Variational Graph-based Normal Integration

Supplementary Material

1. Normal Map Formation: Derivation

Let a depth map indexed by image coordinates as $z(h, g) = e^{d(h,g)}$. The total derivative of z in terms of h at image location u, v yields:

$$\frac{\partial z}{\partial h} = z_h = z_x x_h + z_y y_h, \tag{1}$$

where

$$z_h = z d_h, \tag{2}$$
$$z_x = -\frac{n_x}{n_z}, \tag{3}$$
$$z_y = -\frac{n_y}{n_z}, \tag{4}$$
$$x = \frac{1}{f} z u, \tag{5}$$
$$y = \frac{1}{f} z v, \tag{6}$$
$$u = h \cos\theta_h + g \cos\theta_g, \tag{7}$$
$$x_h = \frac{z}{f}(d_h u + \cos\theta_h), \tag{8}$$
$$y_h = \frac{z}{f}(d_h v + \sin\theta_h). \tag{9}$$

This can be reduced to:

$$f d_h = (z_x u + z_y v) d_h + z_x \cos\theta_h + z_y sin\theta_h. \tag{10}$$

In addition, because

$$u_x = \frac{f}{z}(1 - \frac{x}{z} z_x) = \frac{f}{z}(1 - \frac{u}{f} z_x), \tag{11}$$

the image gradient and depth gradient becomes equivalent (the orthographic case) when this term becomes 1, and this happens when f is sufficiently large, where the FOV narrows so much that only the rays near the principal axis are imaged (u and v are sufficiently small). Moreover, as $\frac{u}{f} \to 0$, z has to be commensurably large so that $u_x \to \frac{f}{z} \to 1$. This also justifies the scaling factor $f - z_x u - z_y v$ becomes a constant f.

2. Gradient Flow and Graph-based Diffusion

We show that the proposed graph-based normal integration inherently performs depth-discontinuity-aware anisotropic diffusion on a graph. It is worth noting that underneath each sampled triplet configuration there is a two-component softmax weighting assignment (*i.e.* sigmoid function), and the Jacobian of the softmax function governs how information propagates from a node to its neighbors, which implements an anisotropic diffusion on a graph. The softmax function:

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \tag{12}$$

maps a vector of logits $\vec{z} \in \mathcal{R}^n$ to a probability distribution over n outcomes. Its Jacobian matrix $J_\sigma(z) = \frac{\partial \sigma}{\partial z}$ is given by

$$J_\sigma(\vec{z}) = \text{diag}(\vec{\sigma}) - \vec{\sigma}\vec{\sigma}^\intercal, \tag{13}$$

where $\vec{\sigma} = \sigma(\vec{z})$. Each off-diagonal entry ($J_\sigma = -\sigma_i$) represents how increasing z_j decreases σ_i, while the diagonal term $J_{\sigma\sigma} = \sigma_i(1 - \sigma_i)$ reflects how sensitive each probability is to its own input. The structure of this matrix behaves similarly to a graph Laplacian over the probability simplex, encoding conservative, diffusion-like coupling among the components of $\vec{\sigma}$. In short, this softmax-induced Laplacian determines how small variation in logits diffuse across output probabilities.

Analogously, anisotropic diffusion on a graph with node feature $\vec{u} \in \mathcal{R}^n$ evolves according to

$$\frac{\partial \vec{u}}{\partial t} = -L(\vec{u})\vec{u}, \tag{14}$$

where $L(\vec{u})$ is a nonlinear, data dependent Laplacian defined by edge weights $w(u_i - u_j)$ for some function g. This process smooths $\vec{u}$ while preserving important discontinuities , adapting to the local structure of the data. J_σ defines a nonlinear, adaptive diffusion operator, where the "anisotropic diffusion conductance" between component i and j is proportional to $\sigma_i \sigma_j$.

Moreover, when we apply a small perturbation $\vec{v}$ to $\vec{z}$, the Jacobian-vector product (JVP) gives the resulting change to the softmax output:

$$\delta\vec{\sigma} = J_\sigma(\vec{z})\vec{v}, \tag{15}$$

which yields

$$\delta\vec{\sigma} = \vec{\sigma} \cdot (\vec{v} - \mathbf{1}(\vec{\sigma^\intercal v})), \tag{16}$$

where $\cdot$ denotes entry-wise multiplication and $\mathbf{1}$ is a vector of ones. It shows that $\vec{v}$ is projected onto the tangent space of the probability simplex, and then rescaled anisotropically by $\vec{\sigma}$. Furthermore, because

$$\frac{\partial \vec{\sigma}}{\partial t} = J_\sigma(\vec{z}) \frac{\partial \vec{v}}{\partial t}, \tag{17}$$

$J_\sigma(\vec{z})$ acts as a state-dependent diffusion tensor that couples the rate of change of logits into a smooth but anisotropic flow over the probability simplex.

3. Diligent under Perspective Projection

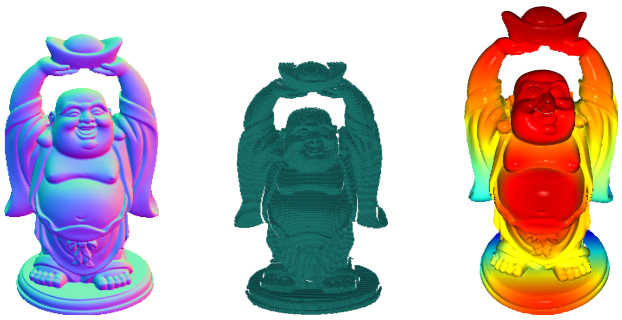
Following the derivation in Section 1, from Figure 1 to 8, we visualize the reconstruction results produced by the proposed graph-based normal integrator for objects in DiLiGent dataset [2] captured under perspective projections. In this work, we only demonstrate the effectiveness of our solution, while technical analysis are left as part of our future work.

(a) normal. (b) point cloud. (c) mesh.

Figure 1. bear under perspective projection.

(a) normal. (b) point cloud. (c) mesh.

Figure 2. cat under perspective projection.



(a) normal. (b) point cloud. (c) mesh.

Figure 3. buddha under perspective projection.

4. Diligent of Grid Representations

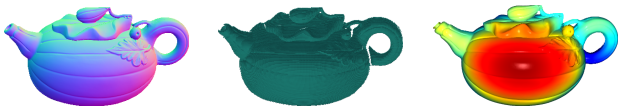
Results summarized in Table 1 are visualized in Figure 16 and Results summarized in Table 2 are visualized in Figure 17, respectively. It can be observed from both figures that our approach outperforms BiNI [1] over several objects by better estimating object parts that are mainly detached from the main body caused by depth discontinuities.

(a) normal. (b) point cloud. (c) mesh.

Figure 4. cow under perspective projection.

(a) normal. (b) point cloud. (c) mesh.

Figure 5. goblet under perspective projection.

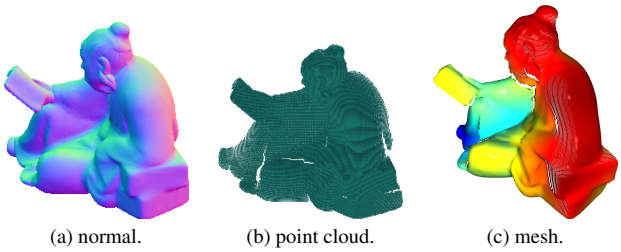


(a) normal. (b) point cloud. (c) mesh.

Figure 6. pot1 under perspective projection.

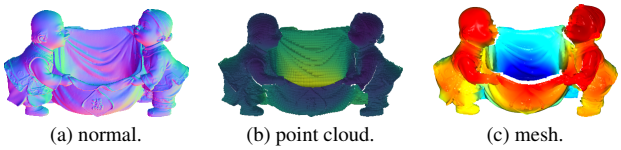
(a) normal. (b) point cloud. (c) mesh.

Figure 7. pot2 under perspective projection.



(a) normal. (b) point cloud. (c) mesh.

Figure 8. reading under perspective projection.

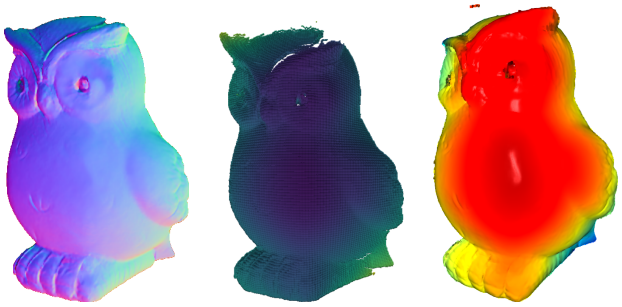


(a) normal. (b) point cloud. (c) mesh.

Figure 9. harvest under orthographic projection.

(a) normal. (b) point cloud. (c) mesh.

Figure 10. human under orthographic projection.



(a) normal. (b) point cloud. (c) mesh.

Figure 11. owl under orthographic projection.

(a) normal. (b) point cloud. (c) mesh.

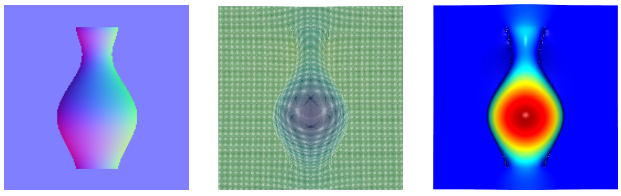
Figure 12. reading under orthographic projection.

(a) normal. (b) point cloud. (c) mesh.

Figure 13. thinker under orthographic projection.

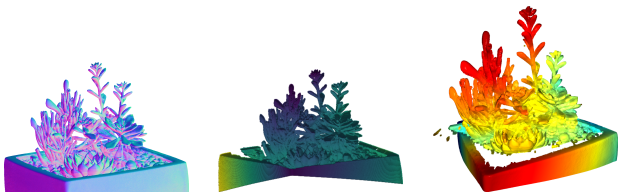
5. Diligent of Scattered Representations

Results summarized in Table 3 are visualized in Figure 18 and Results summarized in Table 4 are visualized in Figure 19, respectively. It can be observed from both figures that our graph-based solution successfully reconstructs object surfaces from the corresponding highly sparse inputs



(a) normal. (b) point cloud. (c) mesh.

Figure 14. vase under orthographic projection.



(a) normal. (b) point cloud. (c) mesh.

Figure 15. plant under orthographic projection.

of oriented points with unstructured distributions. And by comparing the results generated with and without triplet sampling introduced in Section 4.3, it is evident that the introduced sampling strategies effectively detect and preserve the regions surrounded by depth discontinuities. Notably, this performance is consistently aligned with the performance delivered by our solution processing dense grid-based input, verifying both efficacy and effectiveness of our solution.

6. More Objects under Orthographic Projection

Figure 9 to 15 visualize the reconstructions of some other objects whose normal maps are publicly accessible. In general, our method can process various types of complex surface geometry robustly.

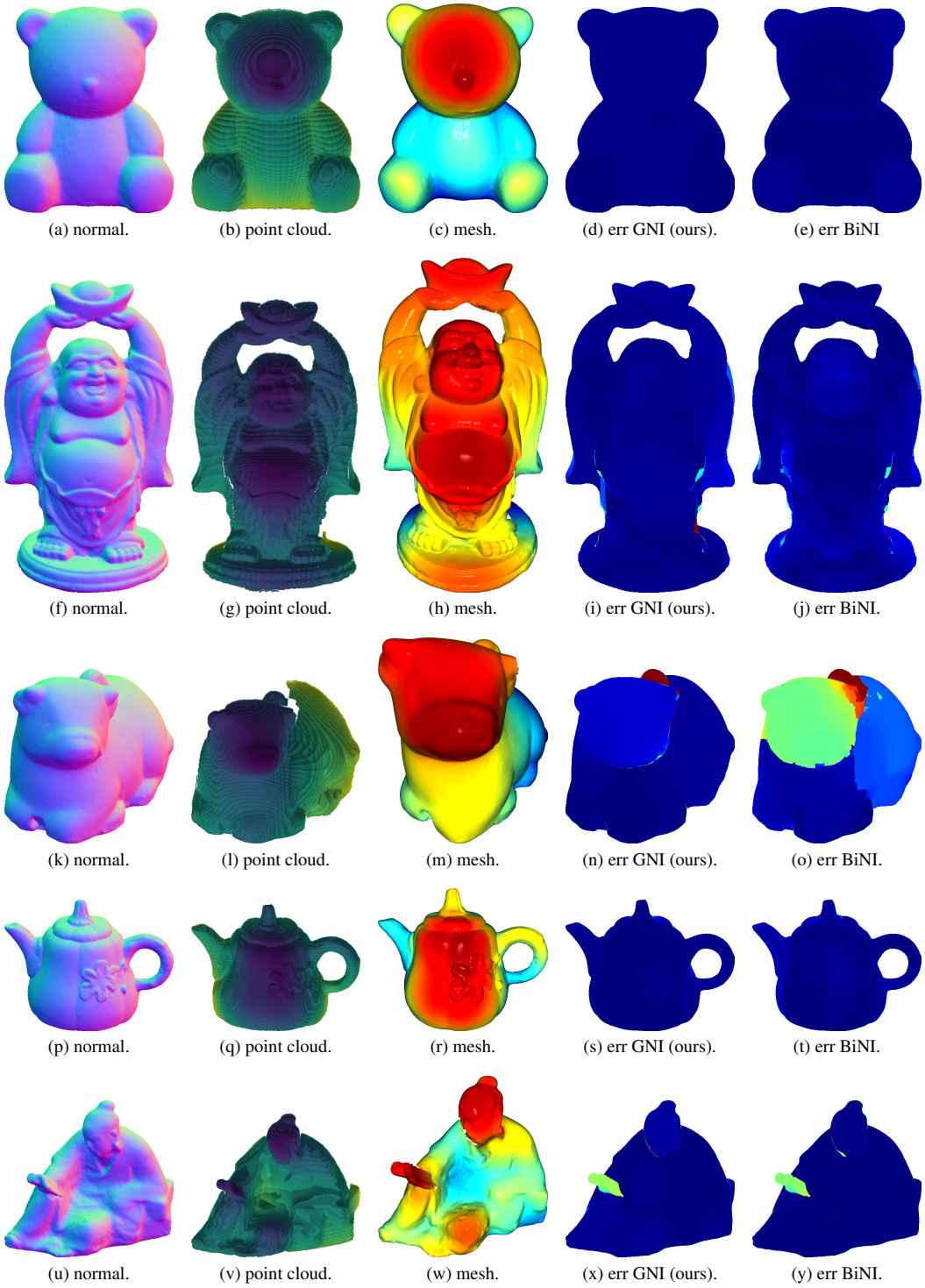


Figure 16. From top to bottom: the grid representations of bear, buddha, cow, pot, reading captured from the front. Statistics are summarized in Table 1.

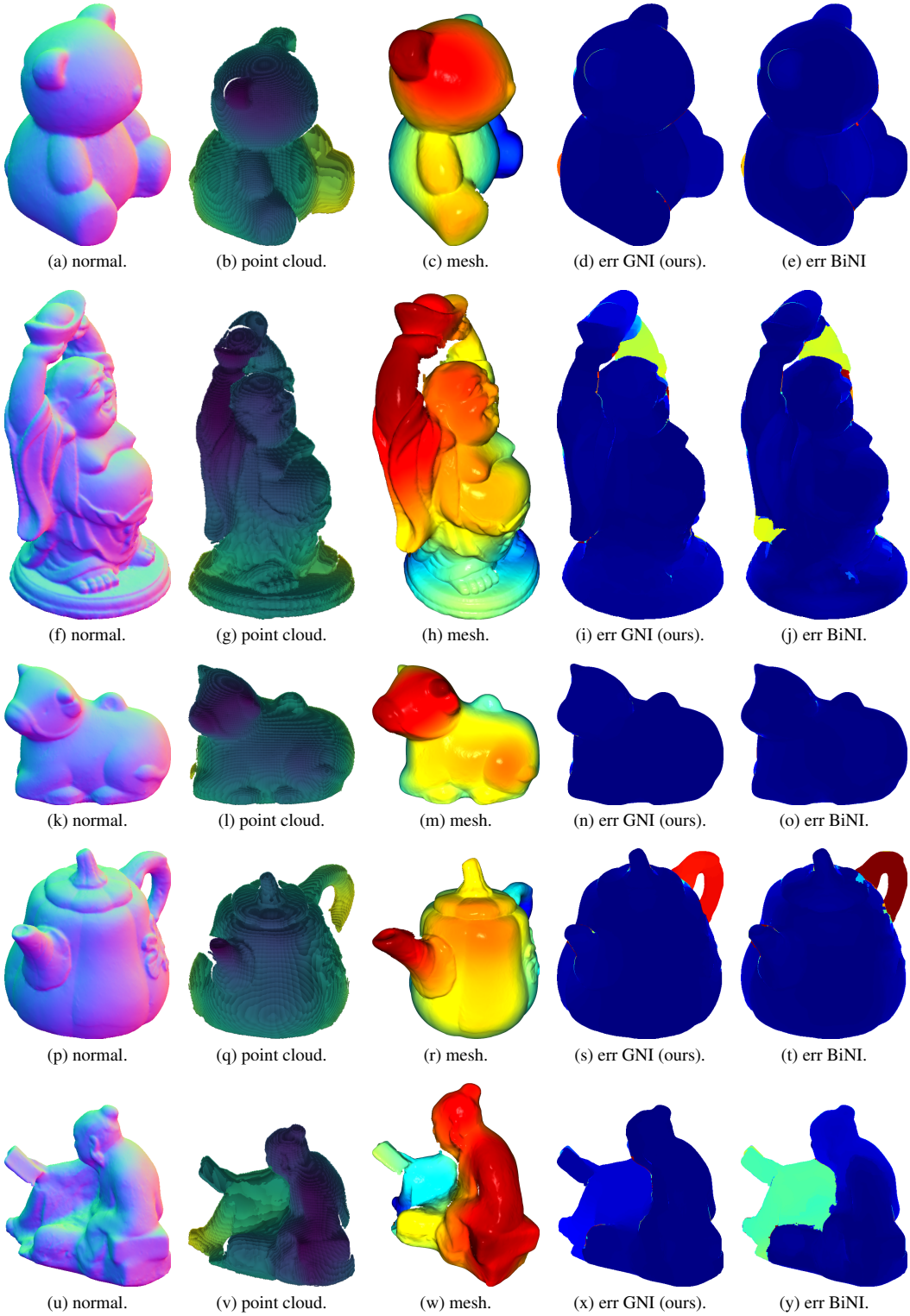


Figure 17. From top to bottom: the grid representations of bear, buddha, cow, pot, reading captured from the side. Statistics are summarized in Table 2.

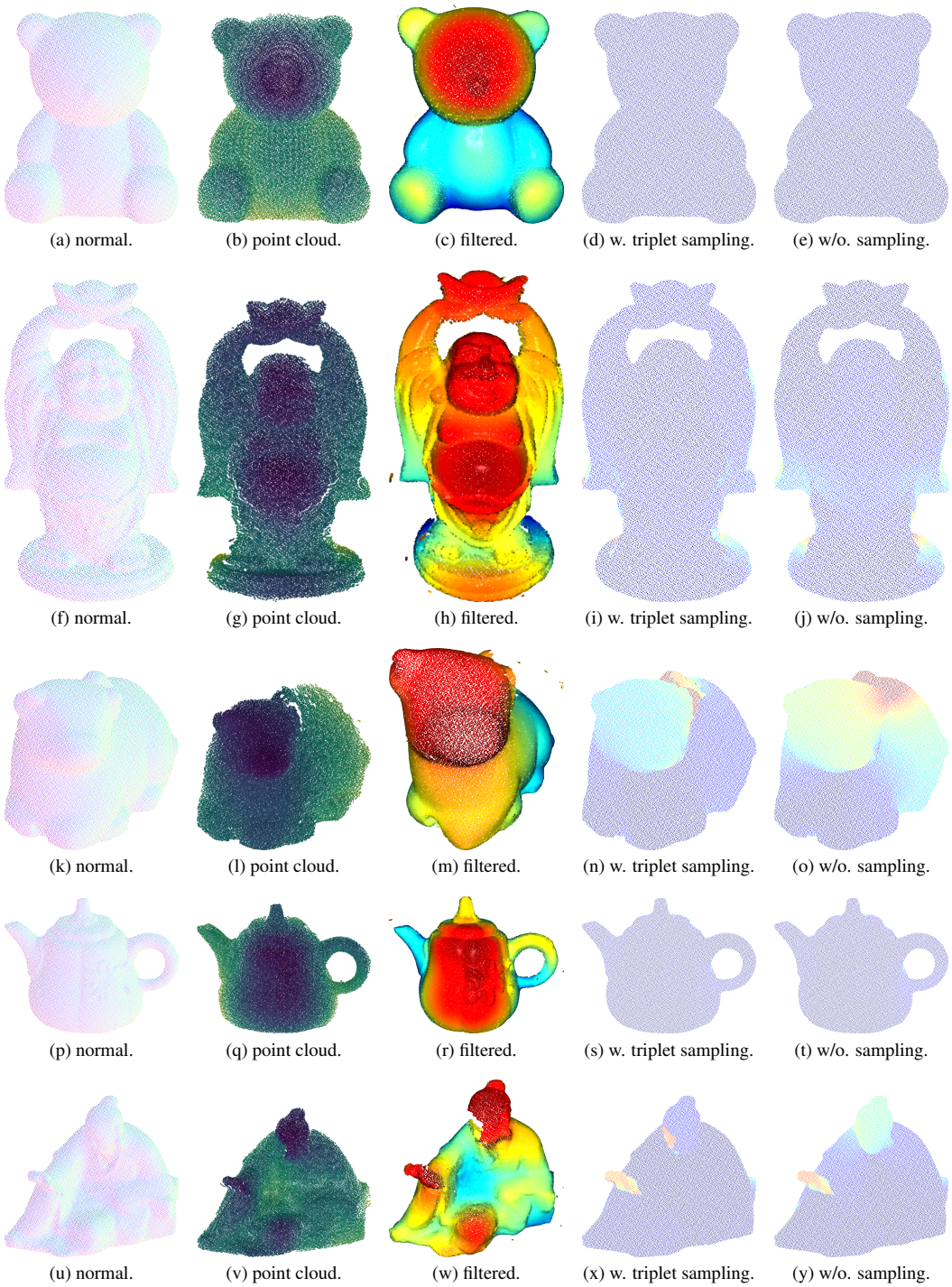


Figure 18. From top to bottom: the scattered representations of bear, buddha, cow, pot, reading captured from the front. Statistics are summarized in Table 3.

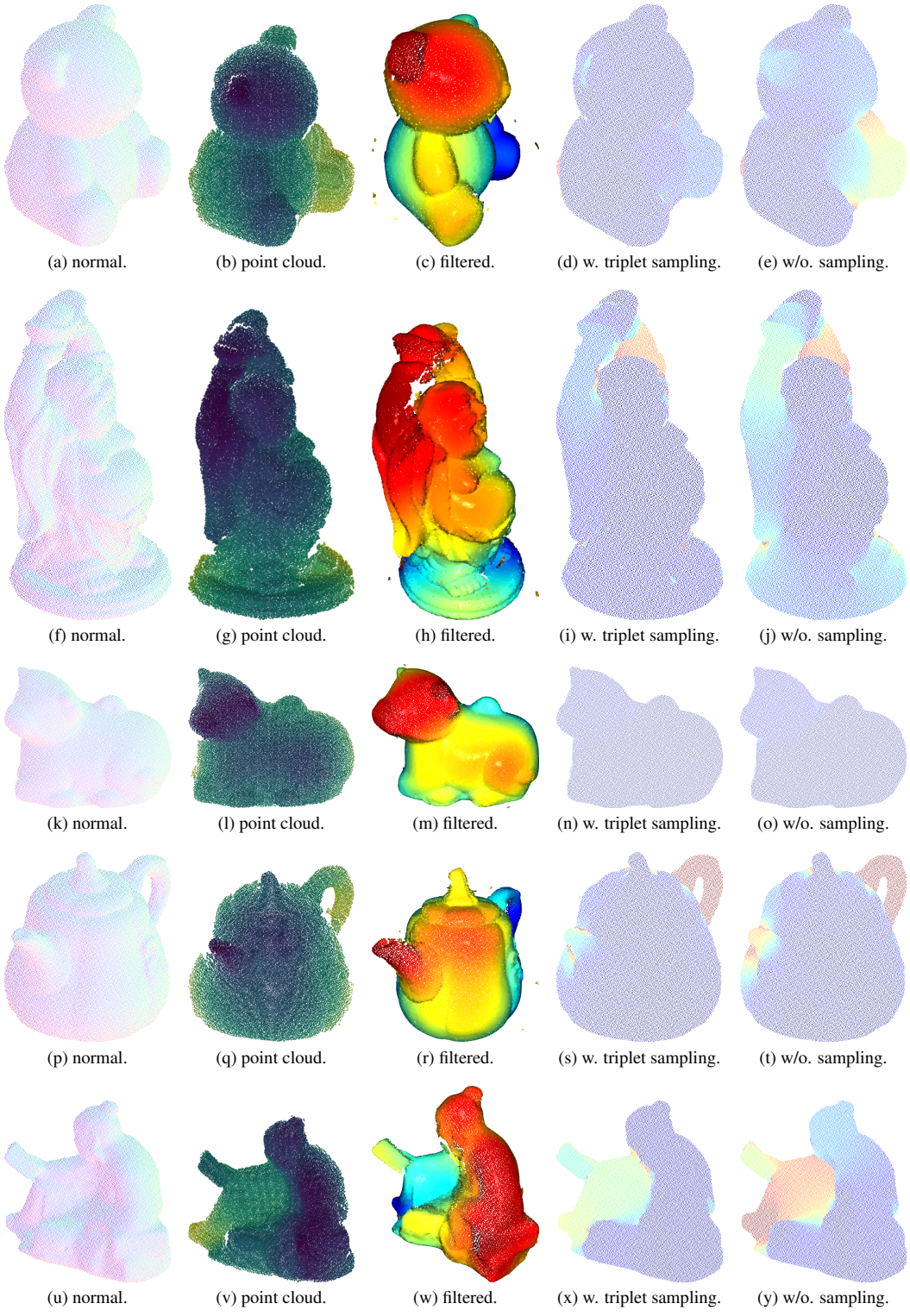


Figure 19. From top to bottom: the scattered representations of bear, buddha, cow, pot, reading captured from the side. Statistics are summarized in Table 4.

References

[1] Xu Cao, Hiroaki Santo, Boxin Shi, Fumio Okura, and Yasuyuki Matsushita. Bilateral normal integration. In *European Conference on Computer Vision*, 2022. 2

[2] Boxin Shi, Zhe Wu, Zhipeng Mo, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016. 2