated by differentiation of GT depth N-MAE and the error between depth and pseudo-depth Z-MZE.

We first of all observe that the far-field method [13] fails to produce accurate results as expected. The two classical optimisation methods [19, 28] are outperformed by the deep learning approaches [33] and [17] on the normal error metric for most scenarios although [19] shows a good consistency and achieves the minimum depth error (which is not surprising as it is primarily a depth rather than normal optimisation method).

Despite the fact that [17] is the best overall performer, we emphasise that it requires knowledge of the setup at train time. In addition, their numerical integration can significantly amplify the depth error on oblique regions (depth real and pseudo) such as the top of the Jar (they achieve best normals and worst depth). [33] achieves the best performance on object with specular materials, because of the use of a patch-based network that extract the most information of the metallic object. It also achieves best MAE on the Bell. The orthographic camera assumption of [33] in terms of error translates to a growing inaccuracy towards the external part of the reconstruction (see Bell, Cup and Jar in Figure 5).

We also notice that the normal predictions are more noisy as opposed to depth prediction. This could be due to actual noisy estimates of the ground truth normals which is inevitable for any laser scanner (see in particular the Ball in Figure 5). As the ground truth depth is more reliable, it is a better evaluation metric compared to the ‘ground truth’ normals. See Figure 6 for depth evaluation.

By looking at Figure 5 it can be seen that even the best methods perform poorly for recovering the geometry of very oblique regions. This is observable for the Jar, Owl and Cup in Figure 5. Therefore these represent quite hard

regions to retrieve.

An interesting observation is that for both CNN-based methods ([17, 33]), the material’s specularity does not seem to be a significant factor of performance. Indeed, convex regions (where self reflections are negligible) are consistently recovered correctly regardless of the material: diffuse head of Queen, bronze Bell, plastic Hippo, wooden Bowl; with the only exception being the aluminium Cup. This is a clear advantage of CNN methods against the classical ones that require diffuse or mostly diffuse materials.

We observe that the hardest regions are the ones containing high frequency details (sharp boundaries) such as House, bottom part of the Squirrel details of the Queen etc. As PS relies on normals, non differentiable regions with non defined surface normals are problematic and this needs to be addressed with more sophisticated surface optimisation methods and/or direct depth retrieval.

6. Conclusion

In this work we proposed the first large scale dataset for the near-field PS problem. Differently from the far-field assumption, when the target object is close to the camera/light setup several non-linear physical effects as anisotropic light propagation, light attenuation and perspective viewing geometry occur. Current PS datasets, mostly consider scenarios where such effects are negligible as they provide directional light coordinates and use quite long-focus lenses giving orthographic viewing geometry.

Recent research trends on 3D reconstruction using PS have shown an increasing interest to deal under near-field settings. However, the lack of a dataset for this topic has

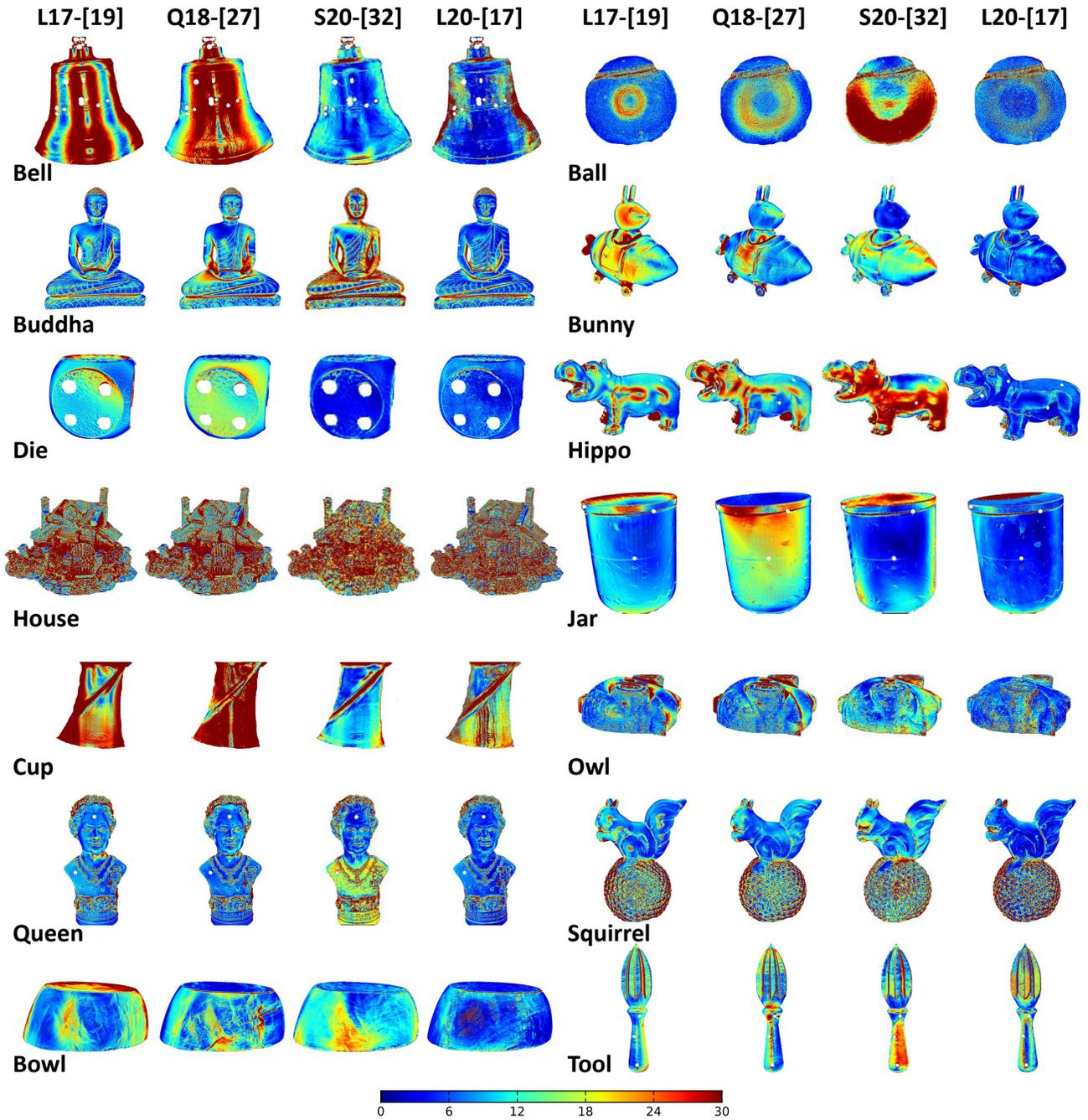


Figure 5. Normal error map comparison for all objects and all near-field methods.

prevented to fairly compare different approaches. For this reason, we benchmark the recent near-field PS approaches and analyse their performance over our dataset which include objects with a wide variety of materials. We also provide a discussion about appropriate ways of evaluation (depth vs normals). In addition, as we noticed that most of the error is expectedly concentrated on the edges and dis-

continuity regions we conclude that future research has to improve the interpretation of the PS imaging data in these specific areas and possibly exploiting networks with edge detection capability to better deal with interreflections.

Finally, it is worth investigating the possibility of using completely raw image data without demosaicing or radial distortion compensation. This requires incorporating

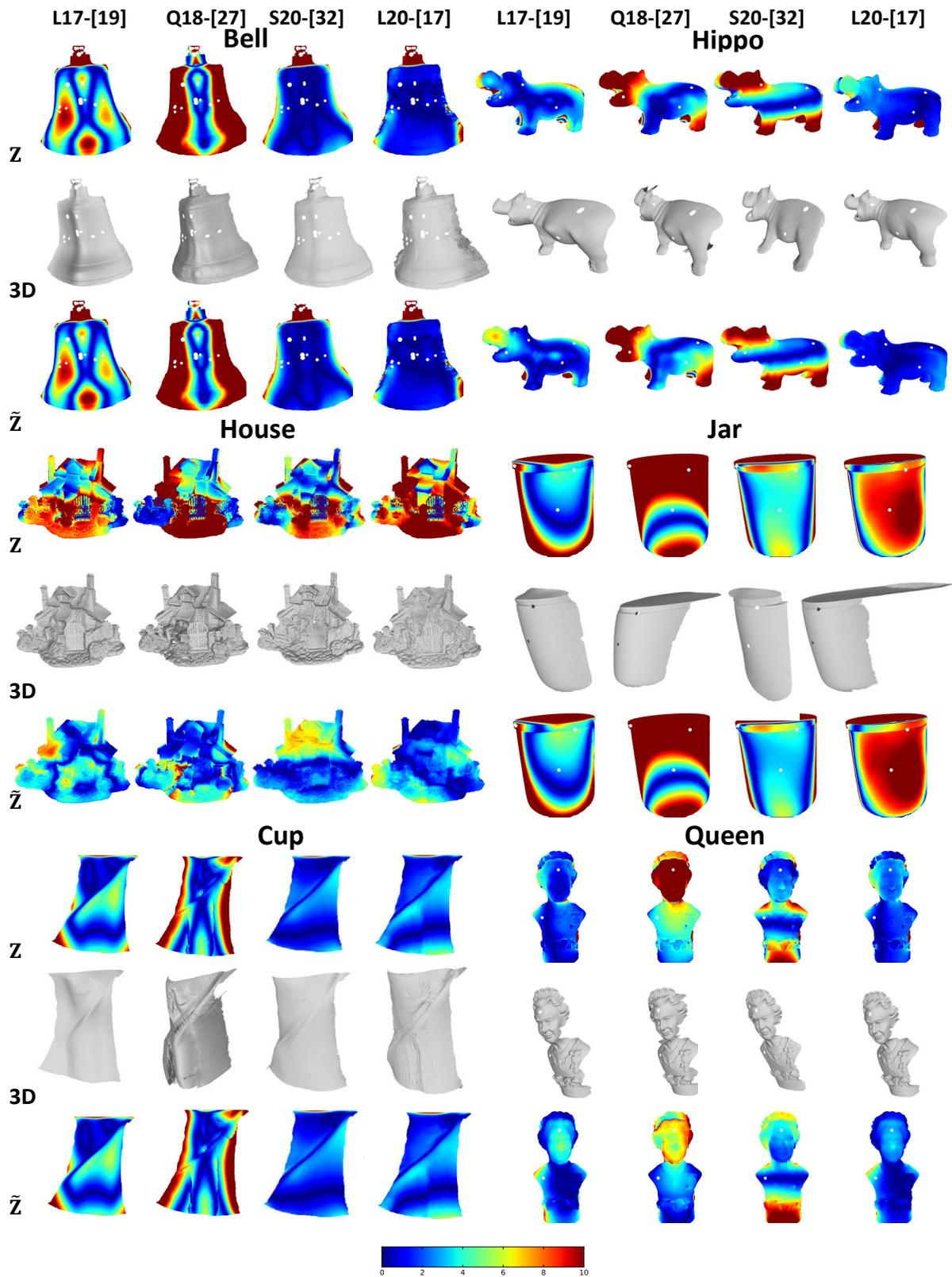


Figure 6. Output surface comparison for 6 objects and all methods. This is show qualitatively through the 3D meshes and well as error maps on depth Z and psedo-depth \tilde{Z} domain (errors in mm).

the radial distortion into the image irradiance equation and treating the images as pure intensity and ignoring the potential of recovering colours. The advantage of skipping these two pre-processing steps is the potential of eliminating some image artefacts, especially around image edges, which currently achieve the least accuracy.

References

- [1] H. Aanæs, A. Lindbjerg Dahl, and K. S. Pedersen. Interesting interest points - A comparative study of interest point performance on a unique data set. *Int. J. Comput. Vis.*, 97(1):18–35, 2012. 2
- [2] H. Aanæs, R. Ramsbøl Jensen, G. Vogiatzis, E. Tola, and A. B. Dahl. Large-scale data for multiple-view stereopsis. *Int. J. Comput. Vis.*, 120(2):153–168, 2016. 2
- [3] J. Ackermann and M. Goesele. A survey of photometric stereo techniques. *Found. Trends Comput. Graph. Vis.*, 9(3-4):149–254, 2015. 2
- [4] N. Alldrin, T. E. Zickler, and D. J. Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *CVPR*, 2008. 2
- [5] M. Chandraker, J. Bai, and R. Ramamoorthi. On Differential Photometric Reconstruction for Unknown, Isotropic BRDFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 35(12):2941–2955, 2013. 2
- [6] G. Chen, K. Han, and K.-Y. K. Wong. Ps-fcn: A flexible learning framework for photometric stereo. In *15th European Conference on Computer Vision (ECCV)*, 2018. 3
- [7] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia. Meshlab: an open-source mesh processing tool. Eurographics, 2008. 4
- [8] J. J. Clark. Active photometric stereo. In *1992 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 29–34, 1992. 1
- [9] T. Collins and A. Bartoli. 3d reconstruction in laparoscopy with close-range photometric stereo. In *MICCAI 2012*, 2012. 1
- [10] A. P. Harrison and D. Joseph. Maximum likelihood estimation of depth maps using photometric stereo. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(7):1368–1380, 2012. 1
- [11] S. Herbot and C. Wöhler. An introduction to image-based 3d surface reconstruction and a survey of photometric stereo methods. *3D Research*, 2(4), 2011. 2
- [12] T. Higo, Y. Matsushita, N. Joshi, and K. Ikeuchi. A handheld photometric stereo camera for 3-D modeling. In *2009 IEEE International Conference on Computer Vision (ICCV)*, pages 1234–1241, 2009. 1
- [13] S. Ikehata. Cnn-ps: Cnn-based photometric stereo for general non-convex surfaces. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 3–18, 2018. 5, 6, 12
- [14] S. Ikehata, D. Wipf, Y. Matsushita, and K. Aizawa. Robust photometric stereo using sparse regression. In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 318–325, 2012. 1
- [15] K. Ikeuchi. Determining surface orientations of specular surfaces by using the photometric stereo method. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 3(6):661–669, 1981. 1
- [16] Y. Iwahori, H. Sugie, and N. Ishii. Reconstructing shape from shading images under point light source illumination. In *International Conference on Pattern Recognition (ICPR)*, pages 83–87, 1990. 1

- [17] F. Logothetis, I. Budvytis, R. Mecca, and R. Cipolla. A cnn based approach for the near-field photometric stereo problem. *BMVC*, 2020. [1](#), [2](#), [3](#), [5](#), [6](#), [12](#)
- [18] F. Logothetis, I. Budvytis, R. Mecca, and R. Cipolla. Px-net: Simple, efficient pixel-wise training of photometric stereo networks. *arXiv preprint arXiv:2008.04933*, 2020. [5](#)
- [19] F. Logothetis, R. Mecca, and R. Cipolla. Semi-calibrated near field photometric stereo. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 3, page 8, 2017. [1](#), [2](#), [3](#), [5](#), [6](#), [11](#), [12](#)
- [20] F. Logothetis, R. Mecca, and R. Cipolla. A differential volumetric approach to multi-view photometric stereo. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1052–1061, 2019. [2](#)
- [21] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM Transactions on Graphics*, 22(3):759–769, 2003. [3](#)
- [22] R. Mecca, Y. Quéau, F. Logothetis, and R. Cipolla. A single lobe photometric stereo approach for heterogeneous material. *SIAM Journal on Imaging Sciences*, 9(4):1858–1888, 2016. [11](#)
- [23] R. Mecca, A. Tankus, A. Wetzler, and A. Bruckstein. A direct differential approach to photometric stereo with perspective viewing. *SIAM Journal on Imaging Sciences*, 7(2):579–612, 2014. [1](#)
- [24] R. Mecca, A. Wetzler, A. Bruckstein, and R. Kimmel. Near Field Photometric Stereo with Point Light Sources. *SIAM Journal on Imaging Sciences*, 7(4):2732–2770, 2014. [1](#), [3](#)
- [25] A. Ngan, F. Durand, and W. Matusik. Experimental analysis of BRDF models. In *Proceedings of the Sixteenth Eurographics conference on Rendering Techniques*, pages 117–126, 2005. [2](#)
- [26] Y. Nie and Z. Song. A novel photometric stereo method with nonisotropic point light sources. In *ICPR 2016*, pages 1737–1742. IEEE, 2016. [1](#), [2](#)
- [27] T. Papadhimetri and P. Favaro. Uncalibrated Near-Light Photometric Stereo. In *British Machine Vision Conference (BMVC)*, 2014. [1](#), [2](#)
- [28] Y. Quéau, B. Durix, Tao Wu, D. Cremers, F. Lauze, and J.-D. Durou. Led-based photometric stereo: Modeling, calibration and numerical solution. *Journal of Mathematical Imaging and Vision (JMIV)*, 60(3):313–340, 2018. [1](#), [2](#), [3](#), [5](#), [6](#), [11](#), [12](#)
- [29] Y. Quéau and J.-D. Durou. Edge-preserving integration of a normal field: Weighted least squares, TV and L1 approaches. In *SSVM*, 2015. [2](#), [4](#), [5](#), [11](#)
- [30] Y. Quéau, F. Lauze, and J.-D. Durou. Solving uncalibrated photometric stereo using total variation. *J. Math. Imaging Vis.*, 52(1):87–107, 2015. [1](#)
- [31] Y. Quéau, R. Mecca, and J.-D. Durou. Unbiased photometric stereo for colored surfaces: A variational approach. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4359–4368, 2016. [2](#)
- [32] Y. Quéau, T. Wu, and D. Cremers. Semi-calibrated near-light photometric stereo. In *International Conference on Scale Space and Variational Methods in Computer Vision (SSVM)*, pages 656–668, 2017. [2](#)
- [33] H. Santo, M. Waechter, and Y. Matsushita. Deep near-light photometric stereo for spatially varying reflectances. In *European Conference on Computer Vision (ECCV)*, 2020. [2](#), [3](#), [5](#), [6](#), [12](#)
- [34] B. Shi, Z. Mo, Z. Wu, D. Duan, S. K. Yeung, and P. Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018. [1](#), [2](#), [5](#)
- [35] A. Tankus and N. Kiryati. Photometric stereo under perspective projection. In *2005 IEEE International Conference on Computer Vision (ICCV)*, pages 611–616, 2005. [1](#)
- [36] L. B Wolff. Relative brightness of specular and diffuse reflection. *Optical Engineering*, 33(1):285–294, 1994. [1](#)
- [37] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):134–144, 1980. [1](#), [2](#)
- [38] Y. Xiong, A. Chakrabarti, R. Basri, S. J. Gortler, D. W. Jacobs, and T. E. Zickler. From shading to local shape. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(1):67–79, 2015. [2](#)

A. Appendix

This appendix provides supplementary material for the main publication. Section A.1 provides additional details of the Photometric Stereo camera and LED setup. Section A.2 provides a discussion on the inherent errors introduced in shape estimation from normals. Section A.3 provides a complete qualitative comparison of all 5 methods evaluated in the main publication.

A.1. Photometric Stereo Setup

This section gives additional details on the PS setup.

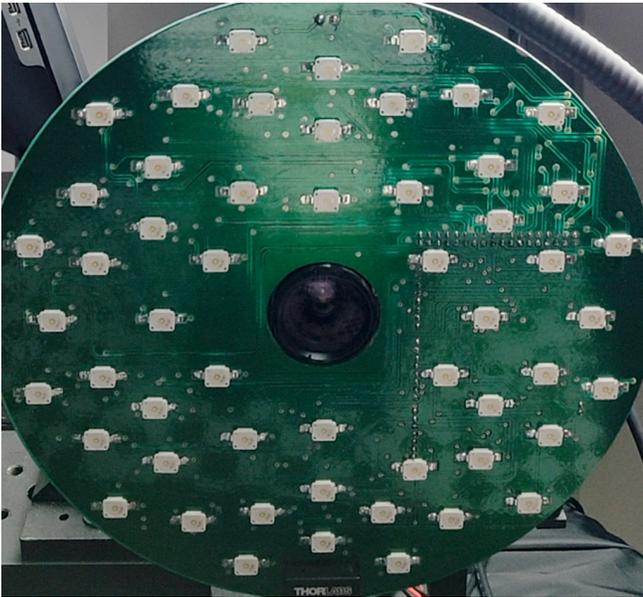


Figure 7. Close-up of the setup used for acquiring Photometric Stereo images.

It consists of the following main components (see Figure 7):

- RGB camera FLIR bfs-u3-32s4c-c with 8mm lens
- 52 LED Golden Dragon OSRAM
- variable voltage for adjustable LED power
- Arduino Mega 2560

The Arduino Mega 2560 controls the LEDs that are turned on and off individually during the Photometric Stereo image capture. The PCB has been designed to accommodate a specific 8mm lens that allows a reasonable wide field of view on a target object placed few centimeters away from the camera. The type of OSRAM LEDs (LW W5SN) are capable of emitting 5600 Kelvin white light up to 191 lumen. The 120° of viewing angle of the LEDs allows a complete lighting of the scene. The LEDs have been distributed on 6 different circumferences (all centered in the camera centre) of radii 35, 45, 55, 65, 75 and 85 mm. Depending on the circumference, the LEDs have been positioned at variable angle of 30° and 60° .

In order to avoid out of focus areas, the aperture of the lens is kept to the minimum ($f/11$). Therefore, variable time of exposures (among objects) have been used to prevent saturations or too dark regions. Note that a separate set of calibration images (see main paper Section 3.1) was captured for each different configuration.

The complete LED parameters (positions, brightness, directions, angular dissipation, see main paper Section 3.1) are included in the dataset.

A.2. Ground Truth Meshes

This section provides supplementary information about the laser-scanned meshes which were used as ground truth for the evaluation of the competing PS approaches. Note that because of the different sizes of the objects and the characteristics of the surfaces, the number of triangles in each object mesh varies from 260K to 6.1M as shown in Figure 8. As the scans can be of questionable quality at some regions due to visibility and/or specular material (i.e. numbers on the Die), manual segmentation was performed on the image domain to only evaluate on the reliable regions.

Ground truth discontinuity and non-differentiability.

Over the years, the majority of PS approaches has assumed that the surface can be described as a continuous and differentiable depth map. It has been acknowledged before [22] that this assumption is violated in practice and some care has been taken to include some robustness to it (e.g. L_1 loss in [19] and Cauchy estimator by [28]). However, we believe that the extent of this issue is underestimated and thus we attempt to quantify it here by computing the following two metrics. Firstly, we compute a normal map through numerical differentiation of the ground truth depth map [29] and compare it with the ground truth normal map. This is shown in Table 1 of the main submission and illustrated visually here in Figure 9. The mean per-pixel angular error of computed and ground truth normals is on average 3.34° with the maximum value being 9.19° on the complex geometry house object. We emphasise that this effect is completely independent of the actual uncertainty of the ‘ground truth’ meshes and it is solely caused by the projection operation (and so the effect would be identical in synthetic data). Note, the ground truth normal map is computed by rendering (i.e. projection, discretisation and occlusion) of the surface normals into the image plane which is quite different than numerical differentiation of the ground truth depth map. Indeed, in Figure 9, the error is concentrated on boundaries.

In addition, we compute the average per-pixel error between the ground truth depth and the depth obtained by numerical integration (using [29]) of the ground truth normals (i.e. pseudo-depth). The observed error is not-negligible

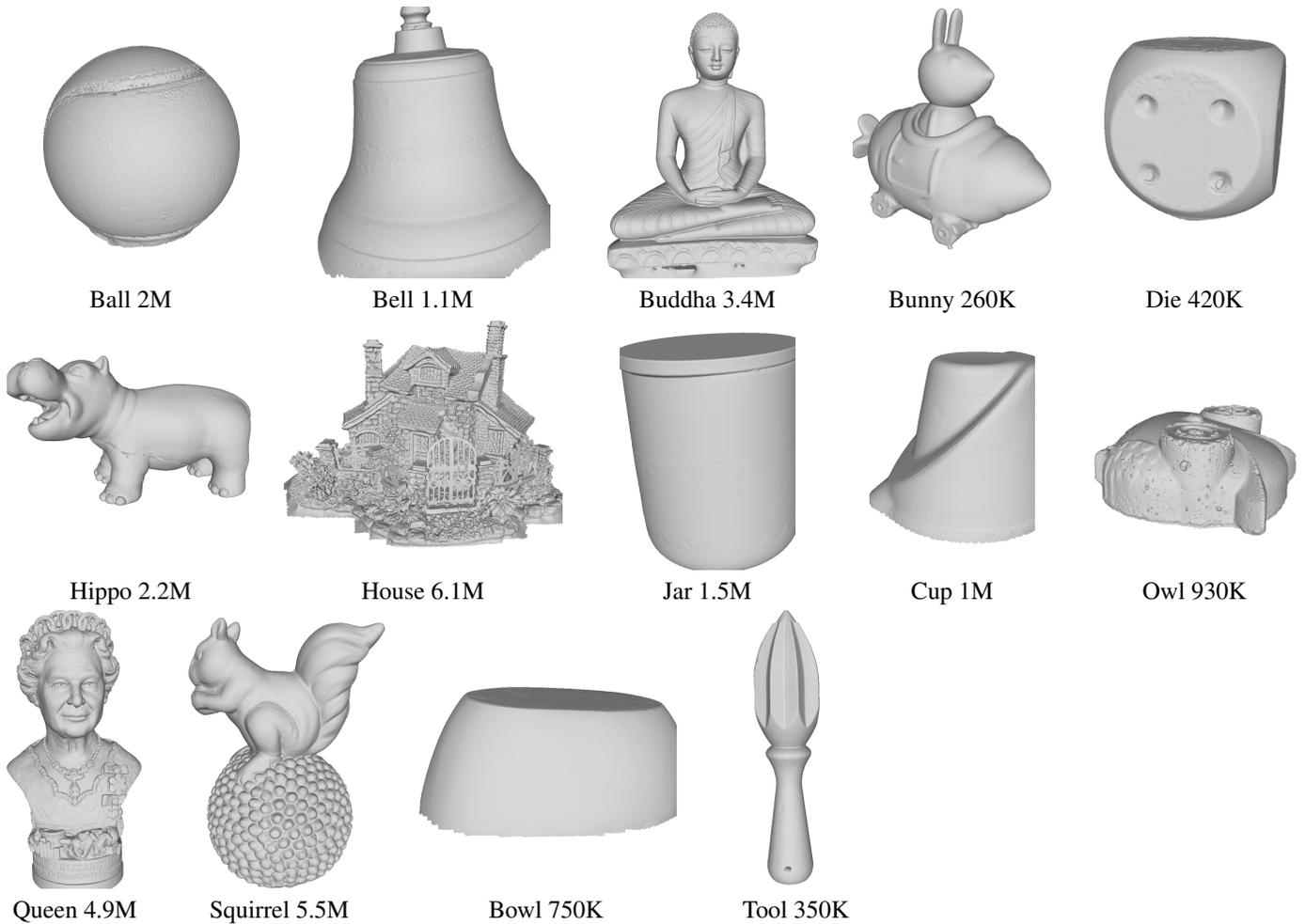


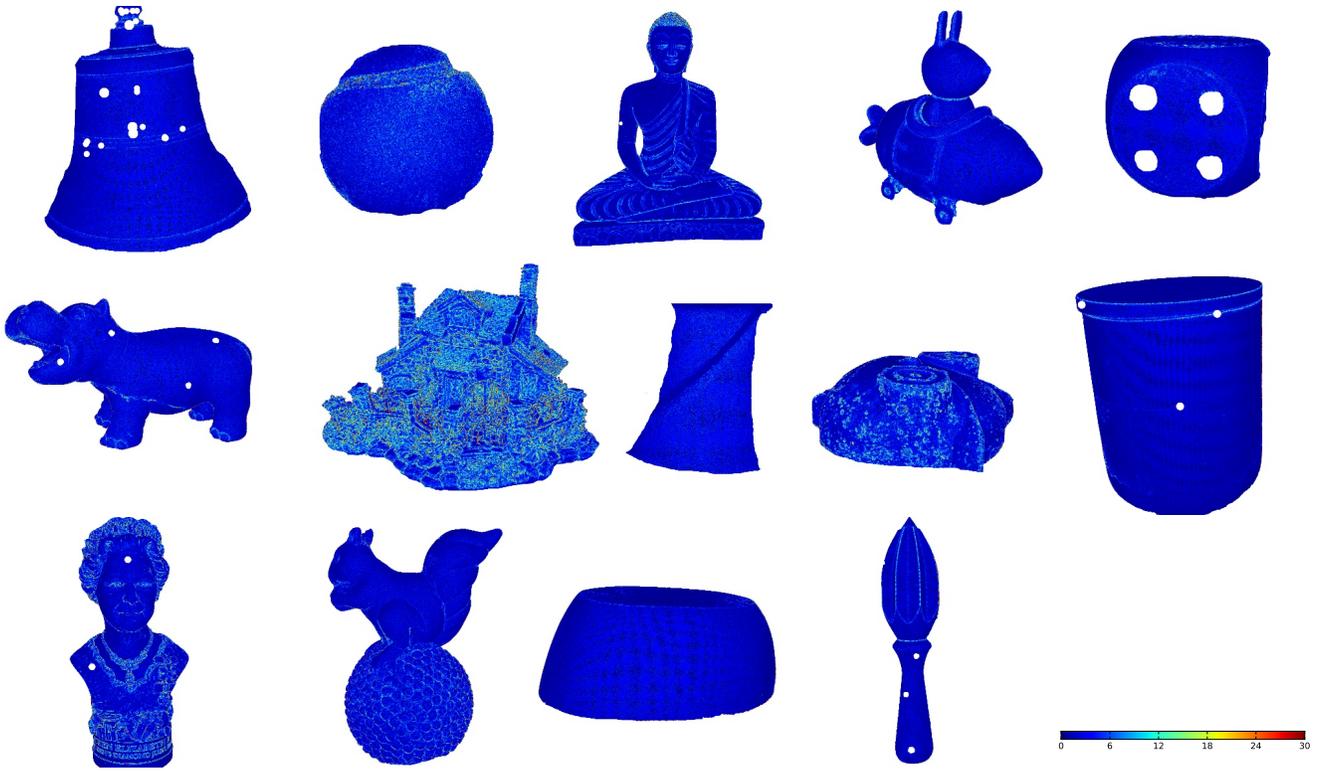
Figure 8. Laser-scanned meshes and their respective number of triangles.

(2.16mm on average) and it propagates outwards from occlusion boundaries (as the numerical integration preserves the actual mean depth).

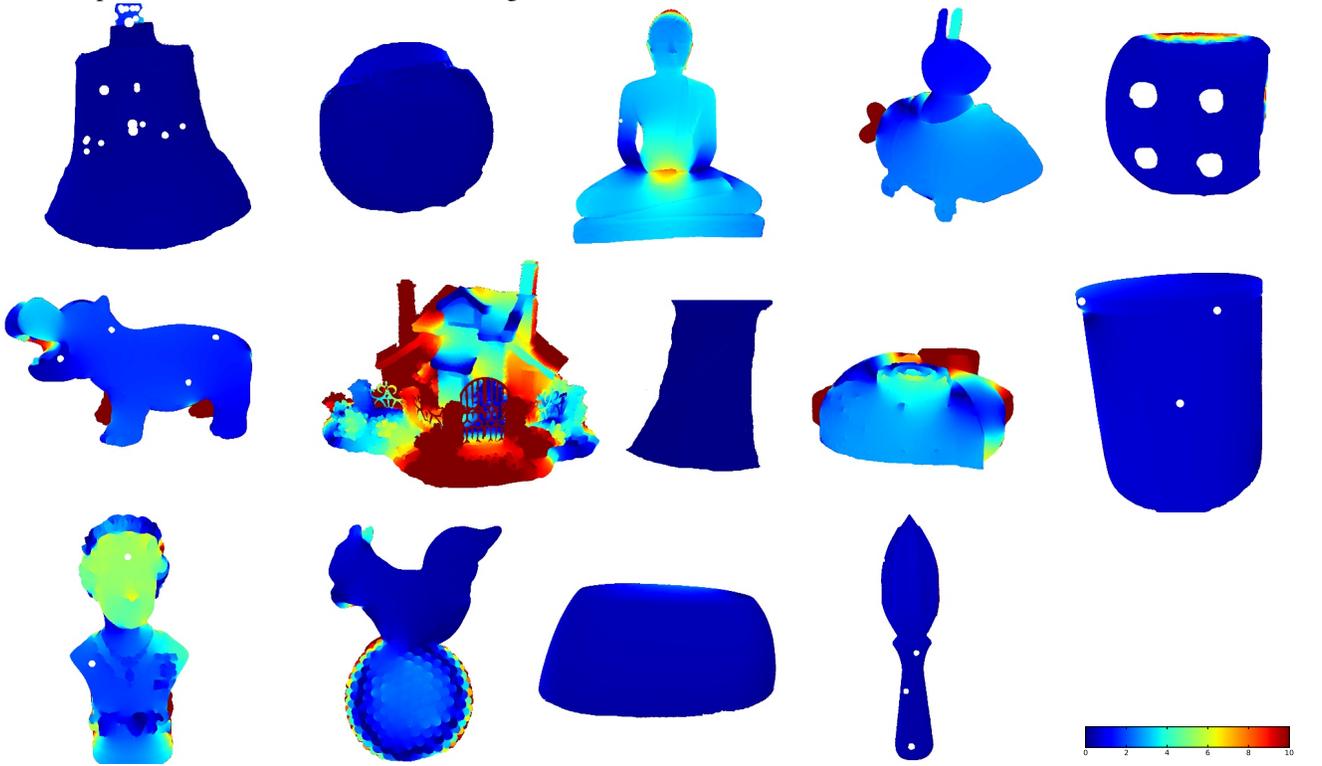
Finally, we note that the two error metrics explained above are likely to be close to the theoretical minimum (for normals and depth respectively) achievable by any approach that is reliant of the differentiable surface assumption. As these error bounds can be non-negligible, we motivate future research that avoids reliance of surface differentiability (e.g. direct depth regression).

A.3. Reconstructions

This section contains a complete qualitative comparison of all 5 methods L17 [19], Q18 [28] I18 [13], S20 [33] and L20 [17]. Estimated 3D surface view as well as depth Z and pseudo-depth \tilde{Z} error maps are provided in Figures 10 to 16. Note that errors of predicted normals are provided in the main publication.



Error maps between differentiated normals and ground truth normals.



Error maps between the pseudo-depth and the ground truth depth.

Figure 9. Normals and discontinuity.

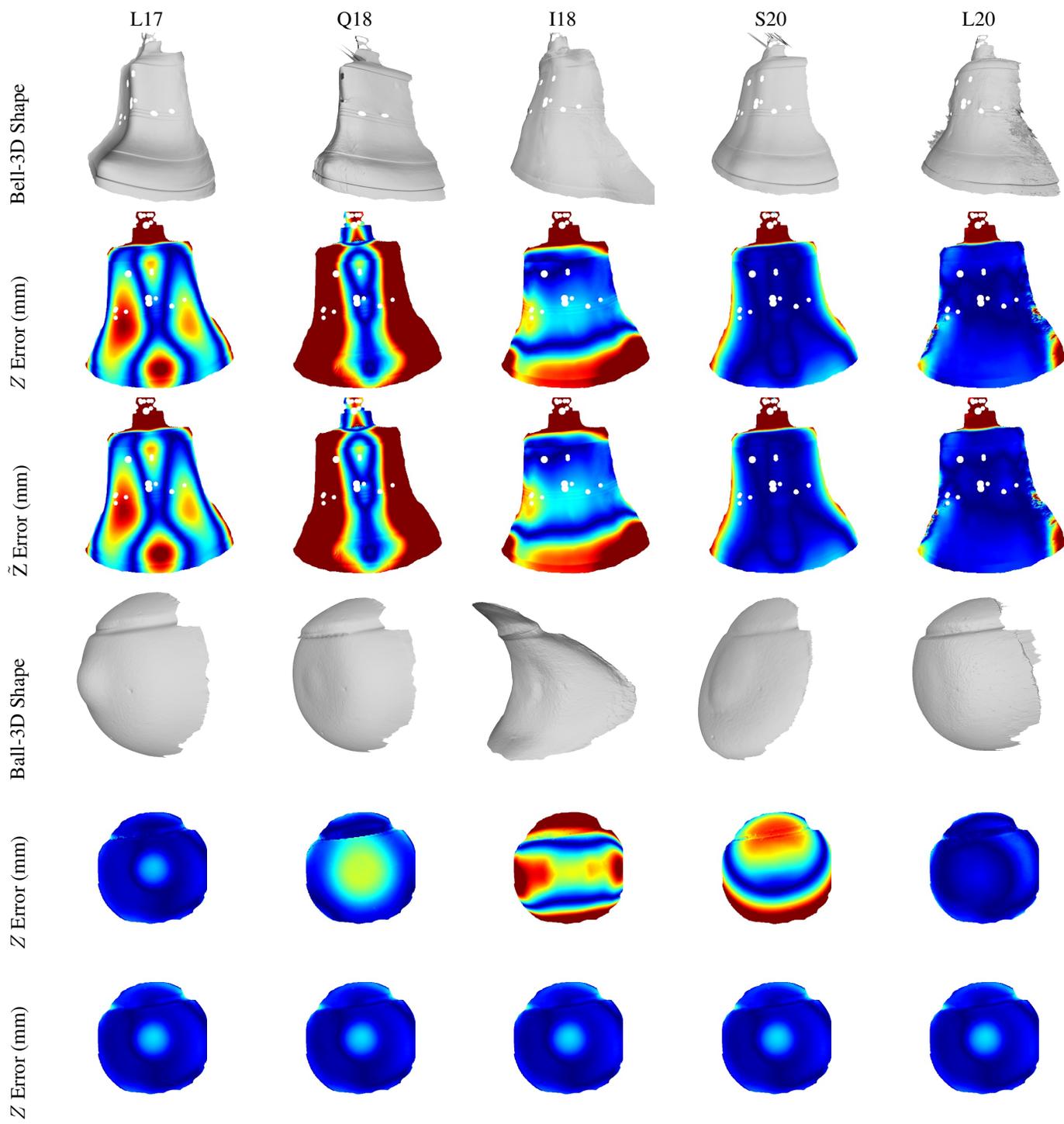


Figure 10. Evaluations 1-2/14

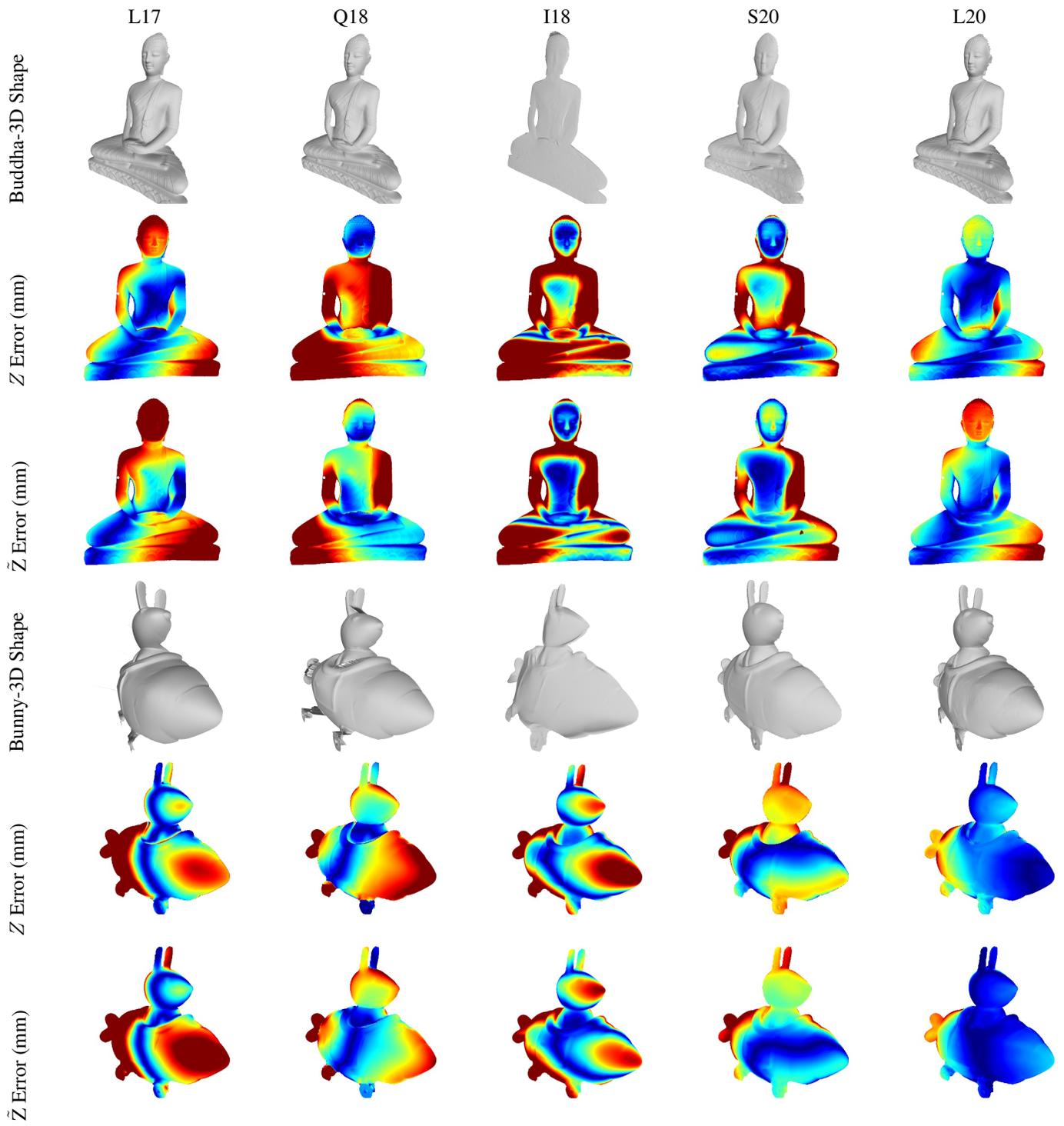


Figure 11. Evaluations 3-4/14

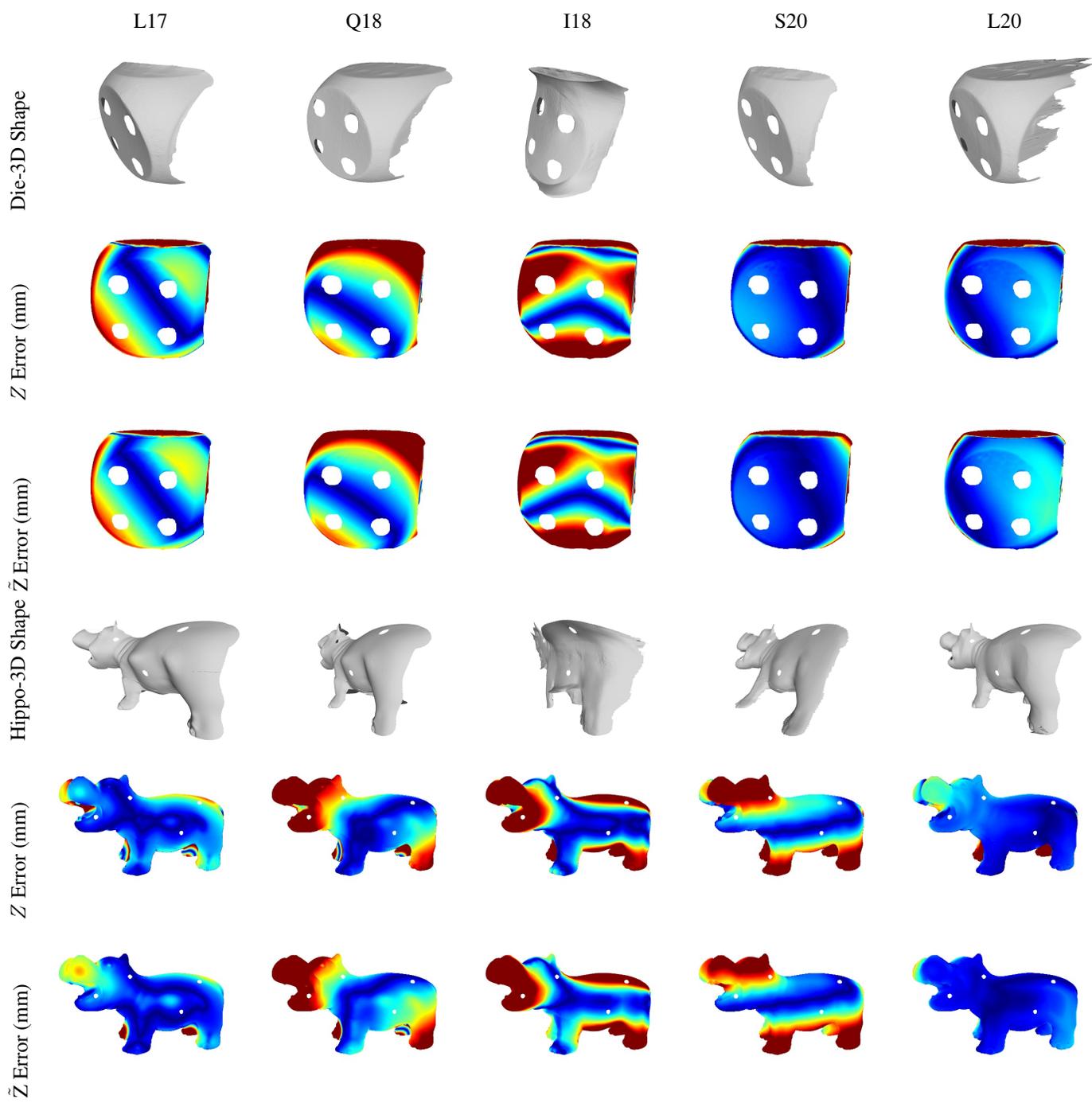


Figure 12. Evaluations 5-6/14

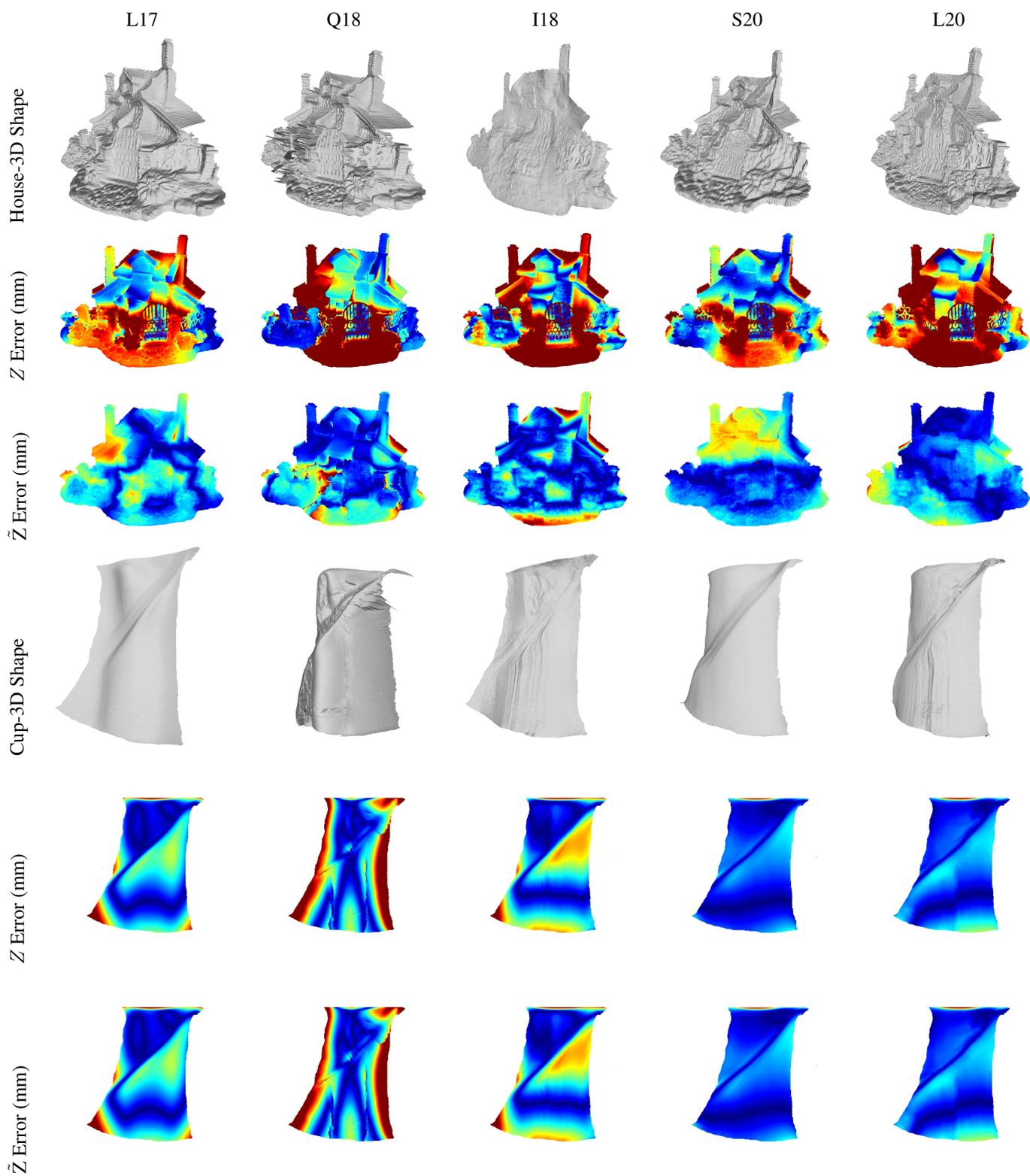


Figure 13. Evaluations 7-8/14

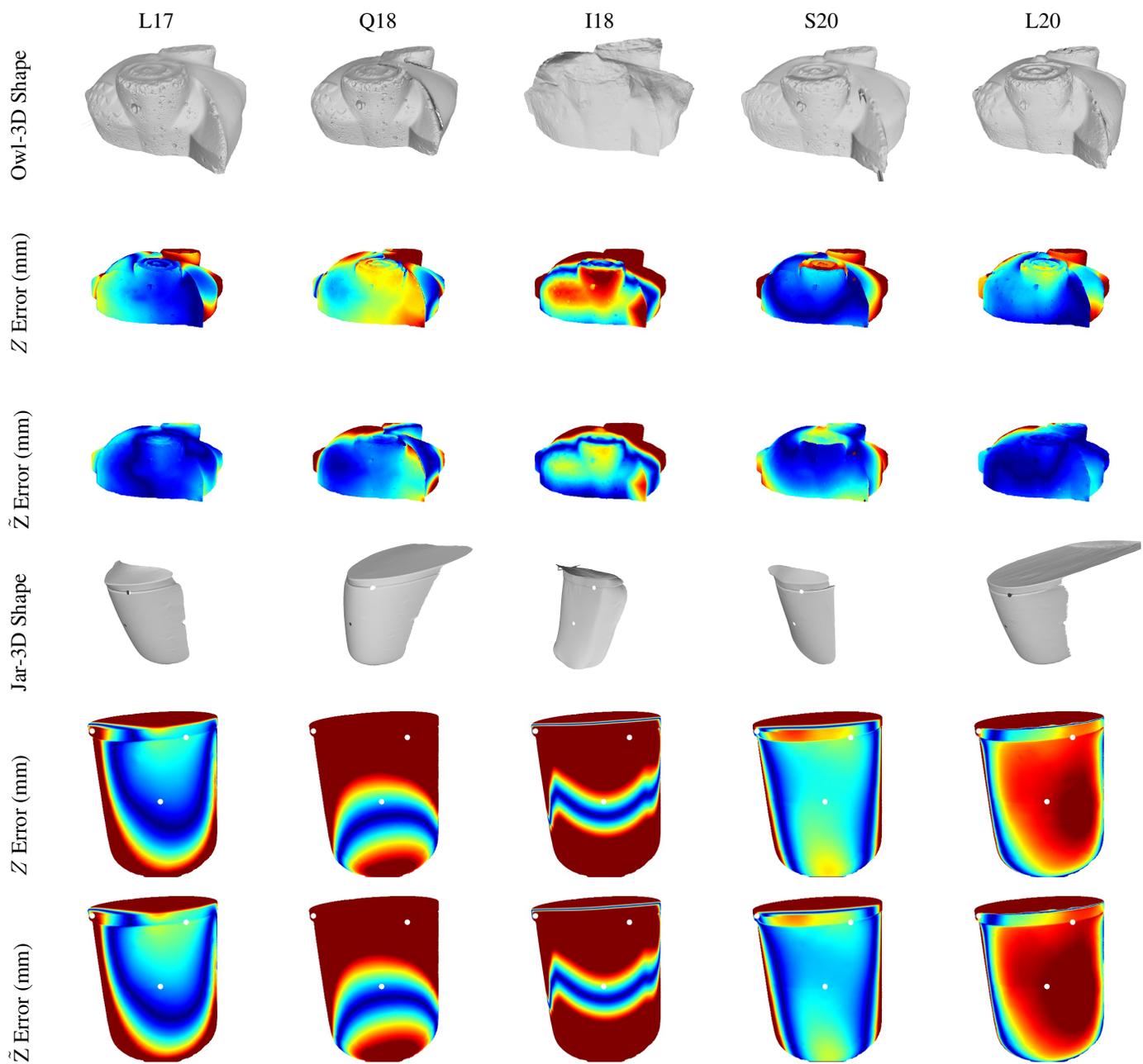


Figure 14. Evaluations 9-10/14

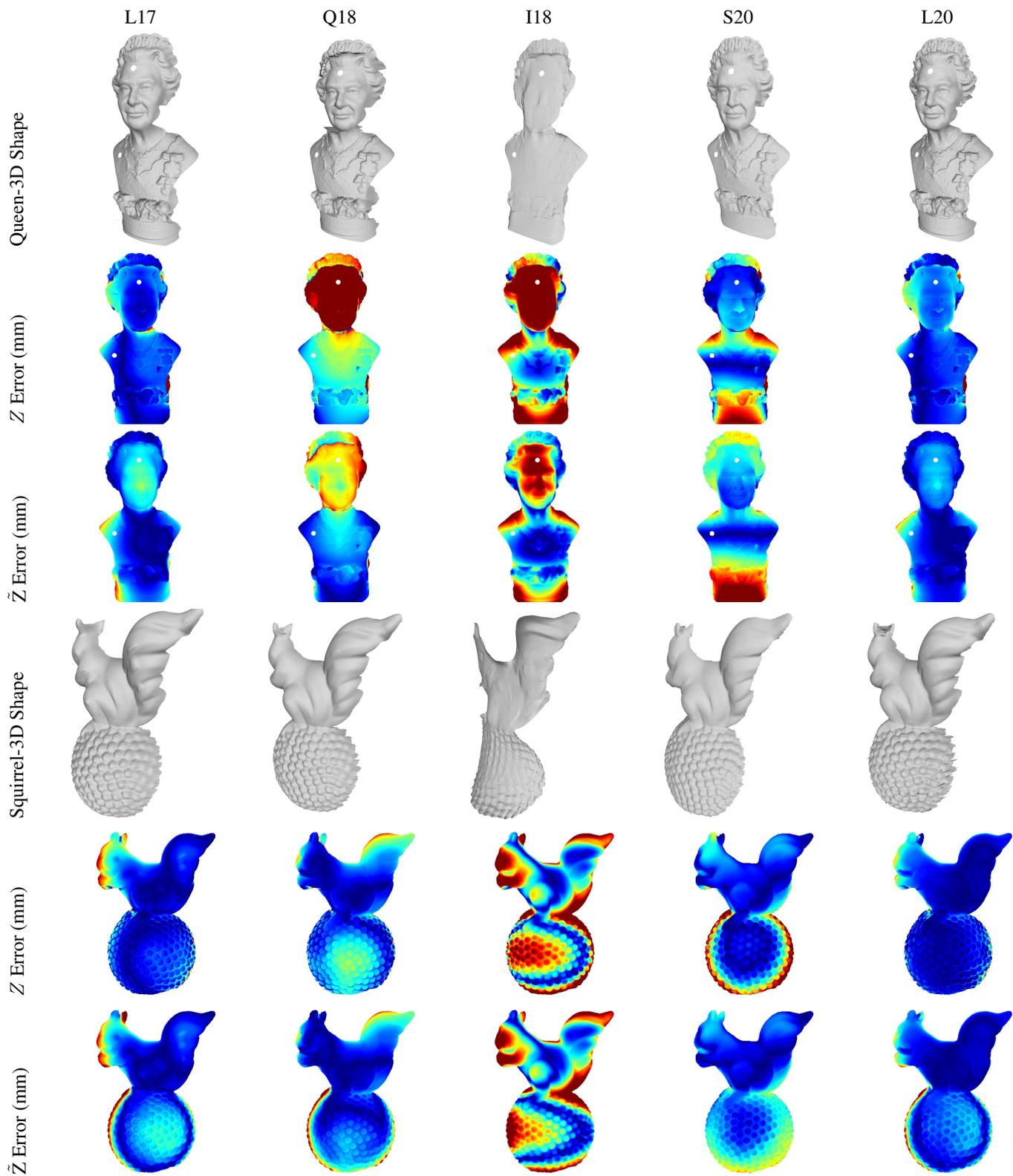


Figure 15. Evaluations 11-12/14

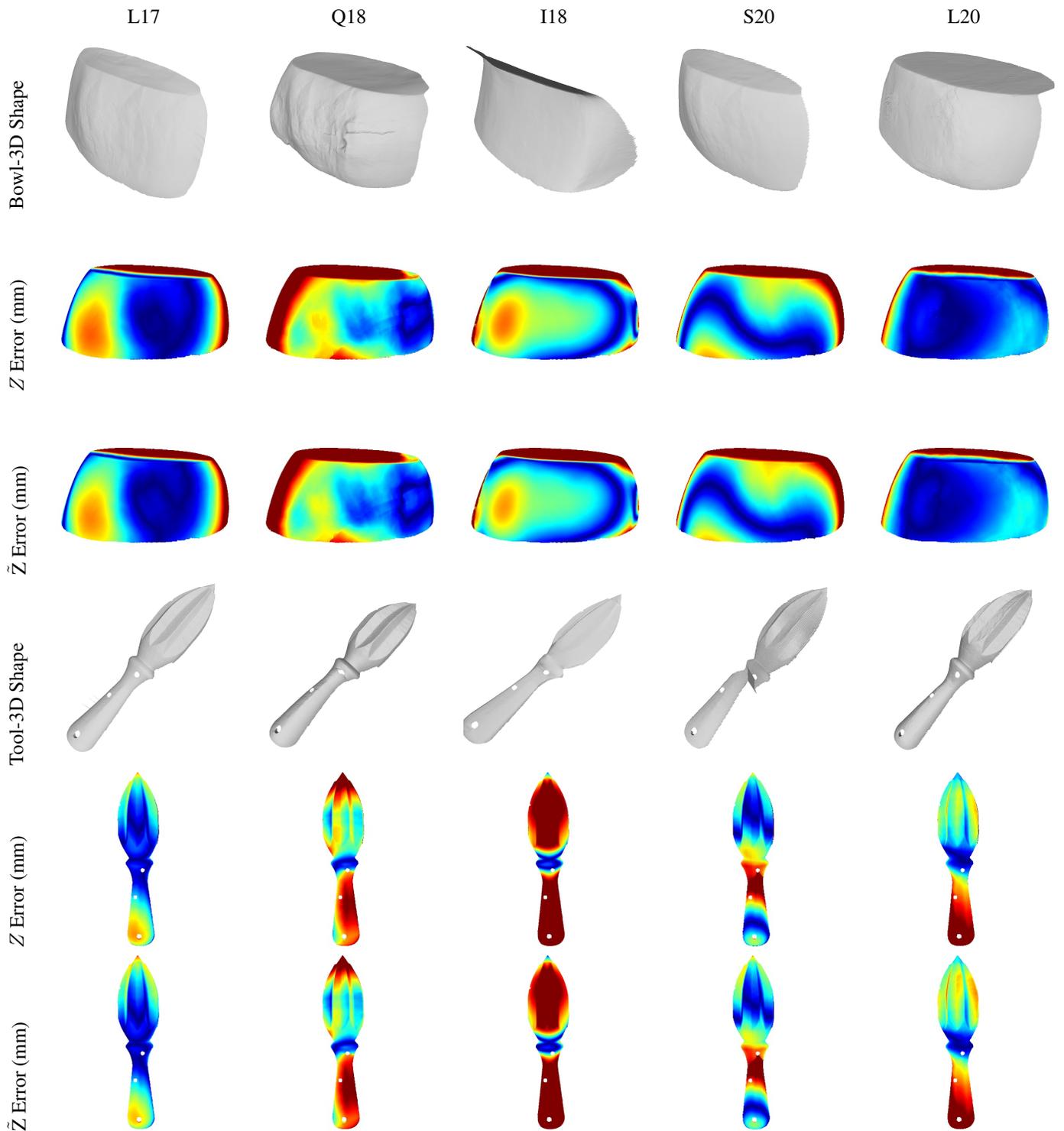


Figure 16. Evaluations 13-14/14