# DiLiGenRT: A Photometric Stereo Dataset with Quantified Roughness and Translucency

Heng Guo[1,†] Jieji Ren[2,†] Feishi Wang[3,4,5†] Boxin Shi[3,4,5*] Mingjun Ren[2*] Yasuyuki Matsushita[6]

[1] School of Artificial Intelligence, Beijing University of Posts and Telecommunications

[2] School of Mechanical Engineering, Shanghai Jiao Tong University

[3] National Key Laboratory for Multimedia Information Processing, School of Computer Science, Peking University

[4] National Engineering Research Center of Visual Technology, School of Computer Science, Peking University

[5] AI Innovation Center, School of Computer Science, Peking University

[6] Graduate School of Information Science and Technology, Osaka University

## Abstract

*Photometric stereo faces challenges from non-Lambertian reflectance in real-world scenarios. Systematically measuring the reliability of photometric stereo methods in handling such complex reflectance necessitates a real-world dataset with quantitatively controlled reflectances. This paper introduces DiLiGenRT, the first real-world dataset for evaluating photometric stereo methods under quantified reflectances by manufacturing* 54 *hemispheres with varying degrees of two reflectance properties:* ***Roughness*** *and* ***Transluency****. Unlike qualitative and semantic labels, such as "diffuse" and "specular," that have been used in previous datasets, our quantified dataset allows comprehensive and systematic benchmark evaluations. In addition, it facilitates selecting best-fit photometric stereo methods based on the quantitative reflectance properties. Our dataset and benchmark results are available at* https://photometricstereo.github.io/diligentrt.html.

## 1. Introduction

Photometric stereo [53] aims at recovering detailed surface normals and reflectances from images captured under varying illuminations from a fixed camera. To measure the applicability of photometric stereo methods in the real world, comprehensive benchmark evaluations on real captured datasets are desired. General reflectance is one of the key challenges for photometric stereo, making it a central consideration when building real-world photometric stereo datasets. By evaluating a photometric stereo method on a real-world dataset, we seek to know the exact reflectance conditions where the method would become trustworthy. Achieving this necessitates a dedicated design of the real-world dataset with quantitatively controlled reflectances.

Previous datasets for photometric stereo are constructed with reflectances defined in a *qualitative* and *semantic* manner. Datasets like DiLiGenT [48], LUCES [36, 40], DiLiGenT10$^2$ [44], and DiLiGenT-II [52] categorize reflectances using a set of semantic labels, such as `specular`, `metallic`, `steel`, and `wood`. These semantic labels are intuitive but at the same time ambiguous. For example, a `steel` surface can exhibit a rough or mirror-like reflectance depending on the manufacturing process. Furthermore, these semantic labels are not explicitly ordered and non-metric; therefore, it is difficult to predict the accuracy of photometric stereo for "in-between" reflectances, *e.g.*, predicting accuracy for the `paper` reflectance from the evaluations based on `steel` and `wood` reflectances. In addition, existing real-world datasets [34, 36, 44, 48, 52] primarily focus on opaque surfaces, like plaster and ceramic. Since a wide range of real-world objects exhibit significant translucency, such as human skin and plant leaves, it is desired to have a dataset with thorough coverage of varying levels of translucency.

To address these problems, we build a new real-world dataset, DiLiGenRT[1], with reflectances controlled in a *quantified* manner by carefully manufacturing materials. DiLiGenRT consists of **54** hemispheres made of materials having **9** levels of roughness and **6** levels of translucency[2], as shown in Fig. 1. The design of the shared geometry allows us to isolate the error source from shape variations and solely attribute normal estimation errors to varying reflectances. At the same time, for the test purpose on shape diversity, we fur-

* labels corresponding authors (Email address: shiboxin@pku.edu.cn, renmj@sjtu.edu.cn). † denotes equally contributed authors.

[1] DiLiGenRT = Directional Lighting with Generalized Roughness and Transluency

[2] Transluency is related to object shape, index of refraction (IOR), and absorptive ratio. DiLiGenRT is built with the same sphere shape and IOR. We refer to "absorptive ratio" as "translucency" in this paper.

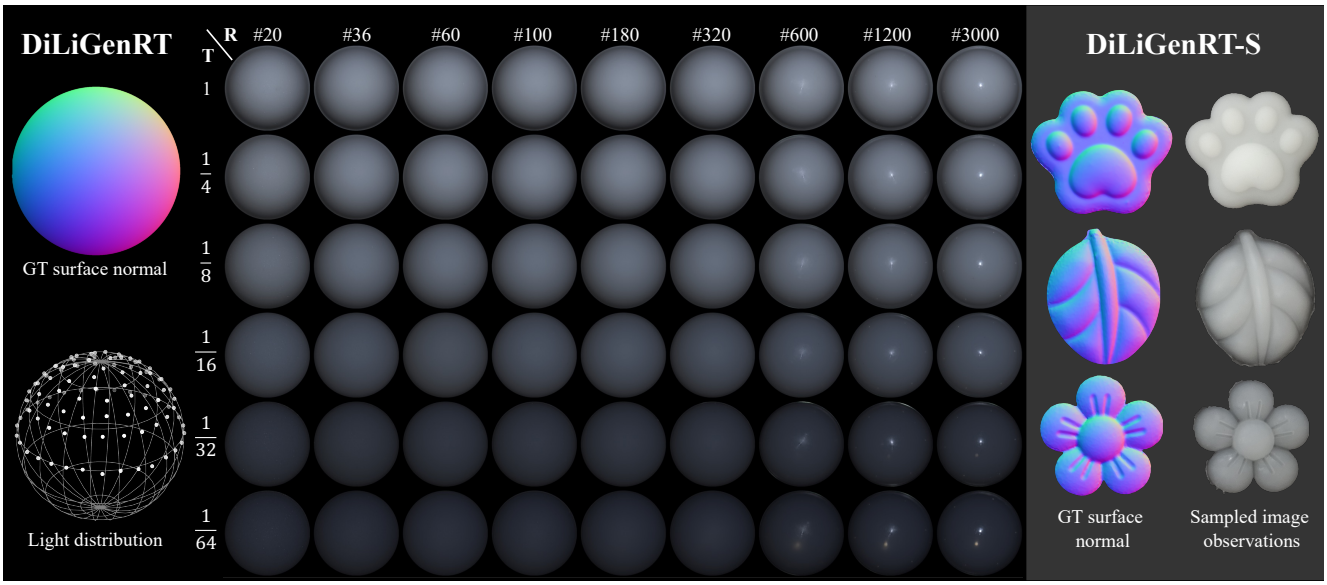



Figure 1. Overview of DiLiGenRT represented as a **translucency-roughness** ($6 \times 9$) matrix and DiLiGenRT-S including 3 distinct shapes. (**Left**) All objects in DiLiGenRT share the same ground truth sphere surface normal and are illuminated by 100 uniformly distributed directional lights. (**Middle**) Image observations of DiLiGenRT under varying degrees of roughness (R) and translucency (T) indexed by the solution concentration and grit size, respectively. The larger the solution concentration or grit size, the lesser the roughness or translucency. From top-left to bottom-right, the surface becomes shinier as roughness decreases, while the image appearance grows darker as translucency increases. (**Right**) The ground truth surface normals and selected image samples are included in DiLiGenRT-S.

ther introduce DiLiGenRT-S, which contains three distinct Shapes, as depicted in Fig. 1 (right).

Using the DiLiGenRT dataset, we conduct benchmark evaluations to provide each photometric stereo method a performance profile matrix under various roughness and translucency levels. As reflectance properties are quantifiable, continuous, and common to various materials, it enables a comprehensive analysis of photometric stereo's working range w.r.t. reflectance, and facilitates selecting best-fit photometric stereo methods based on the reflectance properties. In addition, it offers the capability to infer the performance of photometric stereo methods for materials that are not directly covered in the dataset based on the numerical interpolation of reflectance properties.

**Contributions** In this paper, we present DiLiGenRT, the first real-world photometric stereo dataset manufactured with quantified roughness and translucency. Based on DiLiGenRT, we provide a benchmark result presented as performance profiles of photometric stereo methods, showing their working ranges over quantitative reflectances.

## 2. Related works

This paper's main goal is to provide a benchmark dataset for photometric stereo. We begin by reviewing photometric stereo methods categorized by their reflectance assumptions. After that, we summarize existing synthetic and real-world datasets for photometric stereo.

### 2.1. Photometric stereo methods

We briefly summarize photometric stereo methods based on their reflectance assumptions: Lambertian and more general Bidirectional Reflectance Distribution Functions (BRDFs) for opaque surfaces, and subsurface scattering for translucent surfaces.

**Lambertian reflectance** The Lambertian reflectance model is an ideal diffuse reflectance model that has been extensively used in the literature because of its simplicity and capability of representing diffuse reflectances. Lambertian photometric stereo is well-posed with calibrated distant lights and can be solved in a closed form given three or more images under different light directions [49, 53]. In an uncalibrated light setting, surface normal can be recovered up to a linear ambiguity [17]. With the surface integrability [19], the linear ambiguity can be reduced to a General Bas-Relief (GBR) ambiguity [3]. Existing uncalibrated photometric stereo methods resolve this GBR ambiguity by clustering uniform albedo areas [46], considering specific light distributions (*e.g.*, symmetric light [41], ring light [60], differential light [5]) or perspective camera projections [42]. Papadhimitri *et al.* [43] proposes to detect local diffuse reflectance maxima from image observations, which infers surface normals coinciding with the light direction.

**General BRDFs for opaque surfaces** Photometric stereo for surfaces with general BRDFs is more challenging than

Table 1. Summary of real (top) and synthetic (bottom) photometric stereo datasets organized by number (#) of distinct translucent materials, surface normal maps, lights, and sets (one set means a sequence of photometric stereo images under varying lights).

| Dataset | Reflectance control | # Transluency | # Normal | # Lights | # Sets |
|---|---|---|---|---|---|
| DiLiGenT [48] | Semantic labels | 0 | 10 | 96 | 10 |
| DiLiGenT-MV [34] | Semantic labels | 0 | 100 | 100 | 100 |
| LUCES [36, 40] | Semantic labels | 0 | 14 | 52 | 14 |
| Harvard [56] | Semantic labels | 0 | 7 | 20 | 7 |
| ETHz [31] | Semantic labels | 0 | 3 | 260 | 3 |
| Gourd&Apple [2] | Semantic labels | 0 | 2 | 102/112 | 2 |
| DiLiGenT-II [52] | Semantic labels | 1 | 30 | 100 | 30 |
| DiLiGenT10$^2$ [44] | Semantic labels | 1 | 10 | 100 | 100 |
| **DiLiGenRT** | Quantified properties | 6 | 1 | 100 | 54 |
| **DiLiGenRT-S** | Quantified properties | 3 | 3 | 100 | 3 |
| BlobbyPS [45] | Semantic labels | 0 | 8 | 96 | 800 |
| PS-Sculpture [6] | Semantic labels | 0 | 59,292 | 64 | 59,292 |
| CyclesPS [21] | Quantified properties | 0 | 15 | 1,300 | 45 |
| SymPS [51] | Quantified properties | 0 | 5 | 77 | 9 |
| PS-Relief [52] | Quantified properties | 0 | 127 | 100 | 3,429 |
| PS-Wild [23] | Quantified properties | 0 | 410 | 31 | 10,099 |
| PS-Mix [24] | Quantified properties | 0 | 410 | 10 | 34,921 |

the ideal Lambertian case. Under calibrated settings, existing optimization-based methods assume dominant diffuse reflectance and treat specular highlights in BRDF as sparse outliers, developing robust photometric stereo approaches based on sparse Bayesian regression [25], low-rank matrix completion [54], and positional thresholding strategies [48]. Other methods explicitly represent general BRDFs via bi-polynomial approximation [47], Microfacet reflectance models [9, 14], or linear bases [12, 13, 20].

Since the work of Santo *et al*. [45], learning-based photometric stereo methods have achieved significant progress in handling general BRDFs by finding a direct mapping between input images to the surface normal map. The proposed network structures can be divided into the all-pixel branch [6, 30] and the per-pixel branch [21, 33, 35, 59]. Follow-up works further improve photometric stereo by addressing global illumination effects [16, 35], reducing the number of inputs [33, 50, 51, 59], and combining the merits of per-pixel-based and all-pixel-based methods [22, 57].

Beginning from SDPS-Net [7], learning-based uncalibrated photometric stereo methods under general BRDFs are proposed. Follow-up works focus on analyzing the feature map of deep uncalibrated photometric stereo [8] or applying an inverse rendering module to solve uncalibrated photometric stereo in a self-supervised manner [31, 32]. The most recent work, UniPS [23] and SDM-UniPS [24], introduce neural global light context to recover surface normals under unknown, spatially-varying illumination, which is shown to be effective for surfaces with general BRDFs.

**Subsurface scattering** is non-local light transport beneath a surface, which is observed on translucent surfaces. Due to the complexity of subsurface scattering, only a few photometric stereo methods have been developed for translucent surfaces. Inoshita *et al*. [26] observe that subsurface scattering in optically thick translucent objects can be modeled as a convolution with a blur kernel, thus recasting photometric stereo as a deconvolution problem. Similarly, Dong *et al*. [11] frame photometric stereo with subsurface scattering as a blind deconvolution problem, jointly estimating subsurface scattering parameters and surface normals through an inverse rendering based on the dipole-diffusion model [28]. However, these methods either assume a restrictive Lambertian reflectance [26] or near-planar surface shapes [11], limiting their application scope.

## 2.2. Photometric stereo datasets

Table 1 lists existing synthetic and real-world photometric stereo datasets. Here we discuss their properties in reflectance, shape, and illumination in detail.

**Synthetic dataset** utilizes physics-based rendering engines, like Mitsuba [27] and Blender [10], to produce image observations and corresponding surface normal maps for diverse synthetic scenes. During the rendering, factors like surface normal, illumination, and material are completely controlled. For example, the BlobbyPS dataset [45] contains 8 smooth Blobby shapes [29] with reflectances of measured MERL BRDF dataset [38]. PS-Sculpture [6] extends BlobbyPS by adding $59,292$ diverse surface normals. Both datasets control the reflectance via semantic labels of the MERL dataset [38], such as `blue-paper` and `alum-bronze`. CyclesPS [21] controls continuous parameters of reflectance properties (*e.g*., roughness, metallic) based on the Disney Principled BSDF model [4]. Unlike semantic labels, reflectance properties can be quantitatively assigned, making it possible to create a large number of unique materials (*e.g*., $30,000$ materials are generated in CyclesPS [21]). The follow-up datasets, such as SymPS [51], PS-Wild [23] and PS-Mix [24] build upon CyclesPS [21], increase the dataset scale by including more diverse shapes, materials, and illuminations.

All these synthetic datasets assume an opaque surface without. In reality, translucent surfaces are relevant everywhere, such as human skin and plant leaves; thus, it is desired to have a dataset with controlled translucency.

**Real-world dataset** Real-world datasets complement synthetic ones by bridging the gap between computer graphics rendering and actual image formation. For instance, the Gourd&Apple dataset [1] offers image observations of two objects with spatially-varying isotropic BRDFs. The Harvard dataset [55] includes 7 surfaces with a uniform diffuse reflectance. However, these two datasets do not provide the ground truth surface normals. The DiLiGenT [48] dataset includes 10 objects of different shapes and general reflectances. It introduces benchmark evaluation of photometric stereo for the first time based on the ground truth surface normals from scanned meshes. Subsequent datasets further expand DiLiGenT [48], introducing aspects

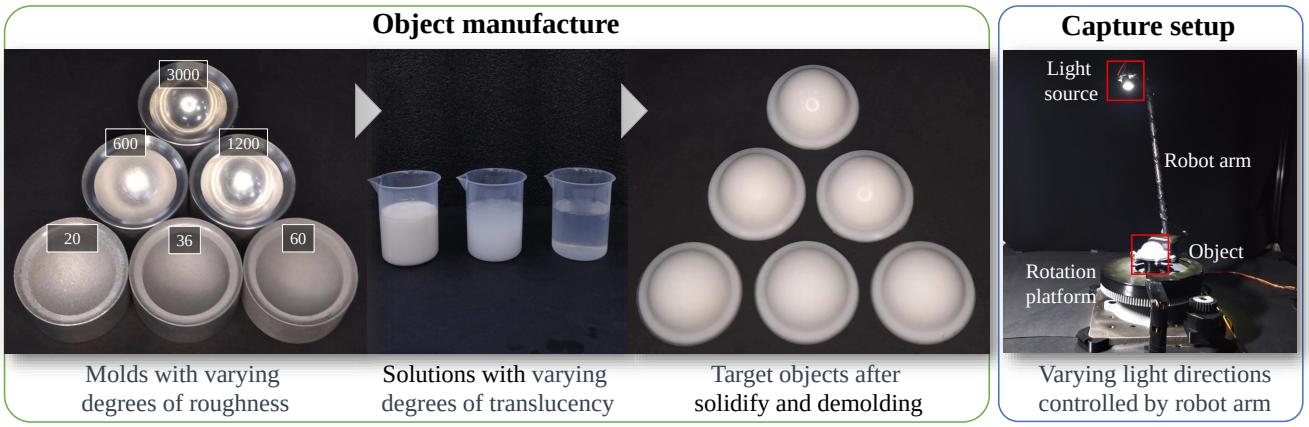



Figure 2. (**Left**) Objects in DiLiGenRT are constructed by first manufacturing molds with varying degrees of roughness through sandblasting and then injecting solutions of different concentrations into the molds, followed by solidifying and de-molding. (**Right**) Images in DiLiGenRT are captured by moving a point light source bundled on a robot arm.

like multi-views (DiLiGenT [48]-MV [34]), near-field illumination (LUCES [40]), environment illumination [15, 18], planar surfaces with rich details (DiLiGenT-Π [52]), and global illumination effects [31].

Closest to our work, to disentangle the error from shape and reflectance, DiLiGenT$10^2$ [44] was proposed by fabricating 10 objects from CAD models with 10 carefully chosen materials (*e.g.*, `polyoxymethylene` and `steel`). Unlike previous datasets where the reflectance is controlled via semantic labels, DiLiGenRT quantitatively controls the reflectance properties by manufacturing 54 objects with 6 translucency levels and 9 roughness levels. In this way, we provide a more comprehensive evaluation of photometric stereo from the reflectance aspect. Besides, compared with DiLiGenT$10^2$ [44] and DiLiGenT-Π [52] providing only one translucent material, DiLiGenRT is more capable of evaluating the influence of subsurface scattering in photometric stereo by including different levels of translucency.

## 3. DiLiGenRT dataset

This section details the manufacturing procedure and capturing process of our real-world photometric stereo dataset, DiLiGenRT. We also introduce two setups for measuring the translucency and roughness of the materials in DiLiGenRT.

### 3.1. Object manufacturing and capturing

As shown in Fig. 1, DiLiGenRT consists of 54 spheres sharing the same shape but different materials. The objects on the same row share the same translucency, while those in the same column have the same roughness.

**Object manufacturing** To create DiLiGenRT, we begin with manufacturing 9 identical aluminum alloy molds. For creating varying roughness levels, the mold surfaces are sandblasted by grit with 9 different sizes ranging from #3000 to #20 (FEPA standard). Figure 2 (left) shows 6 out of the 9 molds, where the mold sandblasted with grit size #3000 exhibits a shinier appearance than the one with grit size #20. To produce distinct levels of translucency, we blend transparent silicone with color paste in different proportions, resulting in 6 solutions with varying concentrations between $1/1$ and $1/64$ as shown in the second column of Fig. 2. By injecting one of these solutions into the molds of different roughness levels and solidifying them, we obtain objects with identical translucency but different roughness levels, as shown in the third column of Fig. 2. Objects with other levels of translucency are also created in a similar manner. In total, we manufacture 54 spherical objects, each having one of 6 translucency levels and one of 9 roughness levels.

**Capturing setup** As shown on the right side of Fig. 2, we use a capture setup similar to DiLiGenT$10^2$ [44], where light directions are controlled by moving a single point light source attached to a robot arm. A lens is mounted in front of the point light source to enhance the directionality and uniformity of illumination. We capture 100 images for each object under different light directions, uniformly distributed over a hemisphere as shown in the left bottom of Fig. 1. Please refer to our supplementary material for further details regarding light calibration and camera settings.

### 3.2. Roughness and translucency measurement

This section describes the procedure for measuring the roughness and translucency of the manufactured materials.

**Roughness measurement** In the mechanical industry, measuring roughness is a well-explored field, leading to various roughness metrics. We choose surface area roughness $Sa$ as our roughness metric, which averages the absolute differences between each point's surface height and the local area's mean surface height. As depicted in Fig. 3, we assume the surface roughness across the entire sphere is uniform

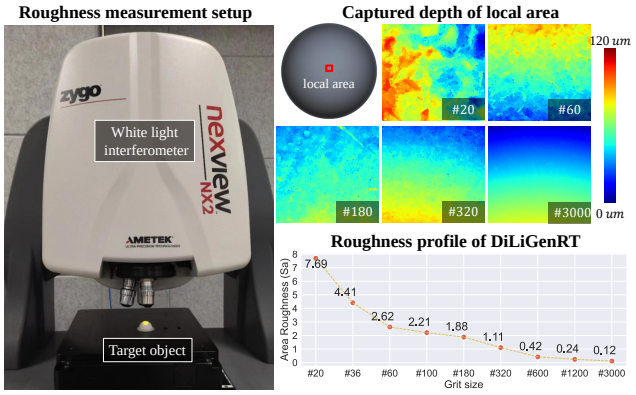



Figure 3. Roughness measurement at varying levels of grit size labeled in gray boxes. The measured roughness profile is shown in the bottom right.

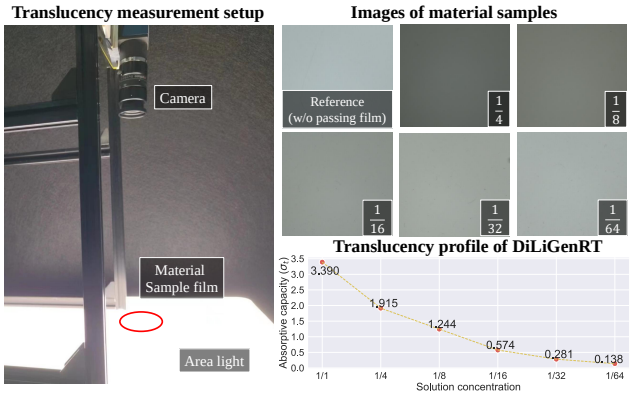



Figure 4. Translucency measurement at varying levels of solution concentration labeled in gray boxes. Observed radiance without a material sample is shown as a reference image. The translucency profile is shown in the bottom right.

and conduct surface height measurements at a highlighted local area. The surface heights are measured using a white light interferometer Zygo Nexview NX2[3] with vertical resolution of $0.15$ [nm]. As the grit size decreases, we observe the reduction in the variance of measured surface heights within the local area, showing that our objects have different levels of roughness. The roughness profile of DiLiGenRT, based on the $Sa$ metric w.r.t. different grit sizes, is shown in Fig. 3 (bottom-right).

**Translucency measurement** Based on the Beer-Lambert law [39], we use the absorptive capacity defined as $\sigma_t = \varepsilon c$ as the metric of translucency, which jointly considers the solution concentration $\varepsilon$ and molar extinction coefficient $c$ within the medium. Smaller $\sigma_t$ corresponds to a higher translucency level. As shown in Fig. 4, we manufacture 6 thin disks with a thickness of 3 [mm], solidified from the same 6 solutions used for DiLiGenRT. We put each thin film onto a screen light source and record the images by a camera with and without the film, whose average pixel values are recorded as $I_0$ and $I_1$, respectively. Due to the subsurface scattering, the received radiance decreases as the concentration of the solution increases. The absorptive capacity is calculated as $\sigma_t = -\log_{10}(\frac{I_0}{I_1})/d$, where $d$ is the thickness of the thin medium [39]. The translucency profile of DiLiGenRT measured by metric $\sigma_t$ w.r.t. different solution concentrations is shown in Fig. 4 (bottom-left).

[3] https://www.zygo.com/products/metrology-systems/3d-optical-profilers/nexview-nx2. Retrieved March 25th, 2024.

## 4. Benchmark evaluations

Similar to the latest photometric stereo benchmark DiLiGenT10$^2$ [44], we choose non-learning-based methods LS [53] and its robust version via position thresholding strategy TH28 [48] and TH46 [48], and representative work ST14 [47] as the baselines. For learning-based methods, we select classical photometric stereo methods PS-FCN [6] and CNN-PS [21], as well as the latest photometric stereo methods MS-PS [16] and NormAttention-PSN [30] (abbreviated as Att-PSN) as the baselines. Besides, we add a photometric stereo method tailored for translucent surfaces: DeconvPS [26]. We also include three uncalibrated photometric stereo methods: PF14 [43], SDPS-Net [7] (abbreviated as SDPS), and SDM-UniPS [24] in our benchmark.

The evaluations of other photometric stereo methods: UniPS [23], GPS-Net [57], PX-Net [35], and SPLINE-Net [59], DeepPS2 [50] are provided in our supplementary material. We implemented DeconvPS [26] based on their paper by utilizing 2D Gaussian functions as the convolution kernel and assigning the kernel size with the inverse of absorptive capacity value empirically. For the remaining baselines, we adopt the code and pre-trained models released by the authors. Following the practice of previous benchmark datasets, we use mean angular error (MAE) in degrees as the metric to evaluate estimated surface normals.

### 4.1. Roughness-translucency error matrix

Figure 5 presents the roughness-translucency error matrices as a heat map for each method evaluated on DiLiGenRT, where the non-learning-based and learning-based photometric stereo methods are shown in the top and bottom rows, respectively. Each cell in the matrix shows the MAE for the material indexed by the translucency and roughness levels based on the $\sigma_t$ and $Sa$ metrics. These matrices provide a performance profile of each photometric stereo method across the quantified reflectances.

**Non-learning-based methods** Opaque materials with a high degree of roughness, as in the matrix's top left corner, exhibit satisfactory results by all photometric stereo methods

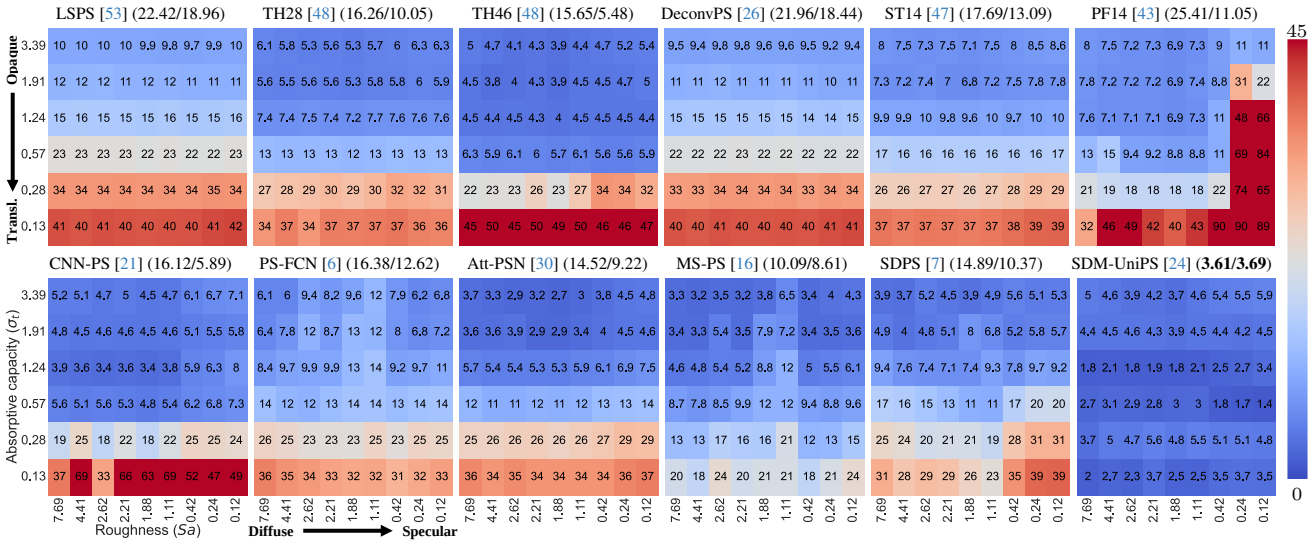



Figure 5. Roughness-translucency MAE matrices for non-learning-based (top) and learning-based (bottom) photometric stereo methods, showing their performance profiles under different levels of reflectance properties. The mean and median of the MAE matrix are presented near the method name. "Att-PSN" and "SDPS" are the abbreviations of NormAttention-PSN [30] and SDPS-Net [7]. The ticks of row and column are $\sigma_t$ and $Sa$. Reducing $\sigma_t$ corresponds to increasing translucency, while lowering $Sa$ is associated with decreased roughness.

because the reflectance is close to Lambertian. The surface normal estimation error tends to grow as translucency increases and roughness decreases. However, translucency appears to have a more significant influence than roughness. DeconvPS [26] enhances LS [53] by addressing subsurface scattering and shows smaller MAE values on translucent surfaces. However, the method is still prone to substantial errors due to attached shadows and the Lambertian assumption.

By using a position thresholding strategy, TH46 [48] outperforms all other non-learning-based photometric stereo methods. It indicates that for the mid-range of translucency levels (*i.e.*, $\sigma_t \in [0.57, 3.39]$), the Lambertian component is still strongly observed, which can be extracted via the simple position thresholding strategy. However, a notable error surge occurs when $\sigma_t \leq 0.28$, corresponding to the solution concentration of $1/32$. The Lambertian component is weaker in these materials due to the strong subsurface scattering. Under uncalibrated light setting, PF14 [43] shows a noticeable performance degradation when surface reflectance crosses certain levels of roughness and translucency.

**Learning-based methods** CNN-PS [21] has a similar performance profile with TH46 [48], which are both per-pixel-based methods, where the MAE increases significantly in high translucency levels. It may be due to that these methods do not consider neighboring pixels, which could be critical when dealing with significant subsurface scattering. The all-pixel-based method Att-PSN [30] shows better accuracy than PS-FCN [6] by addressing blurry normal estimates. This is particularly beneficial for photometric stereo applied to translucent surfaces, where subsurface scattering can blur the surface normal estimates [26].

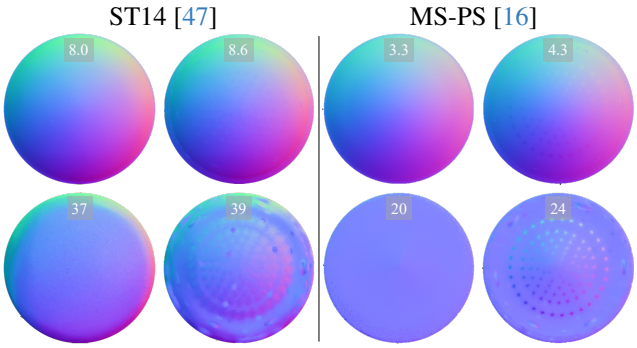



Figure 6. Visualization of estimated surface normals for hemisphere objects at the four corners of the translucency-roughness shown in Fig. 1. The MAEs in degrees are labeled in each sub-figure.

MS-PS [16] uses a multi-scale network structure to learn the global context of images, which contributes to modeling the light transport in subsurface scattering. Using a large-scale training dataset containing diverse materials, MS-PS [16] shows robustness across varying translucency levels. Under the uncalibrated light setting, SDPS [7] shows stabler accuracy than the non-learning-based method PF14 [43] and yields smaller average MAEs. SDM-UniPS [24] achieves the smallest average MAE compared to all other methods even without knowing the light direction. The average MAE is threefold smaller than the one achieved by the second-best method MS-PS [16].

Table 2. Roughness and translucency profiles of DiLiGenRT-S

| Object | PAW | LEAF | FLOWER |
|---|---|---|---|
| Roughness ($Sa$) | 1.845 | 0.698 | 1.189 |
| Absorptive capacity ($\sigma_t$) | 0.974 | 0.407 | 0.272 |

## 4.2. Analysis of estimated normal map

Figure 6 shows estimated surface normal maps using ST14 [47] and MS-PS [16] for surfaces at the four corners of the translucency-roughness matrix shown in Fig. 1. It is observed that the estimated shapes become flatter with increasing the translucency level. This is because subsurface scattering makes the observed radiance at different surface regions more uniform. In addition, with a lower roughness level, the estimated shapes exhibit high-frequency spikes due to specular highlights.

The methods ST14 [47] and MS-PS [16] effectively handle specular highlights on opaque surfaces, as shown by the slight increase in MAE between the even and odd columns in the first row of Fig. 6. However, when it comes to more translucent surfaces, the impact of specular highlights grows more significantly. As such, a photometric stereo method's proficiency in handling specular highlights on opaque surfaces does not necessarily translate to its performance on translucent surfaces.

## 4.3. Evaluation on diverse shapes by DiLiGenRT-S

In addition to sphere shape contained in DiLiGenRT, we also build DiLiGenRT-S to include three different shapes: PAW, LEAF, and FLOWER, as shown in Fig. 1 (right). Their translucency and roughness profiles provided in Table 2 are measured in the same way discussed in Sec. 3.2. The evaluation on DiLiGenRT-S is summarized in Table 3. It shows a consistent trend that the normal estimation error increases with a greater translucency level. The MAEs are generally higher compared to DiLiGenRT, due to cast-shadows and inter-reflections in complex shapes. The MAEs of SDM-UniPS [24] have a significant increase compared to other methods. It is perhaps due to that SDM-UniPS [24] is particularly good at sphere-like shapes because the pre-trained model was likely overfitting to such a shape in their synthetic training dataset PS-Wild [23] and PS-Mix [24]. Therefore, we will exclude SDM-UniPS [24] on our subsequent evaluations on DiLiGenRT hereafter due to the possible overfitting.

## 4.4. Roughness-translucency error over #lights

The accuracy of surface normal estimation is influenced by surface reflectance, the selected photometric stereo method, and the number of varying lights. Figure 7 offers evaluation results under sparse (#10) and dense (#100) light sets uniformly distributed over a hemisphere. Each cell within the roughness-translucency error table displays the best-performing photometric stereo method and its MAE for a given material under a specific number of lights.

Table 3. Evaluation on DiLiGenRT-S by MAE in degrees, where the lowest MAE is labeled in bold. In reference to Fig. 7, the MAE of the top-performing method at the corresponding $Sa$ and $\sigma_t$ is underlined, while the second-best is denoted with an asterisk (*).

| Object | LS [53] | TH28 [48] | TH46 [48] | DeconvPS [26] | ST14 [47] | CNN-PS [21] |
|---|---|---|---|---|---|---|
| PAW | 17.9 | 17.2 | **16.5*** | 17.9 | 17.7 | 17.3 |
| LEAF | 21.4 | 23.4 | 22.3 | 21.1 | 21.3 | 19.6 |
| FLOWER | 27 | 29.2 | 29 | 26.6 | 27 | 23.7* |
| **Object** | **PS-FCN [6]** | **Att-PSN [30]** | **MS-PS [16]** | **SDM-UniPS [24]** | **PF14 [43]** | **SDPS [7]** |
| PAW | 16.7 | 17 | 17 | 18.6 | 20.2 | 18.6 |
| LEAF | **17.9*** | 20.4 | 18.1 | 29.5 | 27.3 | 19.4 |
| FLOWER | 22.6 | 25.2 | **20.7** | 23.2 | 29.1 | 22.6 |

As shown in Fig. 7, with 100 lights, per-pixel-based methods like TH46 [48] and CNN-PS [21] yield superior results for semi-translucent materials (*i.e.*, $\sigma_t$ between 0.57 and 1.24). Conversely, all-pixel-based methods such as Att-PSN [30] and MS-PS [16] excel in handling opaque and highly translucent materials (*i.e.*, $\sigma_t$ above 1.24 or below 0.57), with MS-PS [16] particularly proficient at handling surfaces with high translucency. This trend remains consistent across varying roughness levels.

With a limited number of lights, *e.g.*, 10, all-pixel-based methods (Att-PSN [30] and MS-PS [16]) show superior accuracy than per-pixel-based ones (CNN-PS [21] and TH46 [48]) for 49 of 54 materials. One potential reason is that per-pixel-based methods predict surface normal only from the image measurements at a single pixel position, ignoring the spatial context provided by neighboring pixels, while all-pixel-based methods take them into account. This difference becomes significant under a sparse light setting, where image observations at a single pixel position are insufficient for per-pixel-based methods to accurately predict surface normals under general reflectance, while the broader observations from neighboring pixels benefit all-pixel-based methods for producing better normal estimates. According to the MAE values summarized in Fig. 7, for opaque and semi-translucent surfaces, 10 light sources suffice for photometric stereo given a maximum MAE tolerance of 10°.

**Method selection based on DiLiGenRT** Figure 7 also presents an overview of the minimum MAE bounds for photometric stereo methods and shows their working ranges. It serves as a guide to selecting an appropriate photometric stereo method based on the reflectance and light setting. For example, the closest cell to the LEAF object's roughness and absorptive capacity values shown in Table 2 is $(Sa, \sigma_t) = (0.42, 0.28)$ in DiLiGenRT. The best and second-best methods indexed by these values in Fig. 7 are MS-PS [16] and PS-FCN [6]. The MAEs of these methods shown in Table 3 closely match the smallest MAE observed for the Leaf object. Similar observations are also confirmed on the other two objects of DiLiGenRT-S. In this way, the

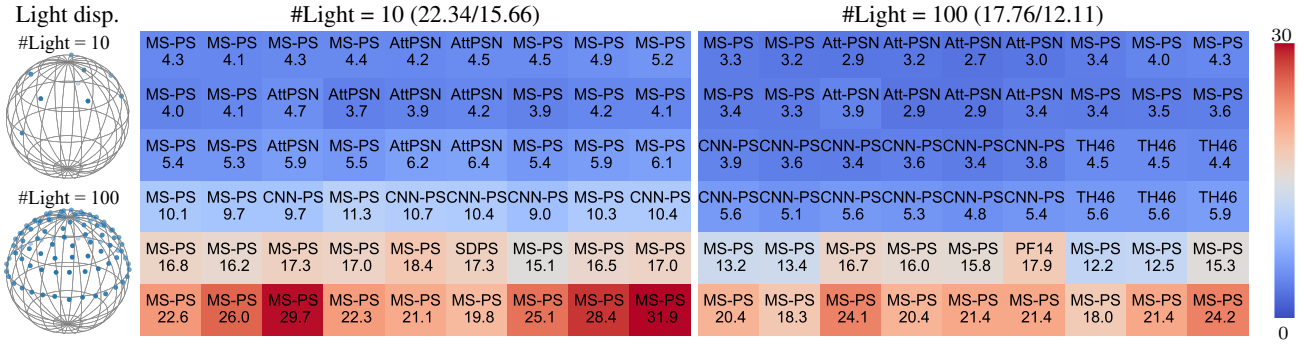



Figure 7. Benchmark evaluation on DiLiGenRT under sparse and dense lights (#10 and #100), summarized by mean/median MAE values. Each cell records the best-performing method for the material along with its MAE.

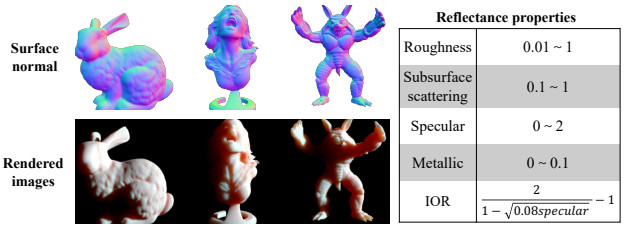



Figure 8. Selected shapes and images of synthetic dataset PS-SSS.

quantified reflectances of DiLiGenRT are useful in recommending the most suitable photometric stereo method given a target object.

### 4.5. Fine-tuning with a synthetic translucent dataset

The benchmark evaluation illustrates the challenges in accurately recovering surface normal for highly translucent surfaces due to severe subsurface scattering effects. One potential reason is that the training dataset used in existing learning-based methods does not include translucent materials. To address the problem, we propose to boost the performance of existing methods through fine-tuning using a synthetic translucent dataset.

As shown in Fig. 8, we render a synthetic dataset specifically with Sub-Surface Scattering, named PS-SSS. To create PS-SSS, we selected 30 shapes and randomly rotated them 100 times, leading to $3,000$ distinct surface normals. Similar to CyclePS [21], we use Blender [10] as our rendering engine and the Disney Principled BSDF [4] as the reflectance model. For each scene, we generate 100 uniformly-distributed light directions [37] and create a unique Principled BSDF whose reflectance properties are randomly assigned within a setting range, as shown in Fig. 8 (right), where the subsurface scattering parameter is adjusted between 0.1 and 1 to control translucency. In total, we generate $3,000$ scenes covering distinct translucent surfaces. Please find the complete shape list of PS-SSS in our supplementary material.

We test incorporating data prior from PS-SSS into existing learning-based methods. As an example, we use PS-FCN [6] as our baseline and fine-tune the normal regression module using our PS-SSS dataset over 9 epochs, where the learning rate is set to 1e-4. As shown in Fig. 9, the mean MAE decreases to 9.92°, which is much smaller compared with 16.38° in the original version presented in Fig. 5. This error reduction is particularly evident in surfaces with high translucency levels ($\sigma_t \leq 0.57$), showing the effectiveness of incorporating data prior for handling subsurface scattering.

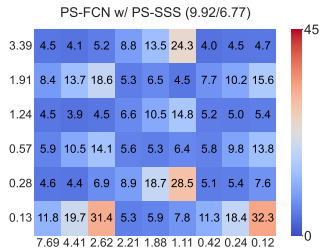



Figure 9. Performance profile of PS-FCN [6] after fine-tuning on PS-SSS.

## 5. Discussion

Our evaluation on DiLiGenRT shows that existing photometric stereo methods can effectively handle varying levels of roughness. However, highly translucent surfaces remain challenging for most methods, where the MAEs generally exceed 20°. Besides leveraging data prior from the synthetic dataset as discussed in Sec. 4.5, explicitly modeling the subsurface scattering and using differentiable rendering [58] could be promising. Moreover, DiLiGenRT is designed to be a method selector based on measured reflectance properties. Algorithms for selecting suitable photometric stereo methods based on input images remain unexplored. More discussions about the limitations and the application ranges of DiLiGenRT can be found in our supplementary material.

## Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant No. 62136001, 62088102, U22A20176, 52175477), JSPS KAKENHI (Grant No. JP23H05491). We thank Dr. Clément Hardy and Dr. Yvain Quéau for evaluating MS-PS [16] on DiLiGenRT.

## References


[1] Neil G. Alldrin, Satya P. Mallick, and David J. Kriegman. Resolving the generalized bas-relief ambiguity by entropy minimization. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007. 3

[2] Neil G. Alldrin, Todd Zickler, and David J. Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008. 3

[3] Peter N. Belhumeur, David J. Kriegman, and Alan L. Yuille. The bas-relief ambiguity. *International Journal of Computer Vision*, 1999. 2

[4] Brent Burley and Walt Disney Animation Studios. Physically-based shading at disney. In *Proc. of SIGGRAPH*, pages 1–7. vol. 2012, 2012. 3, 8

[5] Manmohan Chandraker, Jiamin Bai, and Ravi Ramamoorthi. On differential photometric reconstruction for unknown, isotropic brdfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12):2941–2955, 2012. 2

[6] Guanying Chen, Kai Han, and Kwan-Yee K. Wong. PS-FCN: A flexible learning framework for photometric stereo. In *Proc. of European Conference on Computer Vision (ECCV)*, 2018. 3, 5, 6, 7, 8

[7] Guanying Chen, Kai Han, Boxin Shi, Yasuyuki Matsushita, and Kwan-Yee K. Wong. Self-calibrating deep photometric stereo networks. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 3, 5, 6, 7

[8] Guanying Chen, Michael Waechter, Boxin Shi, Kwan-Yee K Wong, and Yasuyuki Matsushita. What is learned in deep uncalibrated photometric stereo? In *Proc. of European Conference on Computer Vision (ECCV)*, pages 745–762. Springer, 2020. 3

[9] Lixiong Chen, Yinqiang Zheng, Boxin Shi, Art Subpa-Asa, and Imari Sato. A microfacet-based model for photometric stereo with general isotropic reflectance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1):48–61, 2019. 3

[10] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Blender Foundation, Amsterdam, 2021. 3, 8

[11] Bo Dong, Kathleen D Moore, Weiyi Zhang, and Pieter Peers. Scattering parameters and surface normals from homogeneous translucent materials using photometric stereo. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 3

[12] Kenji Enomoto, Michael Waechter, Kiriakos N Kutulakos, and Yasuyuki Matsushita. Photometric stereo via discrete hypothesis-and-test search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2311–2319, 2020. 3

[13] Kenji Enomoto, Michael Waechter, Fumio Okura, Kiriakos N Kutulakos, and Yasuyuki Matsushita. Discrete search photometric stereo for fast and accurate shape estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 3

[14] Dan B Goldman, Brian Curless, Aaron Hertzmann, and Steven M Seitz. Shape and spatially-varying brdfs from photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009. 3

[15] Bjoern Haefner, Songyou Peng, Alok Verma, Yvain Quéau, and Daniel Cremers. Photometric depth super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 4

[16] Clément Hardy, Yvain Quéau, and David Tschumperlé. MS-PS: A multi-scale network for photometric stereo with a new comprehensive training dataset. *arXiv preprint arXiv:2211.14118*, 2022. 3, 5, 6, 7, 8

[17] Hideki Hayakawa. Photometric stereo under a light source with arbitrary motion. *JOSA A*, 1994. 2

[18] Yannick Hold-Geoffroy, Paulo Gotardo, and Jean-François Lalonde. Single day outdoor photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 4

[19] Berthold K.P Horn and Michael J Brooks. The variational approach to shape from shading. *Computer Vision, Graphics, and Image Processing*, 33(2):174–208, 1986. 2

[20] Zhuo Hui and Aswin C Sankaranarayanan. Shape and spatially-varying reflectance estimation from virtual exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(10):2060–2073, 2016. 3

[21] Satoshi Ikehata. CNN-PS: CNN-based photometric stereo for general non-convex surfaces. In *Proc. of European Conference on Computer Vision (ECCV)*, 2018. 3, 5, 6, 7, 8

[22] Satoshi Ikehata. PS-Transformer: Learning sparse photometric stereo network using self-attention mechanism. *Proc. of the British Machine Vision Conference (BMVC)*, 2022. 3

[23] Satoshi Ikehata. Universal photometric stereo network using global lighting contexts. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12591–12600, 2022. 3, 5, 7

[24] Satoshi Ikehata. Scalable, detailed and mask-free universal photometric stereo. *arXiv preprint arXiv:2303.15724*, 2023. 3, 5, 6, 7

[25] Satoshi Ikehata, David Wipf, Yasuyuki Matsushita, and Kiyoharu Aizawa. Robust photometric stereo using sparse regression. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 3

[26] Chika Inoshita, Yasuhiro Mukaigawa, Yasuyuki Matsushita, and Yasushi Yagi. Surface normal deconvolution: Photometric stereo for optically thick translucent objects. In *Proc. of European Conference on Computer Vision (ECCV)*, pages 346–359, 2014. 3, 5, 6, 7

[27] Wenzel Jakob. Mitsuba renderer, 2010. 3

[28] Henrik Wann Jensen, Stephen R Marschner, Marc Levoy, and Pat Hanrahan. A practical model for subsurface light transport. In *Proceedings of Annual Conference on Computer Graphics and Interactive Techniques*, pages 511–518, 2001. 3

[29] Micah K Johnson and Edward H Adelson. Shape estimation in natural illumination. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011. 3

[30] Yakun Ju, Boxin Shi, Muwei Jian, Lin Qi, Junyu Dong, and Kin-Man Lam. Normattention-PSN: A high-frequency region enhanced photometric stereo network with normalized

attention. *International Journal of Computer Vision*, 130(12): 3014–3034, 2022. 3, 5, 6, 7

[31] Berk Kaya, Suryansh Kumar, Carlos Oliveira, Vittorio Ferrari, and Luc Van Gool. Uncalibrated neural inverse rendering for photometric stereo of general surfaces. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 3, 4

[32] Junxuan Li and Hongdong Li. Self-calibrating photometric stereo by neural inverse rendering. In *Proc. of European Conference on Computer Vision (ECCV)*, pages 166–183. Springer, 2022. 3

[33] Junxuan Li, Antonio Robles-Kelly, Shaodi You, and Yasuyuki Matsushita. Learning to minify photometric stereo. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 3

[34] Min Li, Zhenglong Zhou, Zhe Wu, Boxin Shi, Changyu Diao, and Ping Tan. Multi-view photometric stereo: A robust solution and benchmark dataset for spatially varying isotropic materials. *IEEE Transactions on Image Processing*, 29:4159–4173, 2020. 1, 3, 4

[35] Fotios Logothetis, Ignas Budvytis, Roberto Mecca, and Roberto Cipolla. PX-NET: Simple and efficient pixel-wise training of photometric stereo networks. In *Proc. of International Conference on Computer Vision (ICCV)*, pages 12757–12766, 2021. 3, 5

[36] Fotios Logothetis, Roberto Mecca, Ignas Budvytis, and Roberto Cipolla. A CNN based approach for the point-light photometric stereo problem. *International Journal of Computer Vision*, pages 1–20, 2022. 1, 3

[37] George Marsaglia. Choosing a point from the surface of a sphere. *The Annals of Mathematical Statistics*, 43(2):645–646, 1972. 8

[38] Wojciech Matusik, Hanspeter Pfister, Matt Brand, and Leonard McMillan. A data-driven reflectance model. In *Proc. of SIGGRAPH*, 2003. 3

[39] Thomas G Mayerhöfer, Susanne Pahlow, and Jürgen Popp. The bouguer-beer-lambert law: Shining light on the obscure. *ChemPhysChem*, 21(18):2029–2046, 2020. 5

[40] Roberto Mecca, Fotios Logothetis, Ignas Budvytis, and Roberto Cipolla. LUCES: A dataset for near-field point light source photometric stereo. In *Proc. of the British Machine Vision Conference (BMVC)*, 2021. 1, 3, 4

[41] Kazuma Minami, Hiroaki Santo, Fumio Okura, and Yasuyuki Matsushita. Symmetric-light photometric stereo. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2706–2714, 2022. 2

[42] Thoma Papadhimitri and Paolo Favaro. A new perspective on uncalibrated photometric stereo. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1474–1481, 2013. 2

[43] Thoma Papadhimitri and Paolo Favaro. A closed-form, consistent and robust solution to uncalibrated photometric stereo via local diffuse reflectance maxima. *International Journal of Computer Vision*, 2014. 2, 5, 6, 7

[44] Jieji Ren, Feishi Wang, Jiahao Zhang, Qian Zheng, Mingjun Ren, and Boxin Shi. DiLiGenT10$^2$: A photometric stereo benchmark dataset with controlled shape and material variation. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1, 3, 4, 5

[45] Hiroaki Santo, Masaki Samejima, Yusuke Sugano, Boxin Shi, and Yasuyuki Matsushita. Deep photometric stereo network. In *Proc. of International Conference on Computer Vision Workshops (ICCVW)*, 2017. 3

[46] Boxin Shi, Yasuyuki Matsushita, Yichen Wei, Chao Xu, and Ping Tan. Self-calibrating photometric stereo. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010. 2

[47] Boxin Shi, Ping Tan, Yasuyuki Matsushita, and Katsushi Ikeuchi. Bi-polynomial modeling of low-frequency reflectances. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014. 3, 5, 6, 7

[48] Boxin Shi, Zhipeng Mo, Zhe Wu, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation for non-Lambertian and uncalibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 1, 3, 4, 5, 6, 7

[49] William M. Silver. *Determining shape and reflectance using multiple images*. PhD thesis, Massachusetts Institute of Technology, 1980. 2

[50] Ashish Tiwari and Shanmuganathan Raman. DeepPS2: Revisiting photometric stereo using two differently illuminated images. In *Proc. of European Conference on Computer Vision (ECCV)*, pages 129–145. Springer, 2022. 3, 5

[51] Ashish Tiwari and Shanmuganathan Raman. LERPS: lighting estimation and relighting for photometric stereo. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2060–2064. IEEE, 2022. 3

[52] Feishi Wang, Jieji Ren, Heng Guo, Mingjun Ren, and Boxin Shi. DiLiGenT-Pi: Photometric stereo for planar surfaces with rich details-benchmark dataset and beyond. In *Proc. of International Conference on Computer Vision (ICCV)*, pages 9477–9487, 2023. 1, 3, 4

[53] Robert J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 1980. 1, 2, 5, 6, 7

[54] Lun Wu, Arvind Ganesh, Boxin Shi, Yasuyuki Matsushita, Yongtian Wang, and Yi Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *Proc. of Asian Conference on Computer Vision (ACCV)*, 2010. 3

[55] Ying Xiong, Ayan Chakrabarti, Ronen Basri, Steven J Gortler, David W Jacobs, and Todd Zickler. From shading to local shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014. 3

[56] Ying Xiong, Ayan Chakrabarti, Ronen Basri, Steven J. Gortler, David W. Jacobs, and Todd Zickler. From shading to local shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015. 3

[57] Zhuokun Yao, Kun Li, Ying Fu, Haofeng Hu, and Boxin Shi. GPS-Net: Graph-based photometric stereo network. *Proc. of Annual Conference on Neural Information Processing Systems (NeurIPS)*, 33:10306–10316, 2020. 3, 5

[58] Hong-Xing Yu, Michelle Guo, Alireza Fathi, Yen-Yu Chang, Eric Ryan Chan, Ruohan Gao, Thomas Funkhouser, and Jiajun Wu. Learning object-centric neural scattering functions

for free-viewpoint relighting and scene composition. *arXiv preprint arXiv:2303.06138*, 2023. 8

[59] Qian Zheng, Yiming Jia, Boxin Shi, Xudong Jiang, Ling-Yu Duan, and Alex C. Kot. SPLINE-Net: Sparse photometric stereo through lighting interpolation and normal estimation networks. In *Proc. of International Conference on Computer Vision (ICCV)*, 2019. 3, 5

[60] Zhenglong Zhou and Ping Tan. Ring-light photometric stereo. In *Proc. of European Conference on Computer Vision (ECCV)*, 2010. 2

# Supplementary Material for DiLiGenRT: A Photometric Stereo Dataset with Quantified Roughness and Translucency

Heng Guo[1,†] Jieji Ren[2,†] Feishi Wang[3,4,5†] Boxin Shi[3,4,5*] Mingjun Ren[2*] Yasuyuki Matsushita[6]

[1] School of Artificial Intelligence, Beijing University of Posts and Telecommunications

[2] School of Mechanical Engineering, Shanghai Jiao Tong University

[3] National Key Laboratory for Multimedia Information Processing, School of Computer Science, Peking University

[4] National Engineering Research Center of Visual Technology, School of Computer Science, Peking University

[5] AI Innovation Center, School of Computer Science, Peking University

[6] Graduate School of Information Science and Technology, Osaka University

In this supplementary material, we first provide additional technical details about the light calibration and data capture process. Then we describe the limitations of the proposed DiLiGenRT dataset. After that, we analyze the influence of light distribution on the photometric stereo when dealing with diverse materials, followed by providing the evaluation of additional photometric stereo methods, including UniPS [6], GPS-Net [19], PX-Net [11], SPLINE-Net [20], and DeepPS2 [17]. Finally, we provide the complete benchmark results containing estimated surface normals and the corresponding angular error maps on DiLiGenRT dataset for each method.

## A. Light calibration and capture settings

**Light calibration** As shown in Fig. S1 (a), we place 6 mirror spheres near the target object. Following the practice of existing photometric stereo datasets (*e.g.*, DiLiGenT [16], DiLiGenT10$^2$ [14]), the incident light directions can be calibrated via the specular spots on the mirror balls. Specifically, for the $i$-th of 6 mirror spheres, we first use circle Hough transform [8] to detect its projected circle on the image plane, extracting radius $r_i$ in pixel unit and circle center location $\mathbf{c}_i = (u_{0i}, v_{0i})$, as shown in Fig. S1 (c). Then we manually label specular highlight position $\mathbf{p}_i = (u_{pi}, v_{pi})$. Assuming that the world center is aligned with the sphere, the 3D coordinates in pixel unit of the specular spot can be represented as

$$\mathbf{s}_i = \begin{pmatrix} u_{pi} - u_{0i} \\ v_{pi} - v_{0i} \\ \sqrt{r_i^2 - (u_{pi} - u_{0i})^2 - (v_{pi} - v_{0i})^2} \end{pmatrix} \tag{1}$$

As shown in Fig. S1 (right), the surface normal direction at point $\mathbf{s}_i$ on the sphere can be calculated as

$$\mathbf{n}_i = \frac{\mathbf{s}_i}{\|\mathbf{s}_i\|} = \begin{pmatrix} (u_{pi} - u_{0i})/r_i \\ (v_{pi} - v_{0i})/r_i \\ \sqrt{r_i^2 - (u_{pi} - u_{0i})^2 - (v_{pi} - v_{0i})^2}/r_i \end{pmatrix} \tag{2}$$

Thus we have:

$$\begin{aligned} n_{xi} r_i &= u_{pi} - u_{0i}, \\ n_{yi} r_i &= v_{pi} - v_{0i}. \end{aligned} \tag{3}$$

By assuming distant illumination on all of the six mirror spheres, the $\mathbf{n}_i$ on each sphere should be equal. Consequently, we denote an optimized half vector as $\mathbf{n} = [n_x, n_y, n_z]^\top$. Its coordinates under noise can be written as:

$$\begin{aligned} n_x r_i &= u_{pi} - u_{0i} + \epsilon_{xi}, \\ n_y r_i &= v_{pi} - v_{0i} + \epsilon_{yi}, \end{aligned} \tag{4}$$

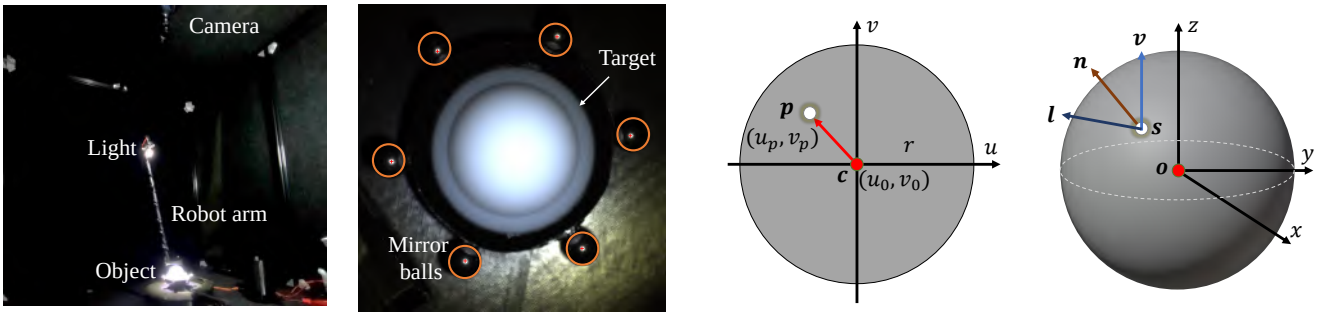
where $\epsilon_{xi}$ and $\epsilon_{yi}$ denote the Gaussian noise terms. Using the principle of maximum likelihood inference, we perform least squares optimization to obtain the optimal values of $n_x$ and $n_y$:

$$\begin{aligned} n_x &= \frac{\sum_{i=1}^{6} (u_{pi} - u_{0i}) r_i}{\sum_{i=1}^{6} r_i^2}, \\ n_y &= \frac{\sum_{i=1}^{6} (v_{pi} - v_{0i}) r_i}{\sum_{i=1}^{6} r_i^2}, \end{aligned} \tag{5}$$

and calculate $n_z$ simply by $n_z = \sqrt{1 - n_x^2 - n_y^2}$. Based on the law of reflection, light direction $\mathbf{l}$, camera view direction $\mathbf{v}$, and surface normal vector $\mathbf{n}$ at specular spot $\mathbf{s}$ follows

$$\mathbf{l} = 2(\mathbf{n}^\top \mathbf{v})\mathbf{n} - \mathbf{v}. \tag{6}$$

* labels corresponding authors (Email address: shiboxin@pku.edu.cn, renmj@sjtu.edu.cn). † denotes equally contributed authors.




(a) Capture setup (b) Light direction calibration setup (c) Specular spot on image plane (d) Specular spot in 3D coordinate

Figure S1. Our capture and calibration setups include 6 mirror balls for light calibration.

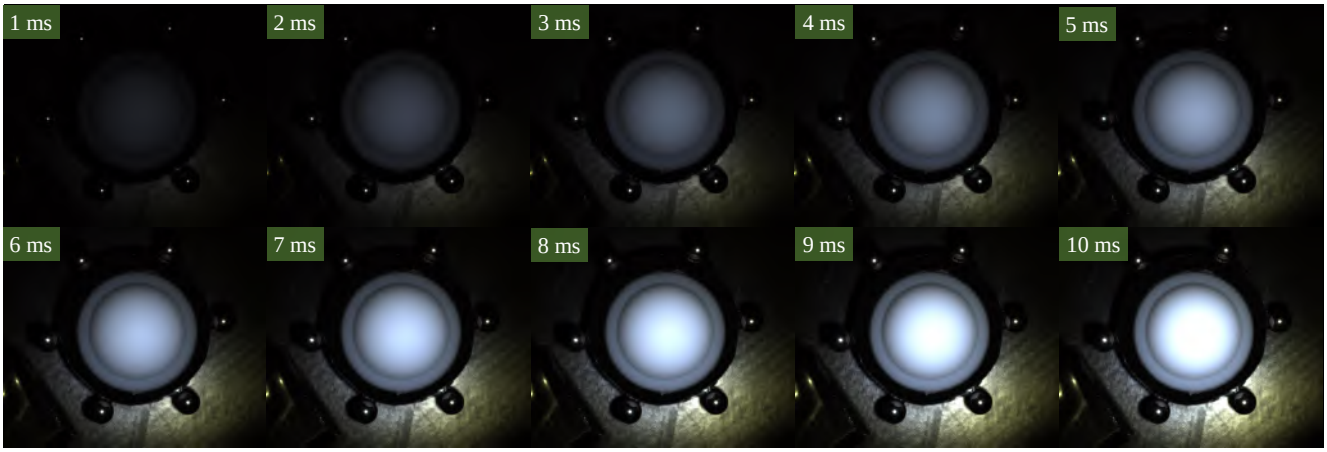



Figure S2. The target scene is captured under multiple exposure times from 1ms to 10ms to composite an HDR image.

As we fix the view direction $\mathbf{v} = [0, 0, 1]^\top$ and know the surface normal $\mathbf{n}$ from the specular spot at $\mathbf{s}$, light direction $\mathbf{l}$ can be calibrated following Eq. (6).

On the other hand, similar to DiLiGenT10$^2$ [14], our capture setup changes the illumination by shifting a single point light source mounted on a robot arm across a hemisphere, with the target object placed at the hemisphere center. This arrangement keeps the distance between the point light and the object approximately constant. As a result, we assume distant illumination, where the light intensities received from various light positions on the object remain uniform.

**Capture settings** To capture DiLiGenRT dataset, we adopt DaHeng Image MER-503-36U3C [1] camera equipped with a 50 mm lens, producing raw images at a resolution of $2448 \times 2048$, as shown in Fig. S1. We crop the images to $960 \times 960$ resolution to focus on the central valid areas. For each target object, we first position the point light source via the robot arm to illuminate the scene, followed by capturing images at 10 distinct exposure times that range from 1ms to 10ms, as illustrated in Fig. S2. Subsequently, these low dynamic range images are amalgamated to compose a high dynamic range (HDR) image [13]. In this way, the images captured under various illuminations within DiLiGenRT are in HDR format, avoiding the impact of image saturation and low-albedo pixels on photometric stereo.

[1] DaHeng camera: https://en.daheng-imaging.com/show-107-2044-1.html. Retrieved March 25th, 2024.

## B. Limitations of DiLiGenRT

This paper focuses on assessing photometric stereo under quantified reflectance properties. There are several limitations in DiLiGenRT dataset.

**Shape diversity.** DiLiGenRT contains sphere shape only. Since translucency is also related to the shape, the evaluation results under different translucency levels in DiLiGenRT dataset could be biased towards the sphere shape. Also, cast shadows and inter-reflections are important factors affecting the performance of photometric stereo methods. However, these phenomena are not included in DiLiGenRT due to the convex sphere shape. Therefore, it is desired to add more

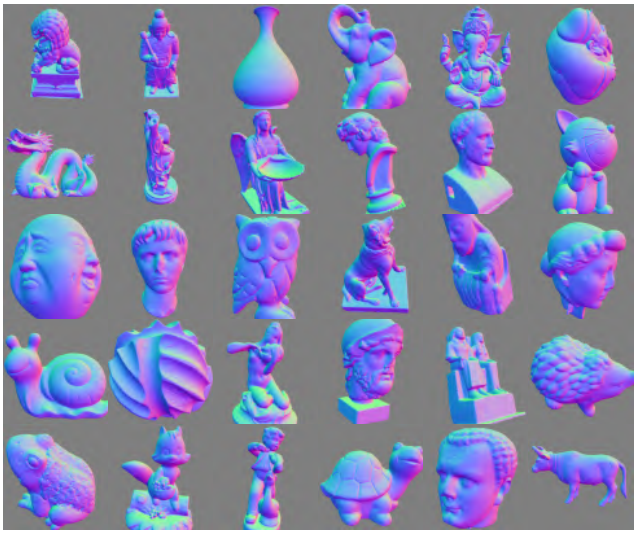


Figure S3. Subset of the 3000 surface normals of our synthetic dataset PS-SSS.

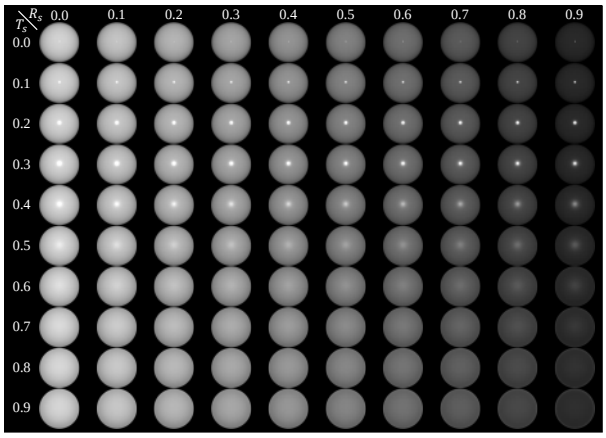



Figure S4. Synthetic dataset PS-Sphere indexed by the roughness $R_s$ and transmittance $T_s$ defined in the Principled BSDF model.

shapes to the dataset.

**Reflectance diversity.** DiLiGenRT contains dielectric objects with isotropic albedo and roughness. The IOR related to the translucency is set to be the same for all the objects in DiLiGenRT. It is desired to further add metallic objects, and dielectric objects with spatially varying albedos, anisotropic roughness, and different levels of IORs so that we can enrich the diversity of the reflectance contained in the dataset. The challenge is that systematically manufacturing surfaces with controlled levels of anisotropic roughness is not straightforward. Also, it is hard to quantify the influence of spatially varying albedo on photometric stereo.

In our future work, we plan to enlarge the scale of DiLiGenRT dataset to address the shape and reflectance diversity for a more comprehensive evaluation of photometric stereo methods.

## C. Diverse shapes in PS-SSS

As shown in Fig. S3, we select 30 objects from Sketchfab [2] For each object, we randomly rotate it for 100 times, leading to $3,000$ diverse surface normal maps. Given the rotated shapes and diverse materials controlled by Principled BSDF model [1], we render $3,000$ sets of images to create PS-SSS.

## D. DiLiGenRT vs synthetic sphere dataset

Compared to the labor-intensive manufacturing process of DiLiGenRT, an alternative way is rendering a synthetic sphere dataset (denoted as PS-Sphere) by adjusting the roughness and transmission metrics $Rs$ and $Ts$ defined in the Disney Principle BSDF model [1], ranging from 0 to 1. We provide such a synthetic dataset and conduct a similar benchmark evaluation like DiLiGenRT, as shown in Figs. S4 and S5, respectively.

To the best of our knowledge, there is no mapping between the synthetic roughness $R_s$ and the real-world *measurable* roughness $Sa$. Therefore, the evaluation results shown in Fig. S5 cannot be used to select best-fit photometric stereo methods as we have no device to access $R_s$ of a real-world object. This also applies to the case of $T_s$. On the other hand, we observed that the mean angular errors (MAE) of photometric stereo methods evaluated on PS-Sphere are much smaller than those on DiLiGenRT shown in Fig. 5 of the main paper, despite that the observed images from Fig. S4 and Fig. 1 of the main paper are similar. Therefore, there could be a domain gap between real-world reflectance and the parametric reflectance model, which highlights the necessity of DiLiGenRT for accurately assessing photometric stereo performance in real-world scenarios.

## E. Analysis on the light distribution

As shown in Fig. 7 of the main paper, we provide the evaluation results of photometric stereo methods under 10 and 100 lights. In Fig. S6, we further show the evaluation results under 20 and 50 uniformly distributed lights and present the best-performing method under different reflectance settings.

Increasing the number of input lights generally reduces the MAEs, as supported by the summarized mean and median MAE values in Fig. S6. However, we find 50 uniformly distributed lights is optimal on the DiLiGenRT dataset. Adding lights to 100 shows only a marginal improvement for opaque and semi-translucent surfaces across various roughness levels, but could be even harmful for surfaces with higher translucency levels. For instance, CNN-PS [5]

[2] https://sketchfab.com. Retrieved March 25th, 2024.

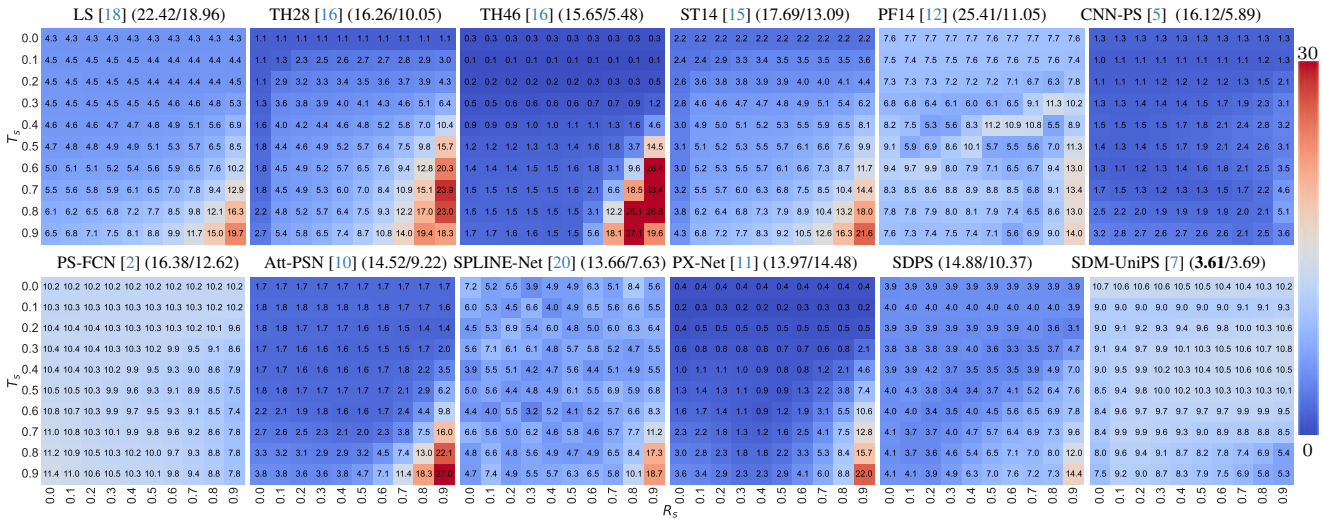



Figure S5. Roughness-translucency MAE matrices for 16 photometric stereo methods, where the ticks of row and column are $\sigma_t$ and $Sa$. The mean and median of the MAE matrix are presented near the method name, showing method's performance profile under different level of reflectance properties. 'Att-PSN' and 'SDPS' are the abbreviations of NormAttention-PSN [10] and SDPS-Net [3].

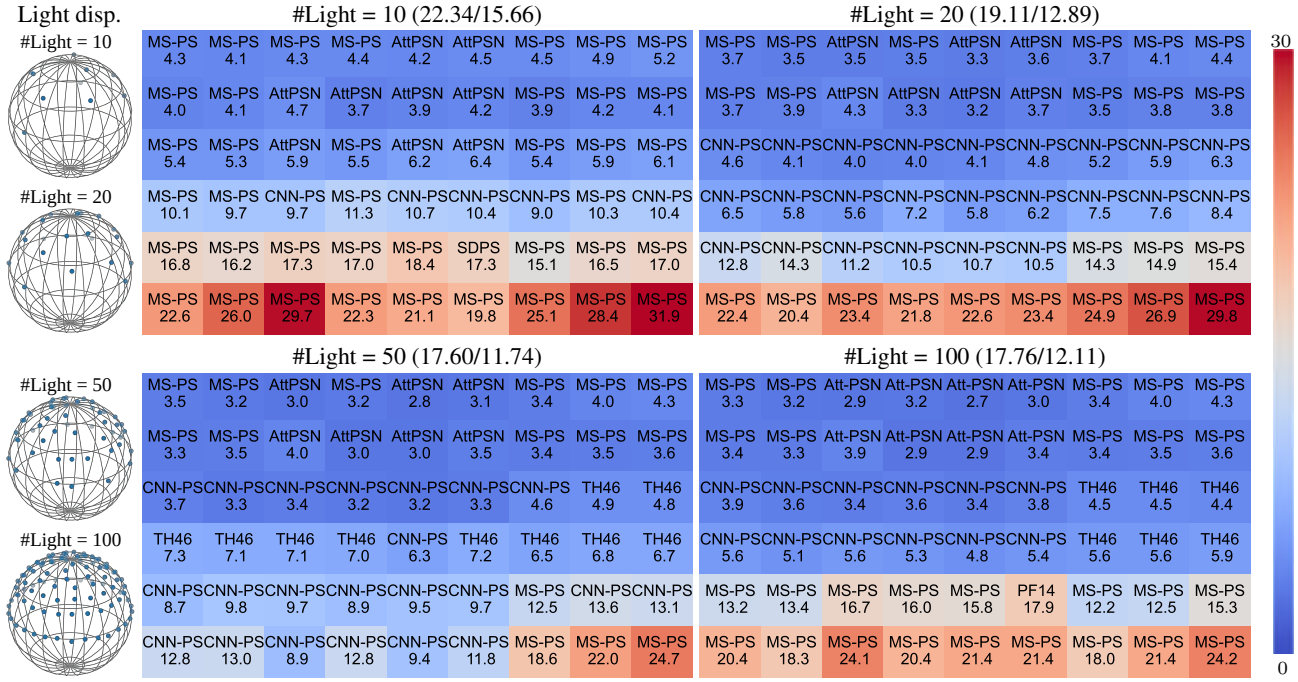



Figure S6. Benchmark evaluation on DiLiGenRT under different number of lights (#10 to #100) distributed uniformly, summarized by mean/median MAE values. Each cell records the best performing method for the material in that cell, along with the associated lowest MAE value.

achieves much smaller MAE on surfaces whose translucency measurement ($\sigma_t$) is 0.13, if reducing the number of input lights from 100 to 50. This effect could be attributed not only to the decrease in the amount of light but also to alterations in the distribution of light directions.

To demonstrate this, we conducted an experiment with CNN-PS [5] on two target spheres with differing degrees of roughness and translucency, as shown in Fig. S7. The number of input lights was fixed at 50, but their distribution was manipulated to be either uniform or biased, as illustrated

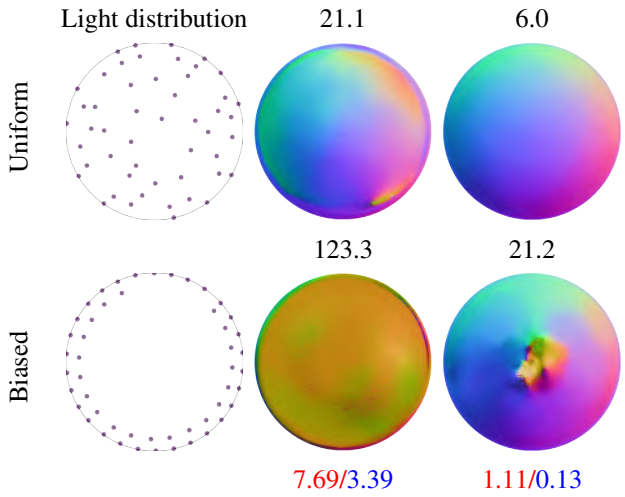



Figure S7. The influence of light distribution on surface normal estimation, tested on two spheres whose roughness and translucency measured by $Sa/\sigma_t$ are shown at the bottom. The normal estimation error measured by MAE is displayed at the top of the surface normal estimates.

on the left side of Fig. S7. We set the hyper-parameter $K$ in CNN-PS [5] as 1 to avoid the influence of averaging rotated surface normals.

As shown in Fig. S7, when more light directions are concentrated near the equator, the estimated surface normals from CNN-PS [5] exhibit greater MAEs compared to those achieved under uniformly distributed light directions. Furthermore, the sensitivity of surface normal estimation w.r.t. the amount and distribution of the incident lights increases when surfaces exhibit a higher level of translucency and a lower level of roughness. Therefore, the amount and distribution of lights serve as an important role in improving the accuracy of photometric stereo methods. It is desired to conduct adaptive illumination planning corresponding to varying reflectance.

## F. Performance profiles for additional photometric stereo methods

Besides the 12 photometric stereo methods evaluated in the main paper, this supplementary material offers evaluations on 5 additional cutting-edge photometric stereo methods: PX-Net [11], SPLINE-Net [20], UniPS [6], DeepPS2 [17], and GPS-Net [19], along with their performance results on DiLiGenRT. GPS-Net [19] published on NeurIPS 2020 combines the merits of all-pixel-based and per-pixel-based photometric stereo method via a graph neural network. SPLINE-Net [20] and PX-Net [11], presented at ICCV 2019 and 2021 respectively, are enhancements to the per-pixel-based method CNN-PS [5], focusing on sparse inputs and global illuminations. DeepPS2 [17] published at ECCV 2022 further reduces the sparse light input to only 2 distinct directional lights. UniPS [6] introduced at CVPR 2022 are built under uncalibrated universal illumination. As illustrated in Fig. S8, we display the performance profiles of all 17 photometric stereo methods.

## G. Complete benchmark results

From Fig. S9 to Fig. S25, we present the complete benchmark evaluations for 17 photometric stereo methods using DiLiGenRT dataset, including the 12 methods outlined in the main paper, as well as 5 additional methods detailed in the supplementary material. For each method, we provide a $6 \times 9$ matrix of their estimated surface normal map alongside their corresponding angular error distribution map. For better visualization, the maximal MAE is truncated at $10^\circ$ for UniPS [6] and SDM-UniPS [7] and $45^\circ$ for other methods. The $x$ and $y$ axes in the matrix denote the translucency and roughness measurements, measured by $\sigma_t$ and $Sa$ respectively.

## References


[1] Brent Burley and Walt Disney Animation Studios. Physically-based shading at disney. In *Proc. of SIGGRAPH*, pages 1–7. vol. 2012, 2012. 3

[2] Guanying Chen, Kai Han, and Kwan-Yee K. Wong. PS-FCN: A flexible learning framework for photometric stereo. In *Proc. of European Conference on Computer Vision (ECCV)*, 2018. 4, 6, 13

[3] Guanying Chen, Kai Han, Boxin Shi, Yasuyuki Matsushita, and Kwan-Yee K. Wong. Self-calibrating deep photometric stereo networks. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 4, 6, 20

[4] Clément Hardy, Yvain Quéau, and David Tschumperlé. MS-PS: A multi-scale network for photometric stereo with a new comprehensive training dataset. *arXiv preprint arXiv:2211.14118*, 2022. 6, 18

[5] Satoshi Ikehata. CNN-PS: CNN-based photometric stereo for general non-convex surfaces. In *Proc. of European Conference on Computer Vision (ECCV)*, 2018. 3, 4, 5, 6, 12

[6] Satoshi Ikehata. Universal photometric stereo network using global lighting contexts. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12591–12600, 2022. 1, 5, 6, 22

[7] Satoshi Ikehata. Scalable, detailed and mask-free universal photometric stereo. *arXiv preprint arXiv:2303.15724*, 2023. 4, 5, 6, 23

[8] John Illingworth and Josef Kittler. The adaptive hough transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (5):690–698, 1987. 1

[9] Chika Inoshita, Yasuhiro Mukaigawa, Yasuyuki Matsushita, and Yasushi Yagi. Surface normal deconvolution: Photometric stereo for optically thick translucent objects. In *Proc. of European Conference on Computer Vision (ECCV)*, pages 346–359, 2014. 6, 10

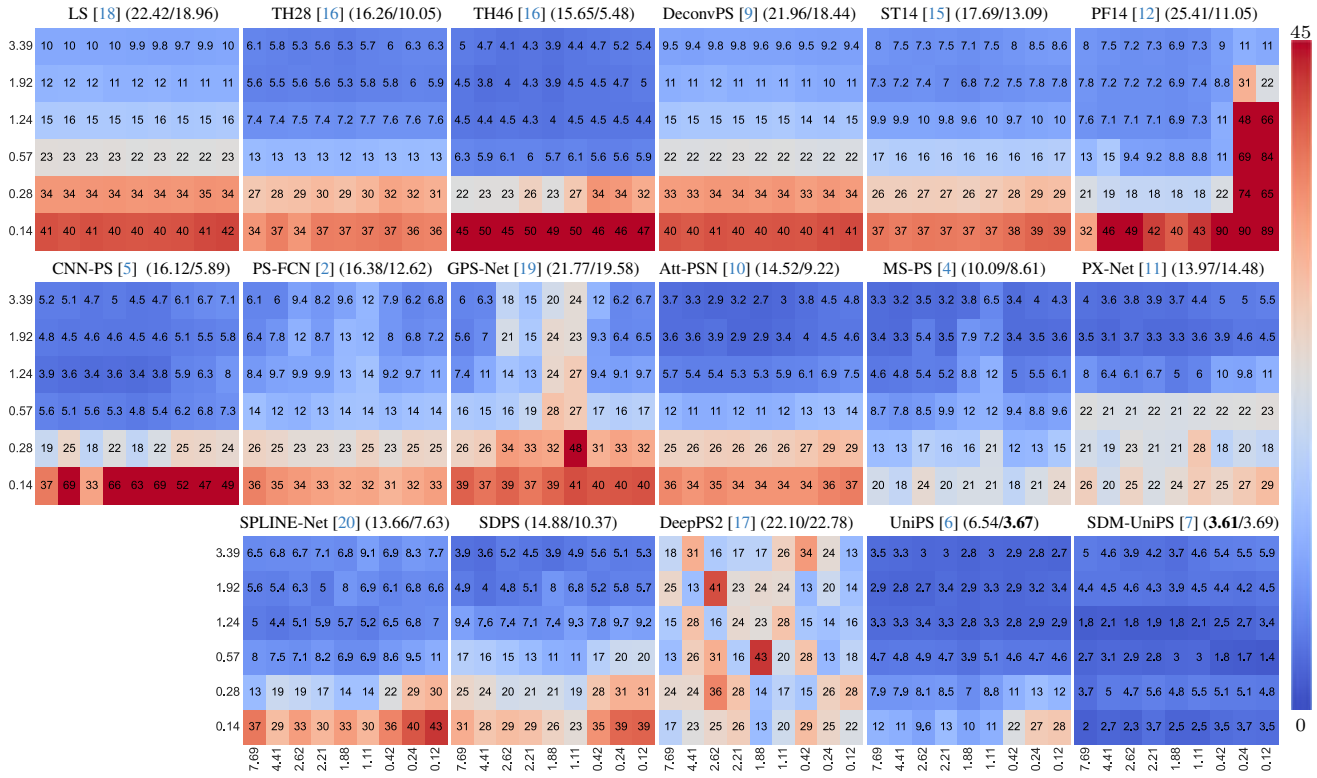



Figure S8. Roughness-translucency MAE matrices for 16 photometric stereo methods, where the ticks of row and column are $\sigma_t$ and $Sa$. The mean and median of the MAE matrix are presented near the method name, showing method's performance profile under different level of reflectance properties. 'Att-PSN' and 'SDPS' are the abbreviations of NormAttention-PSN [10] and SDPS-Net [3].


[10] Yakun Ju, Boxin Shi, Muwei Jian, Lin Qi, Junyu Dong, and Kin-Man Lam. Normattention-PSN: A high-frequency region enhanced photometric stereo network with normalized attention. *International Journal of Computer Vision*, 130(12): 3014–3034, 2022. 4, 6, 15

[11] Fotios Logothetis, Ignas Budvytis, Roberto Mecca, and Roberto Cipolla. PX-NET: Simple and efficient pixel-wise training of photometric stereo networks. In *Proc. of International Conference on Computer Vision (ICCV)*, pages 12757–12766, 2021. 1, 4, 5, 6, 17

[12] Thoma Papadhimitri and Paolo Favaro. A closed-form, consistent and robust solution to uncalibrated photometric stereo via local diffuse reflectance maxima. *International Journal of Computer Vision*, 2014. 4, 6, 19

[13] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010. 2

[14] Jieji Ren, Feishi Wang, Jiahao Zhang, Qian Zheng, Mingjun Ren, and Boxin Shi. DiLiGenT10$^2$: A photometric stereo benchmark dataset with controlled shape and material variation. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1, 2

[15] Boxin Shi, Ping Tan, Yasuyuki Matsushita, and Katsushi Ikeuchi. Bi-polynomial modeling of low-frequency reflectances. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014. 4, 6, 11

[16] Boxin Shi, Zhipeng Mo, Zhe Wu, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation for non-Lambertian and uncalibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 1, 4, 6, 8, 9

[17] Ashish Tiwari and Shanmuganathan Raman. DeepPS2: Revisiting photometric stereo using two differently illuminated images. In *Proc. of European Conference on Computer Vision (ECCV)*, pages 129–145. Springer, 2022. 1, 5, 6, 21

[18] Robert J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 1980. 4, 6, 7

[19] Zhuokun Yao, Kun Li, Ying Fu, Haofeng Hu, and Boxin Shi. GPS-Net: Graph-based photometric stereo network. *Proc. of Annual Conference on Neural Information Processing Systems (NeurIPS)*, 33:10306–10316, 2020. 1, 5, 6, 14

[20] Qian Zheng, Yiming Jia, Boxin Shi, Xudong Jiang, Ling-Yu Duan, and Alex C. Kot. SPLINE-Net: Sparse photometric stereo through lighting interpolation and normal estimation networks. In *Proc. of International Conference on Computer Vision (ICCV)*, 2019. 1, 4, 5, 6, 16

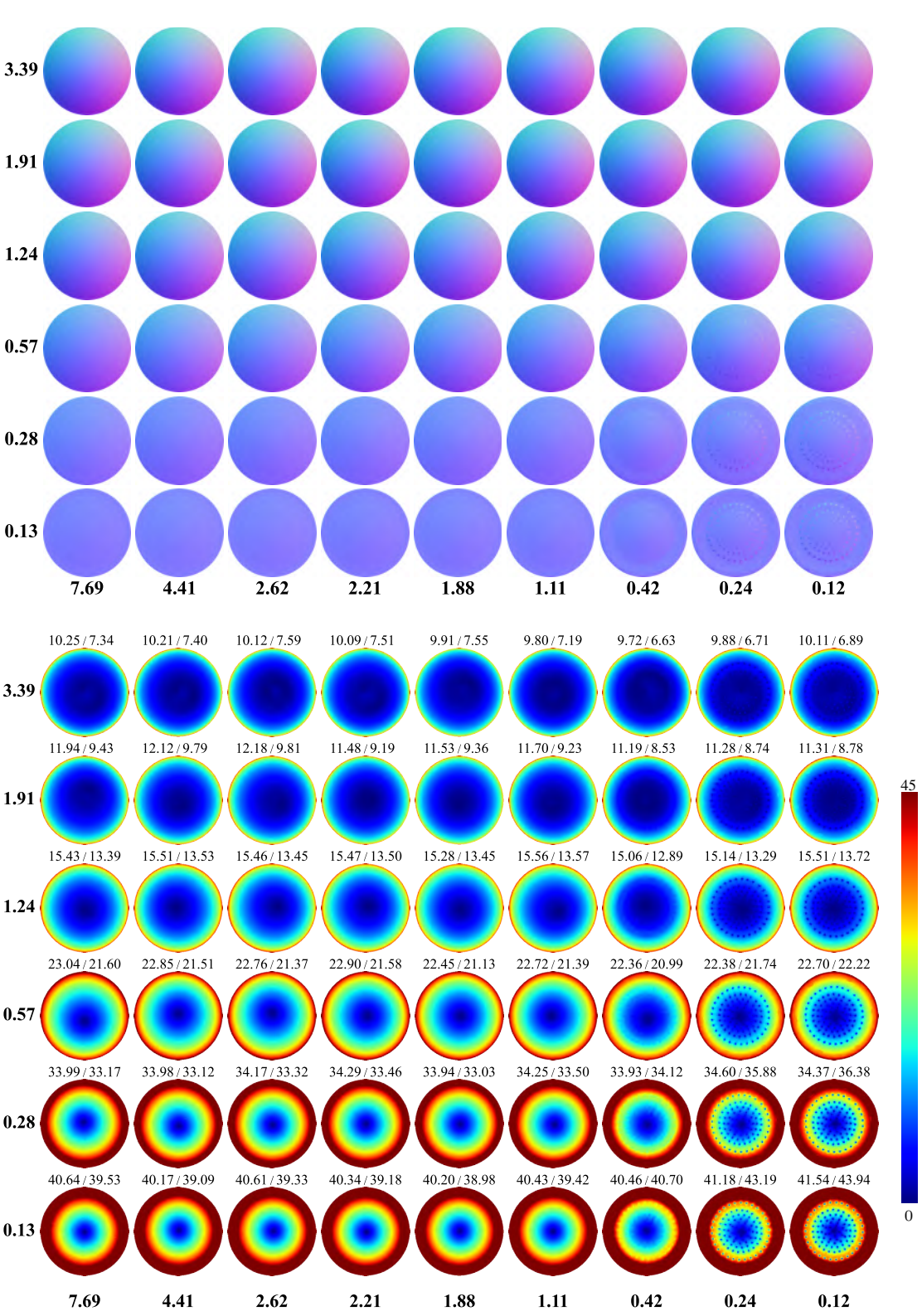



Figure S9. Estimated normal maps (top) and the corresponding angular error maps (bottom) of LS [18]. The mean and median errors for each material are displayed at the top of each error map.

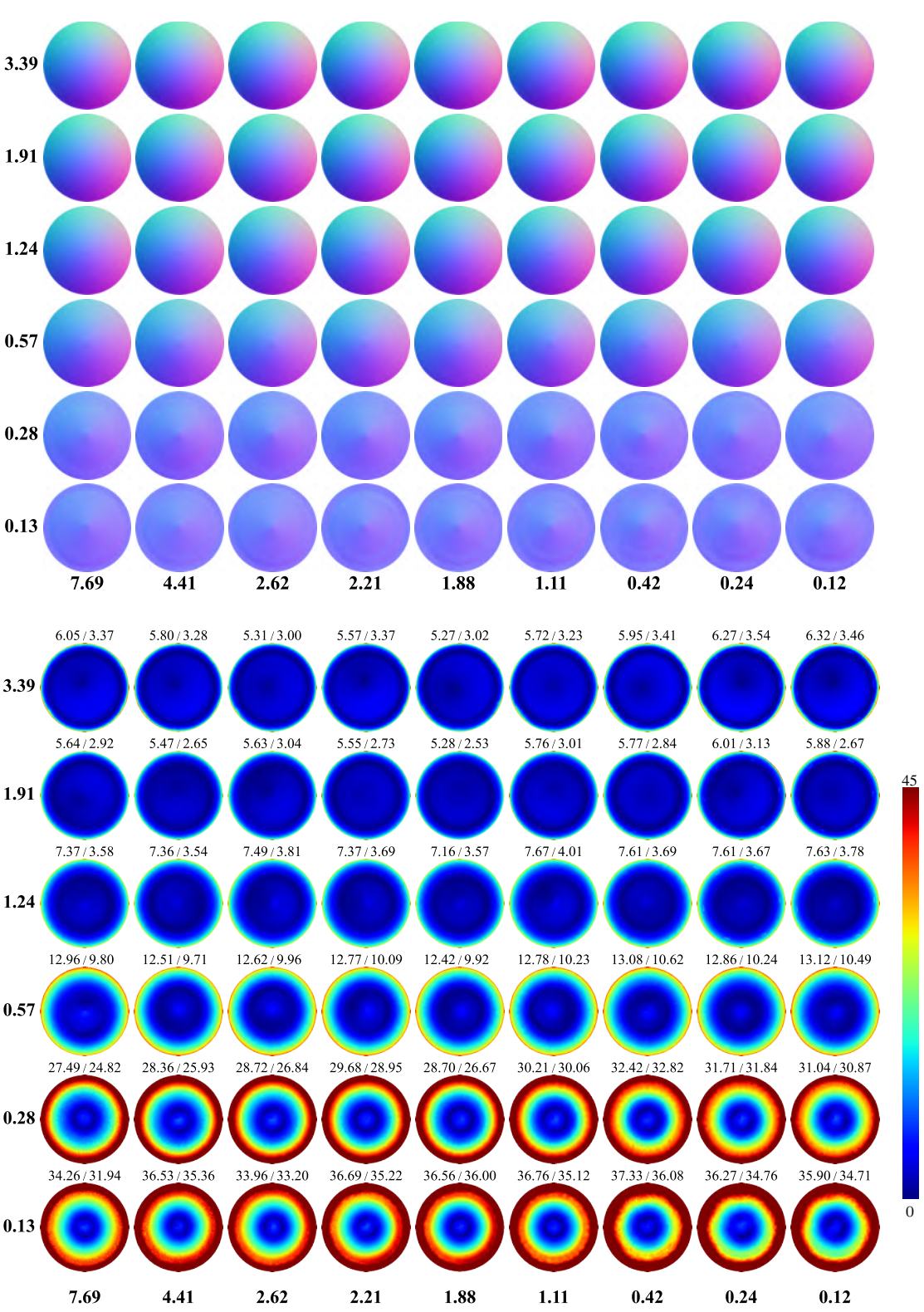



Figure S10. Estimated normal maps (top) and the corresponding angular error maps (bottom) of TH28 [16]. The mean and median errors for each material are displayed at the top of each error map.

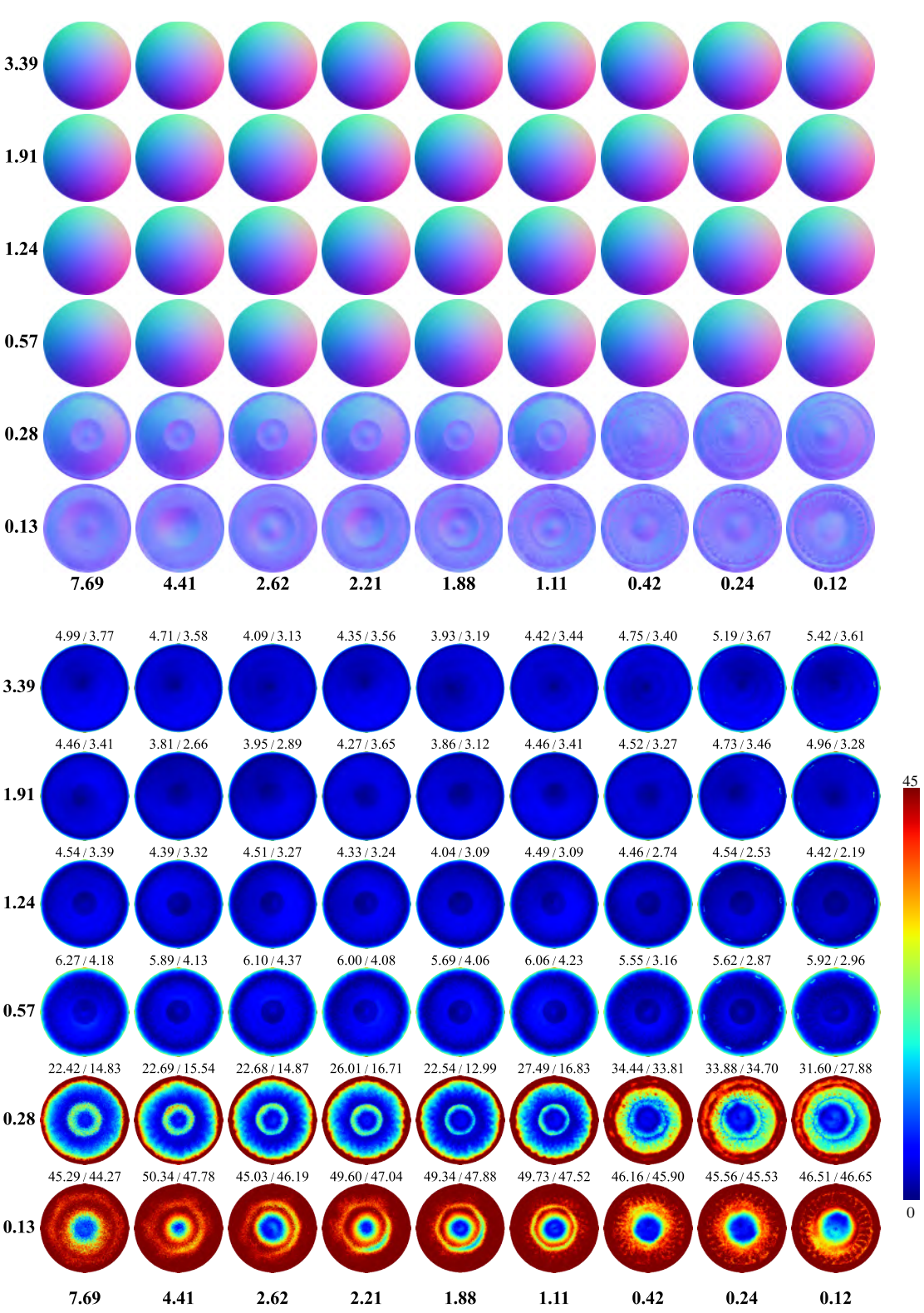



Figure S11. Estimated normal maps (top) and the corresponding angular error maps (bottom) of TH46 [16]. The mean and median errors for each material are displayed at the top of each error map.

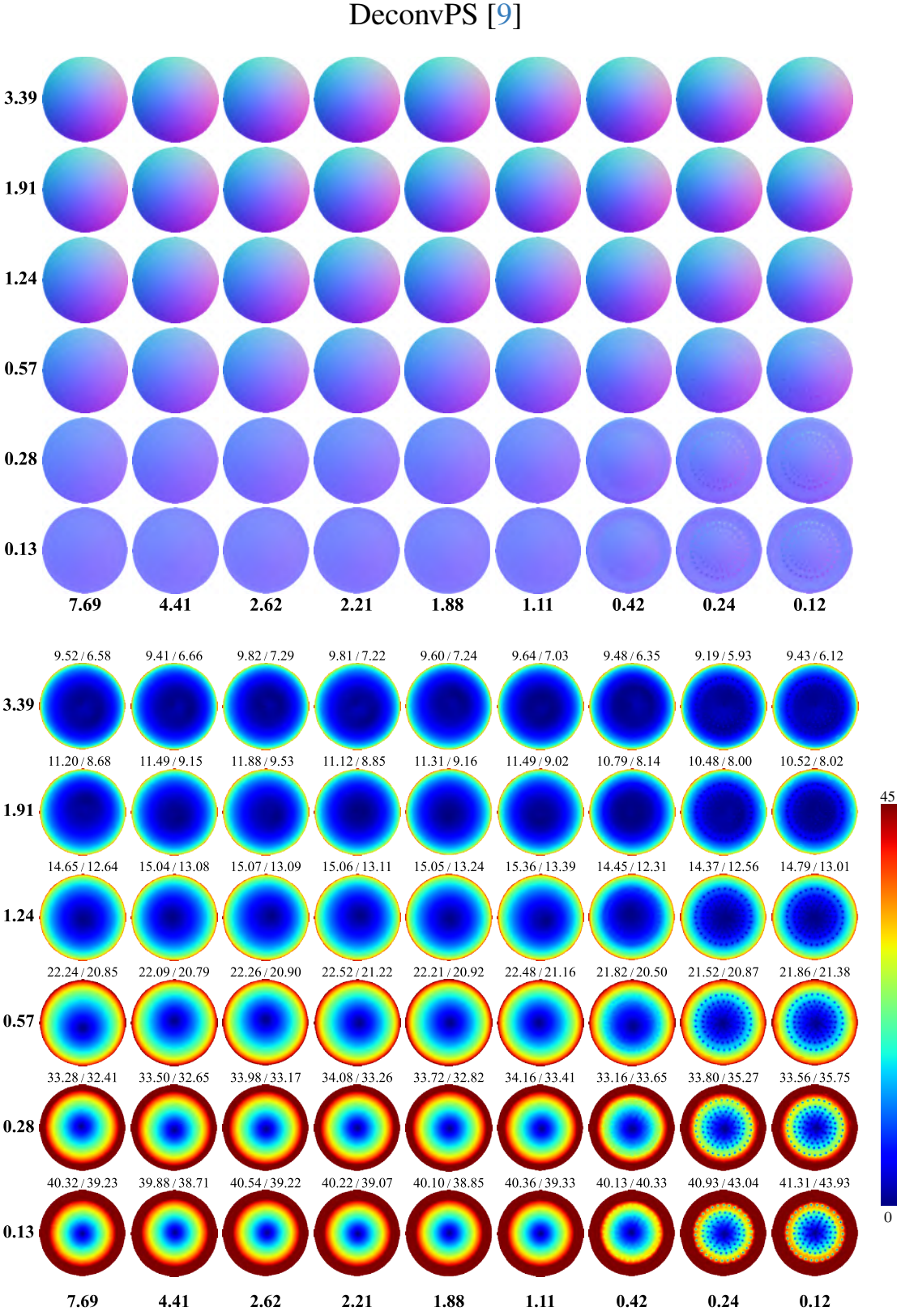



Figure S12. Estimated normal maps (top) and the corresponding angular error maps (bottom) of DeconvPS [9]. The mean and median errors for each material are displayed at the top of each error map.

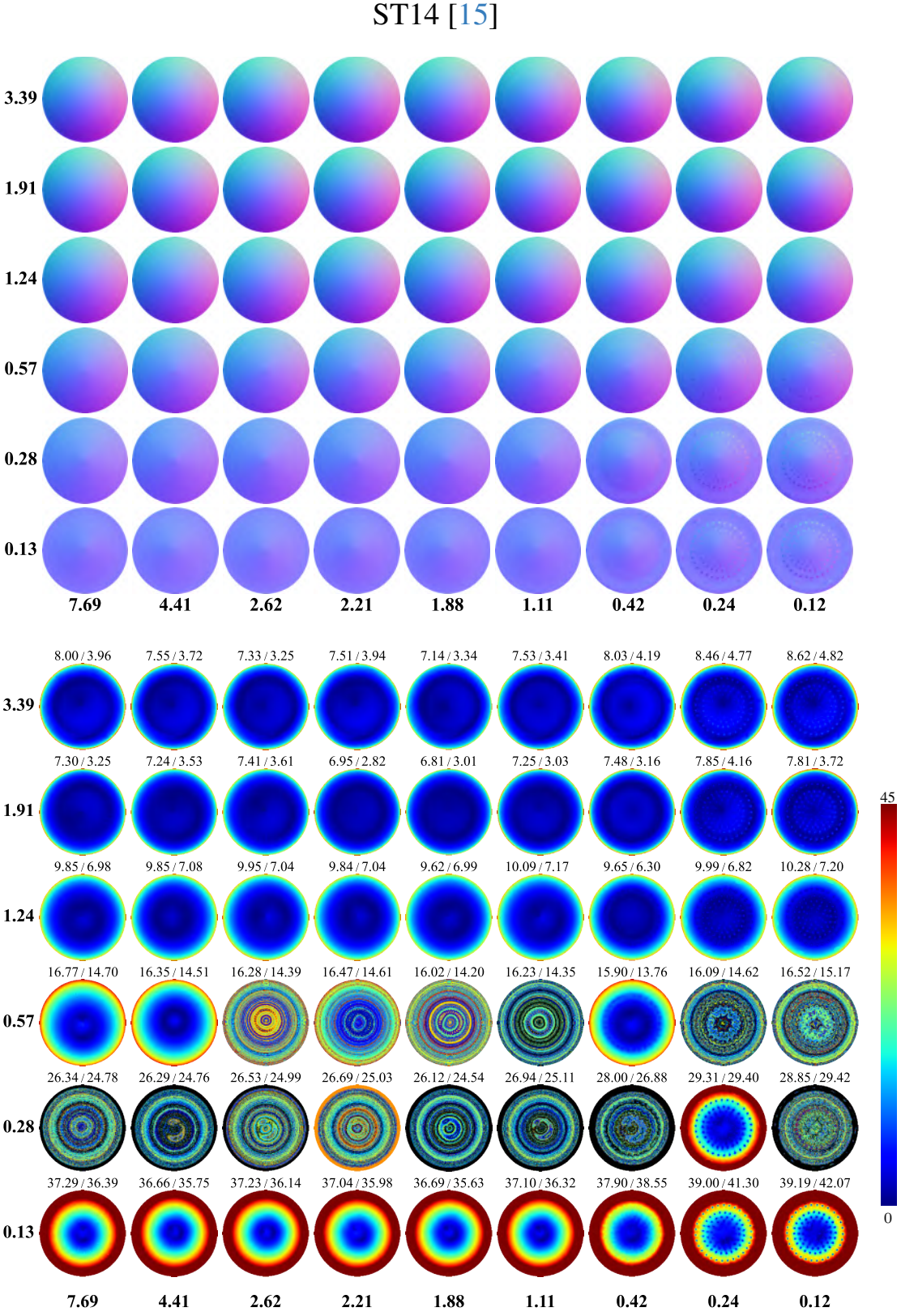



Figure S13. Estimated normal maps (top) and the corresponding angular error maps (bottom) of ST14 [15]. The mean and median errors for each material are displayed at the top of each error map.

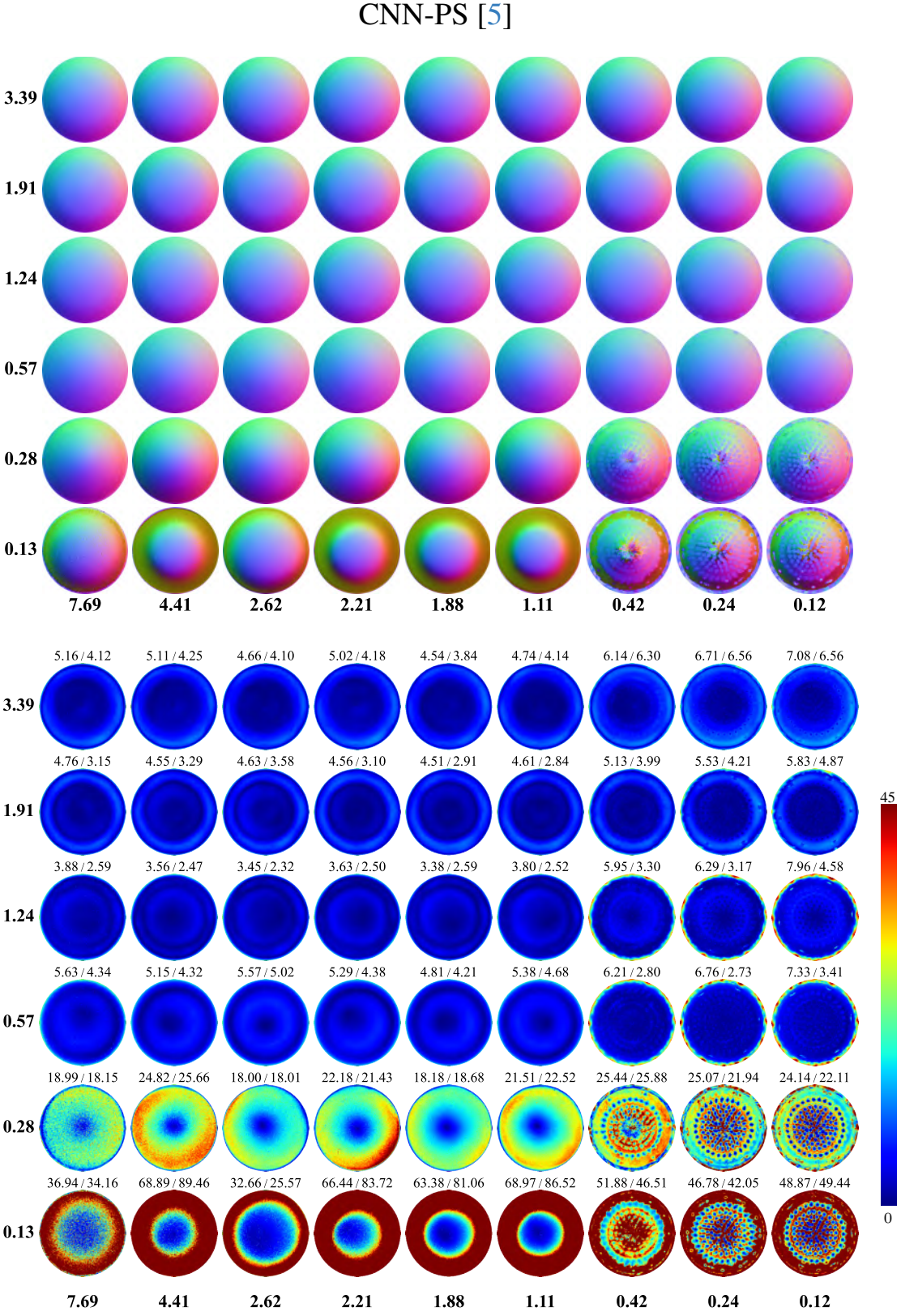



Figure S14. Estimated normal maps (top) and the corresponding angular error maps (bottom) of CNN-PS [5]. The mean and median errors for each material are displayed at the top of each error map.

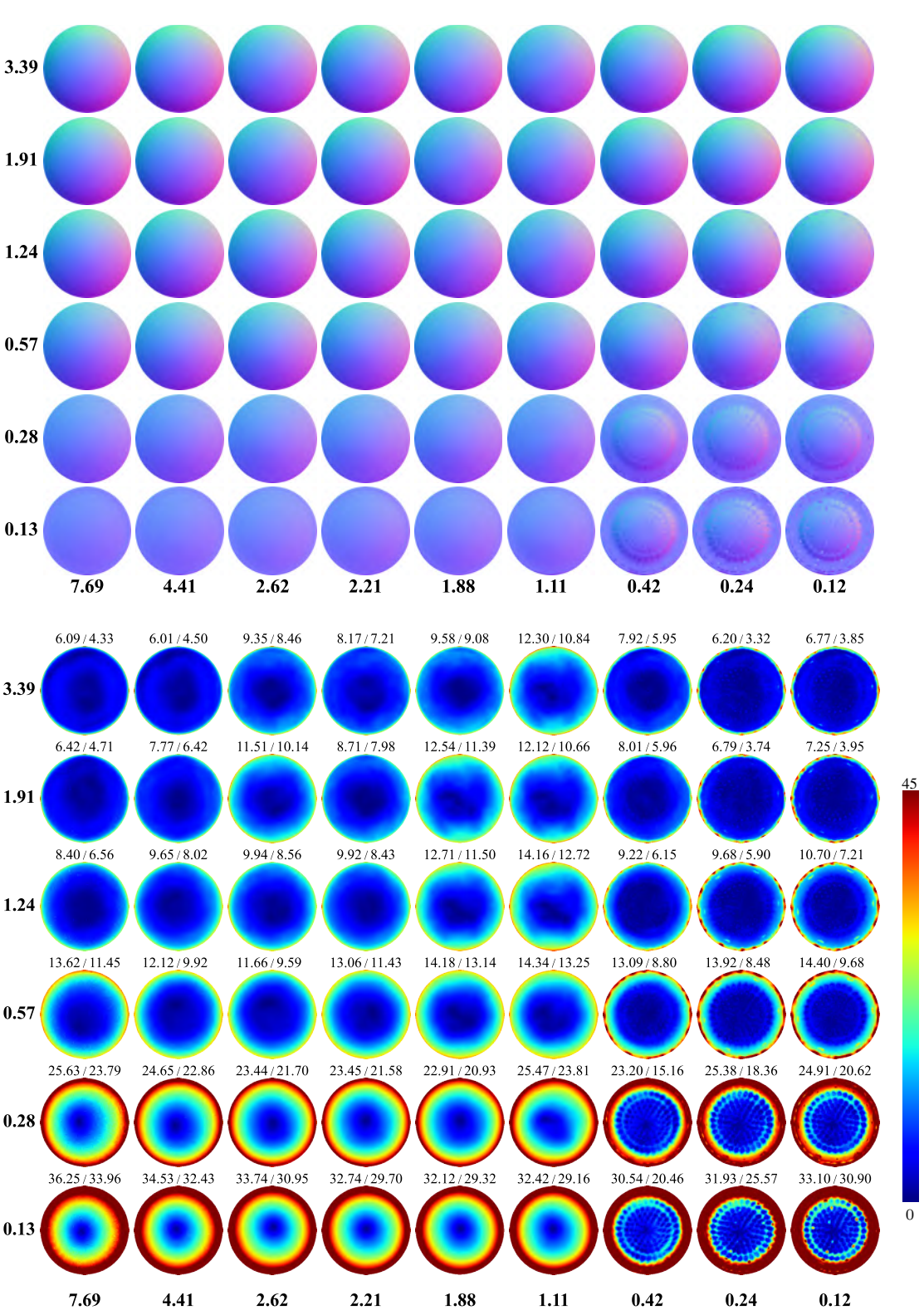



Figure S15. Estimated normal maps (top) and the corresponding angular error maps (bottom) of PS-FCN [2]. The mean and median errors for each material are displayed at the top of each error map.

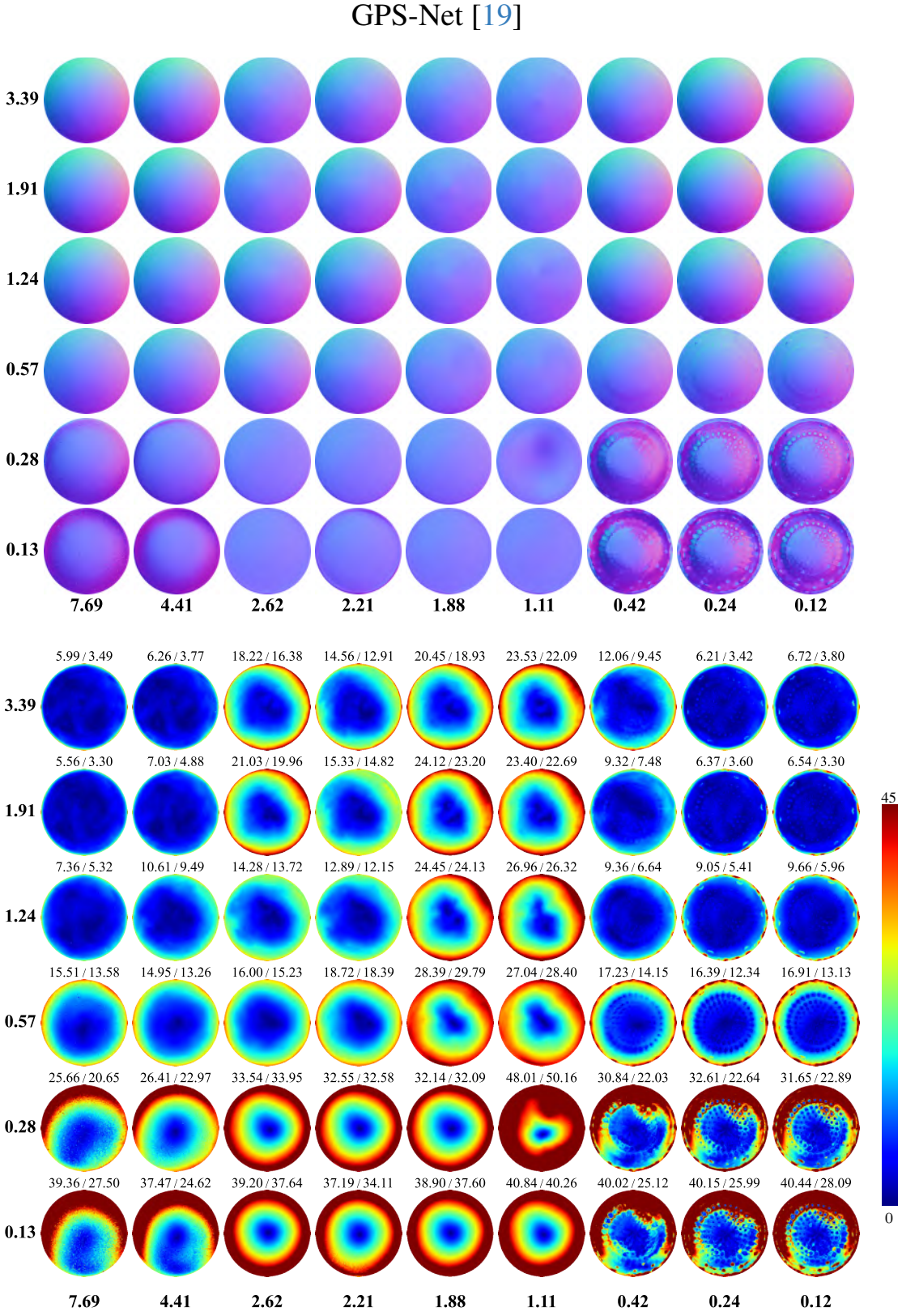



Figure S16. Estimated normal maps (top) and the corresponding angular error maps (bottom) of GPS-Net [19]. The mean and median errors for each material are displayed at the top of each error map.

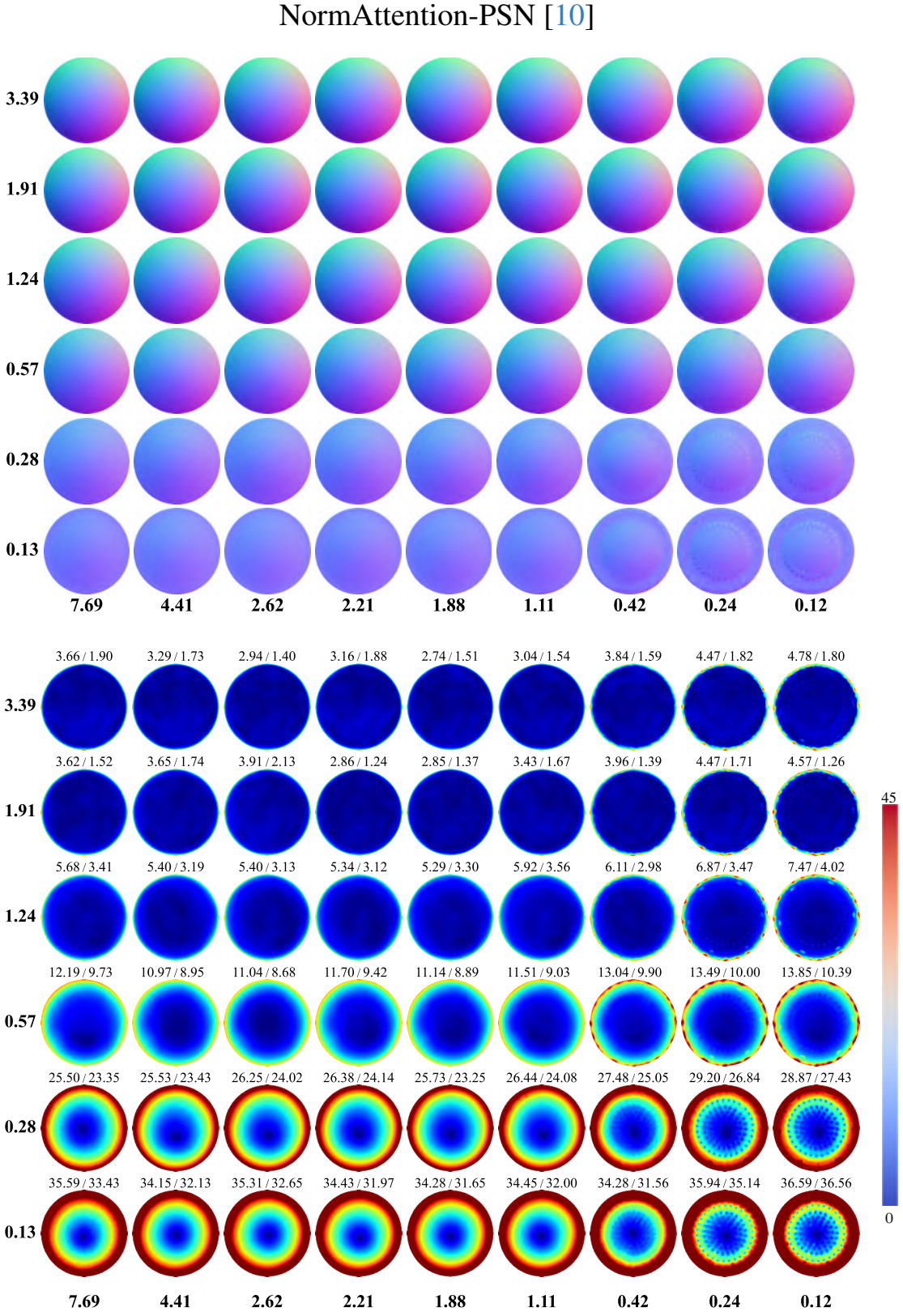



Figure S17. Estimated normal maps (top) and the corresponding angular error maps (bottom) of NormAttention-PSN [10]. The mean and median errors for each material are displayed at the top of each error map.

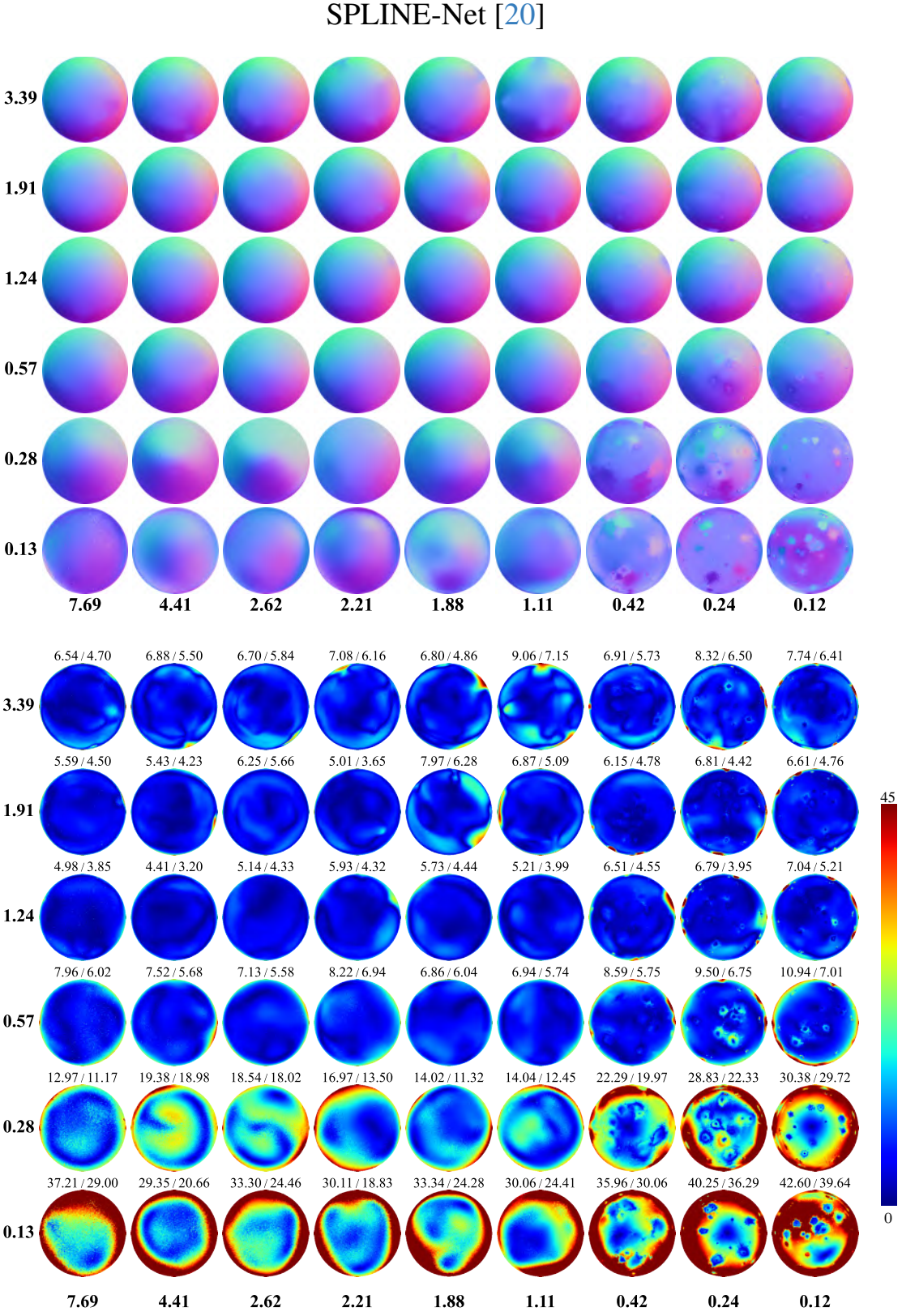



Figure S18. Estimated normal maps (top) and the corresponding angular error maps (bottom) of SPLINE-Net [20]. The mean and median errors for each material are displayed at the top of each error map.

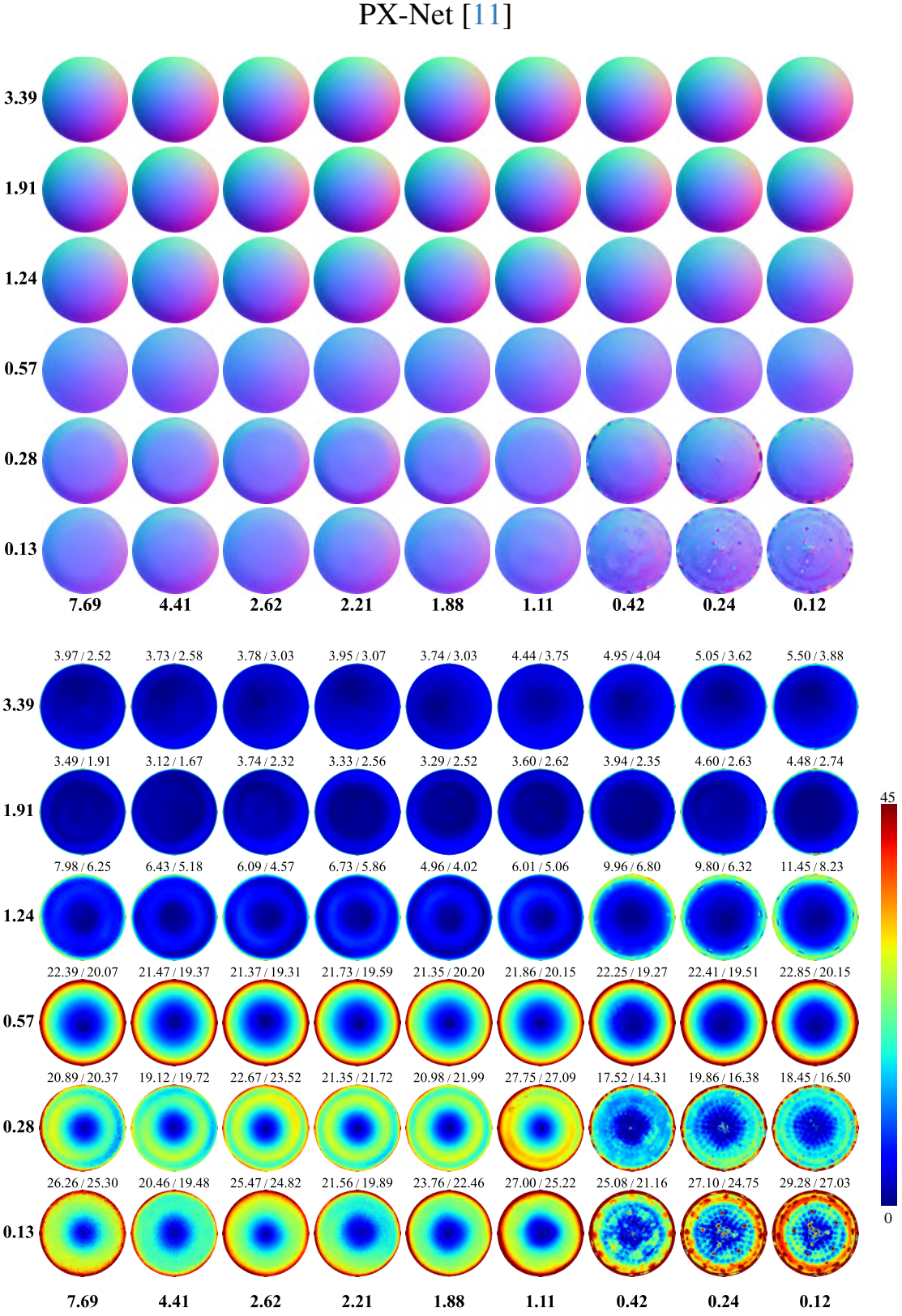



Figure S19. Estimated normal maps (top) and the corresponding angular error maps (bottom) of PX-Net [11]. The mean and median errors for each material are displayed at the top of each error map.

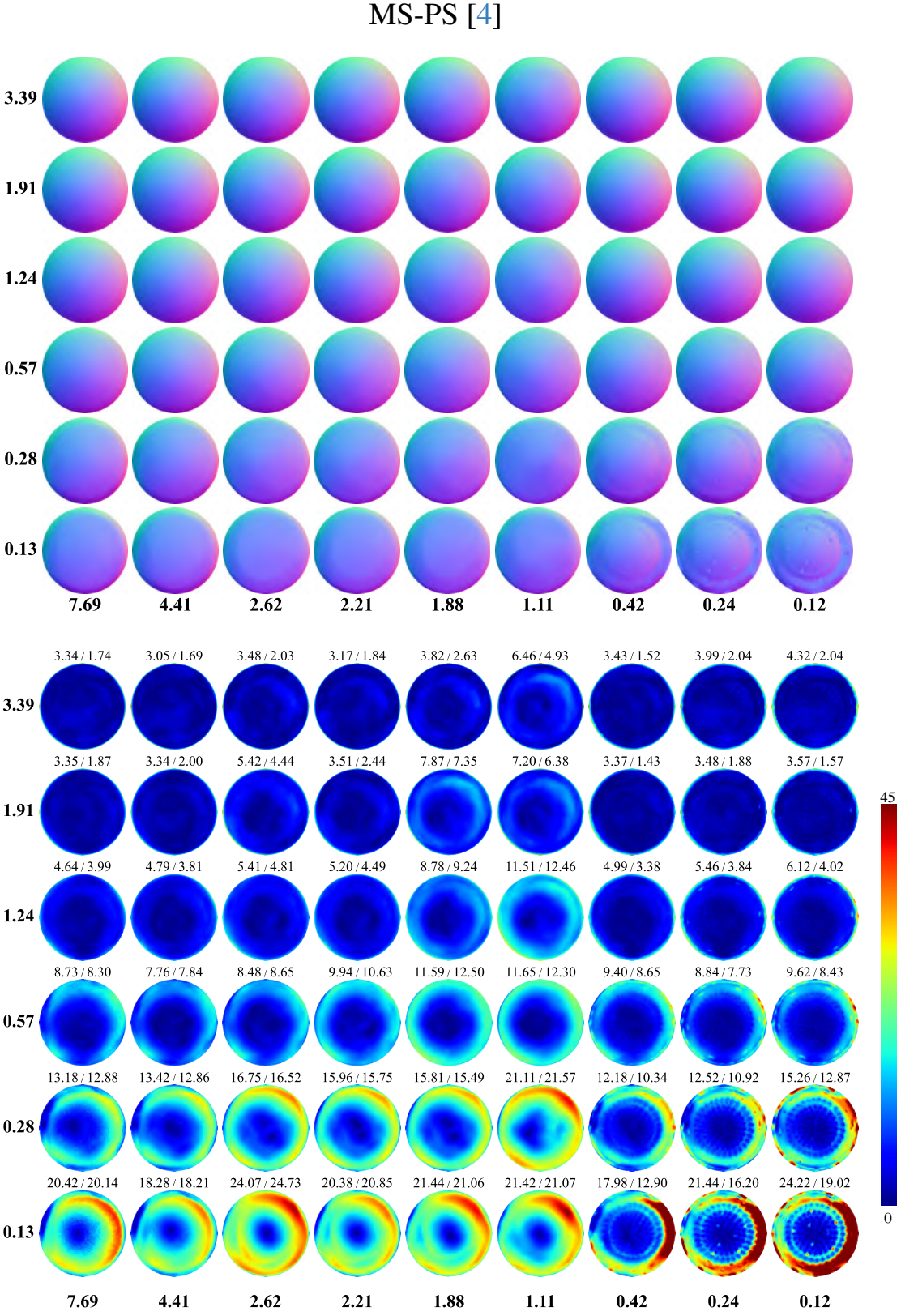



Figure S20. Estimated normal maps (top) and the corresponding angular error maps (bottom) of MS-PS [4]. The mean and median errors for each material are displayed at the top of each error map.

PF14 [12]

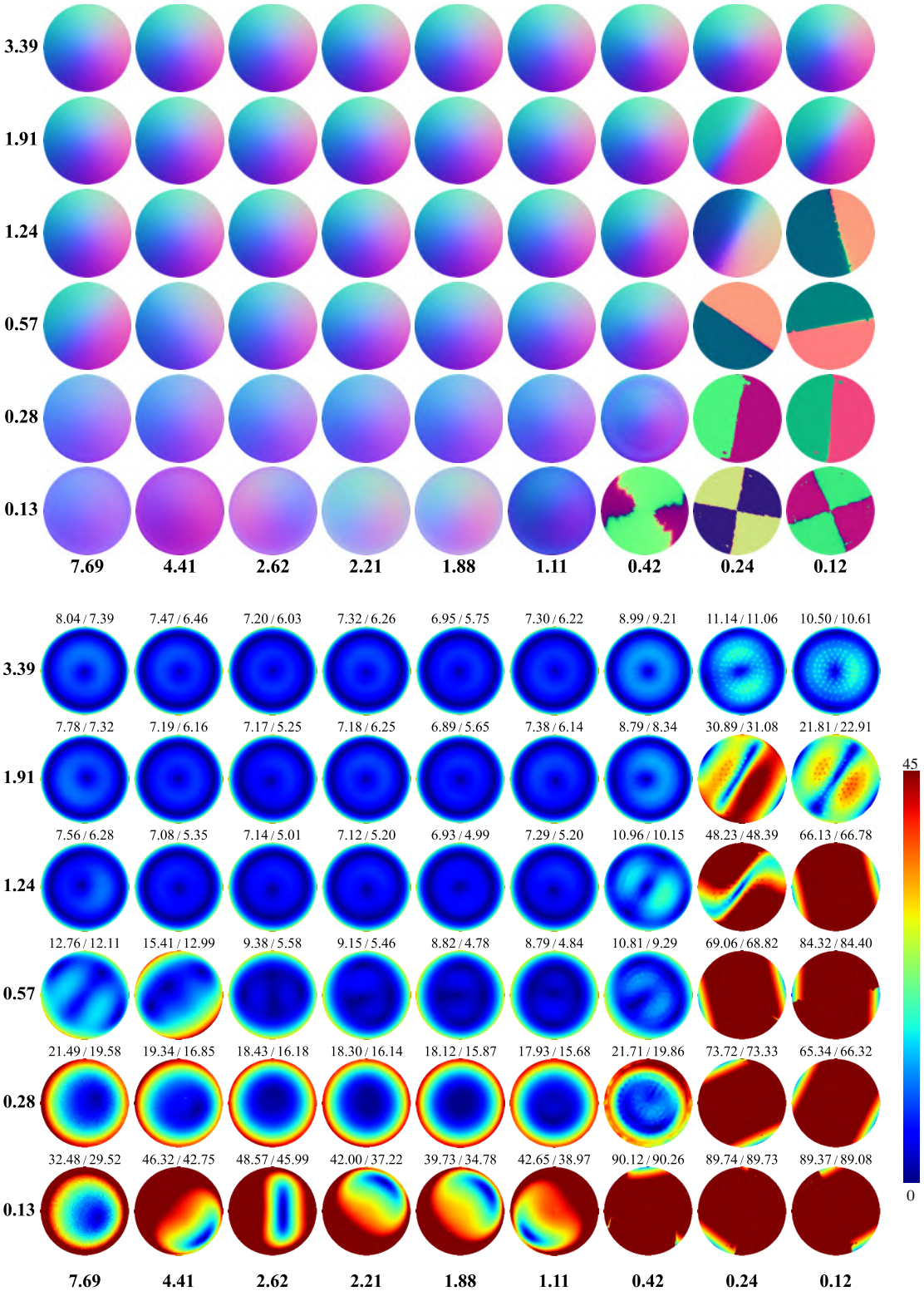



Figure S21. Estimated normal maps (top) and the corresponding angular error maps (bottom) of PF14 [12]. The mean and median errors for each material are displayed at the top of each error map.

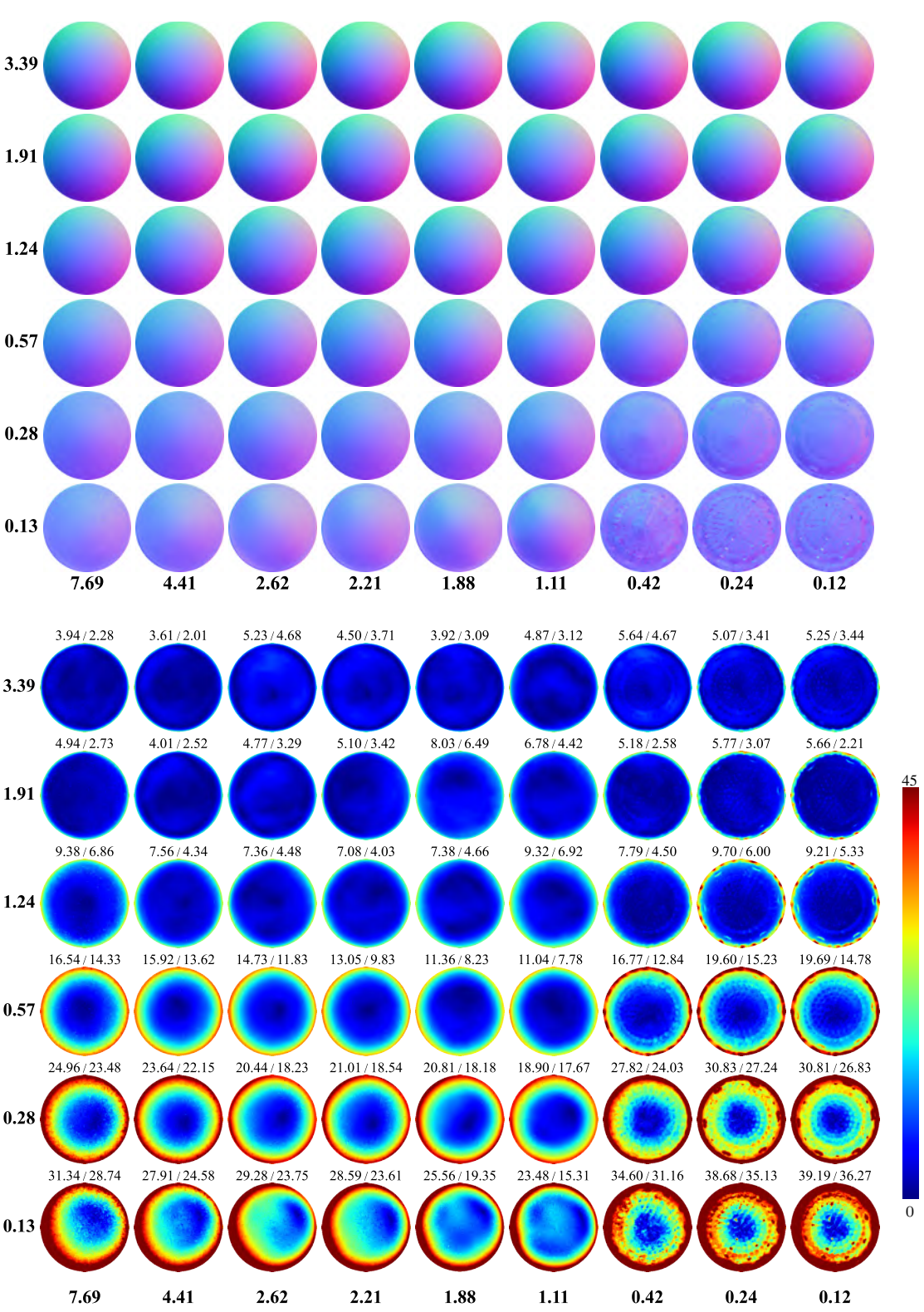



Figure S22. Estimated normal maps (top) and the corresponding angular error maps (bottom) of SDPS-Net [3]. The mean and median errors for each material are displayed at the top of each error map.

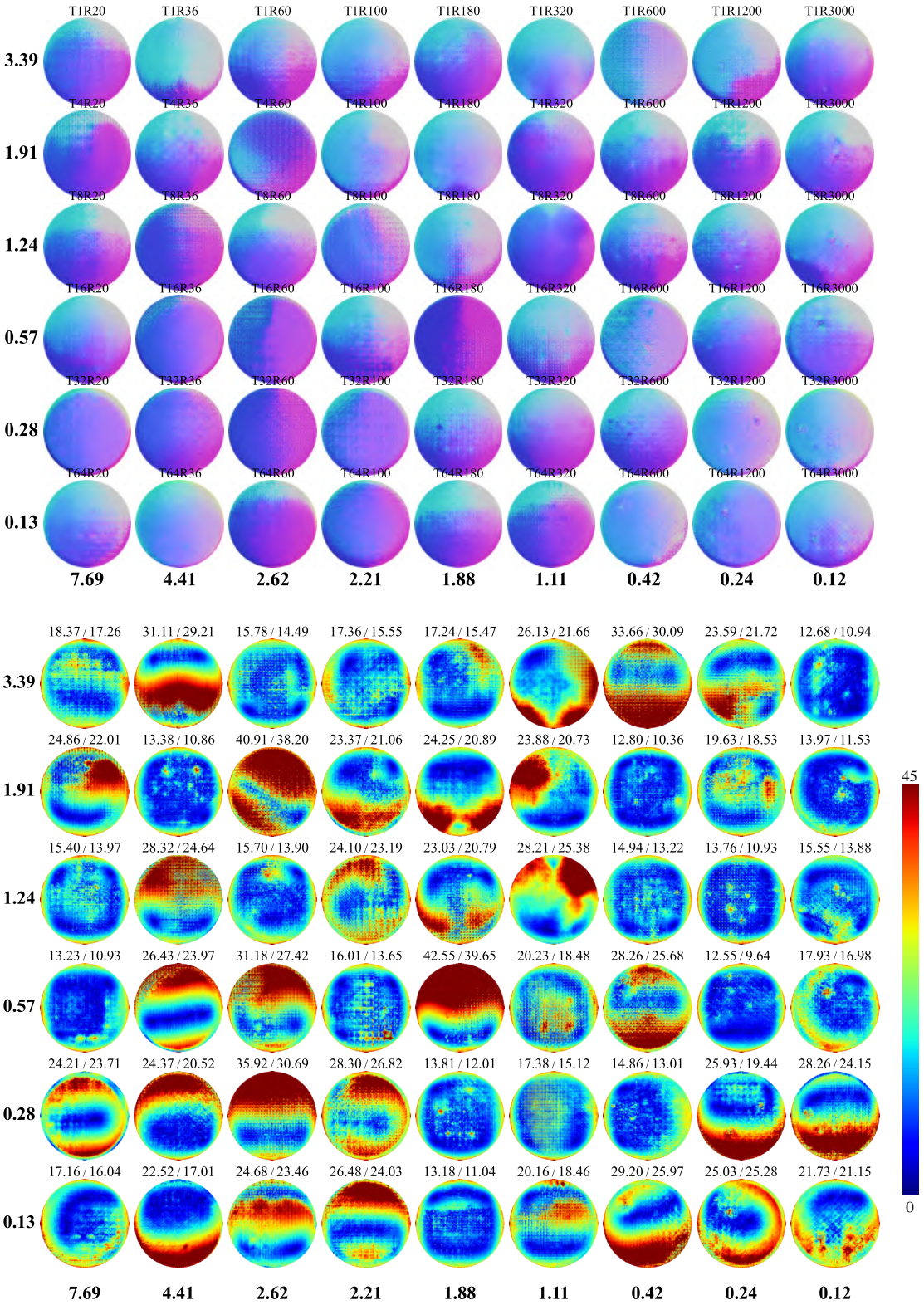



Figure S23. Estimated normal maps (top) and the corresponding angular error maps (bottom) of DeepPS2 [17]. The mean and median errors for each material are displayed at the top of each error map.

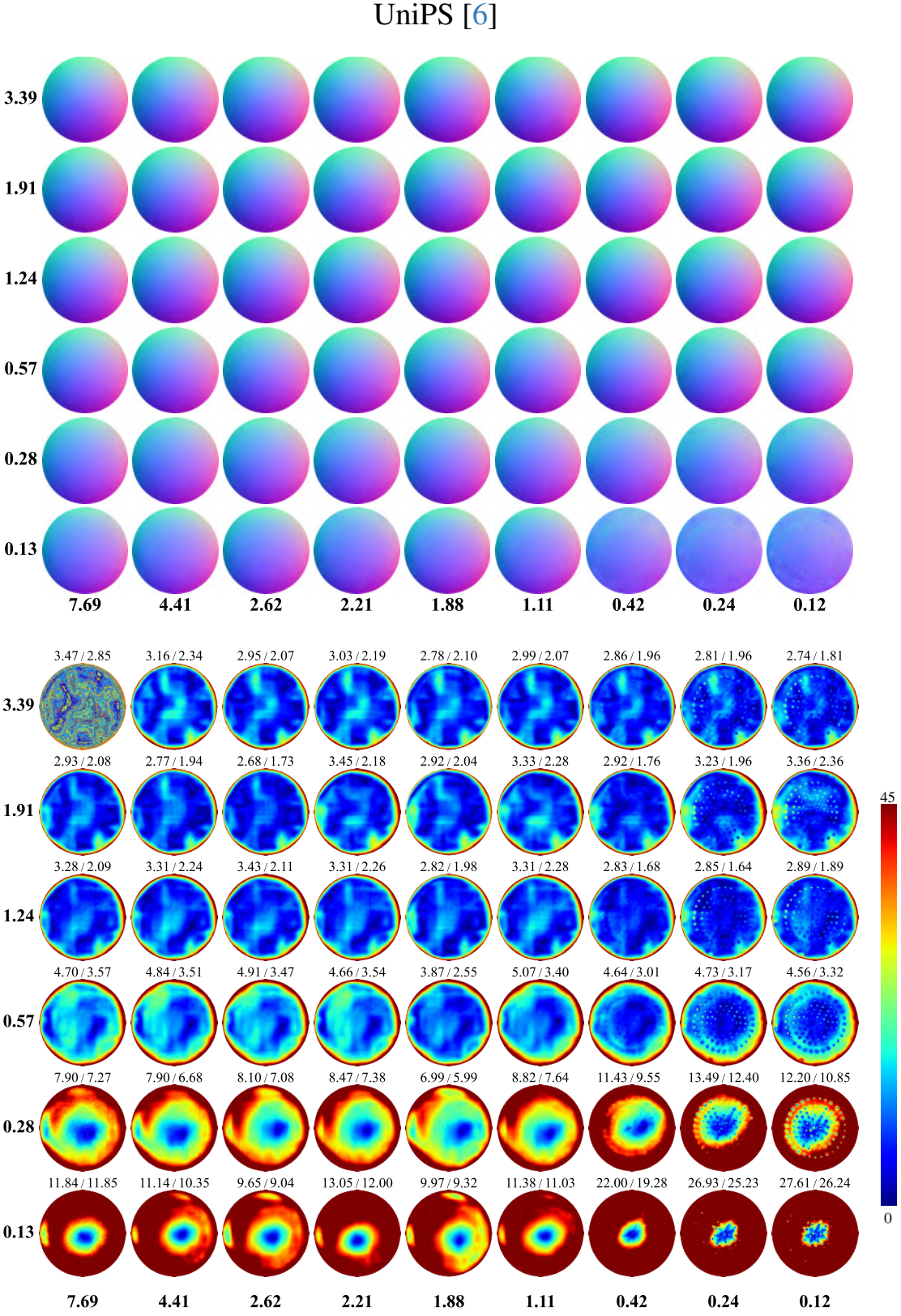



Figure S24. Estimated normal maps (top) and the corresponding angular error maps (bottom) of UniPS [6]. The mean and median errors for each material are displayed at the top of each error map.

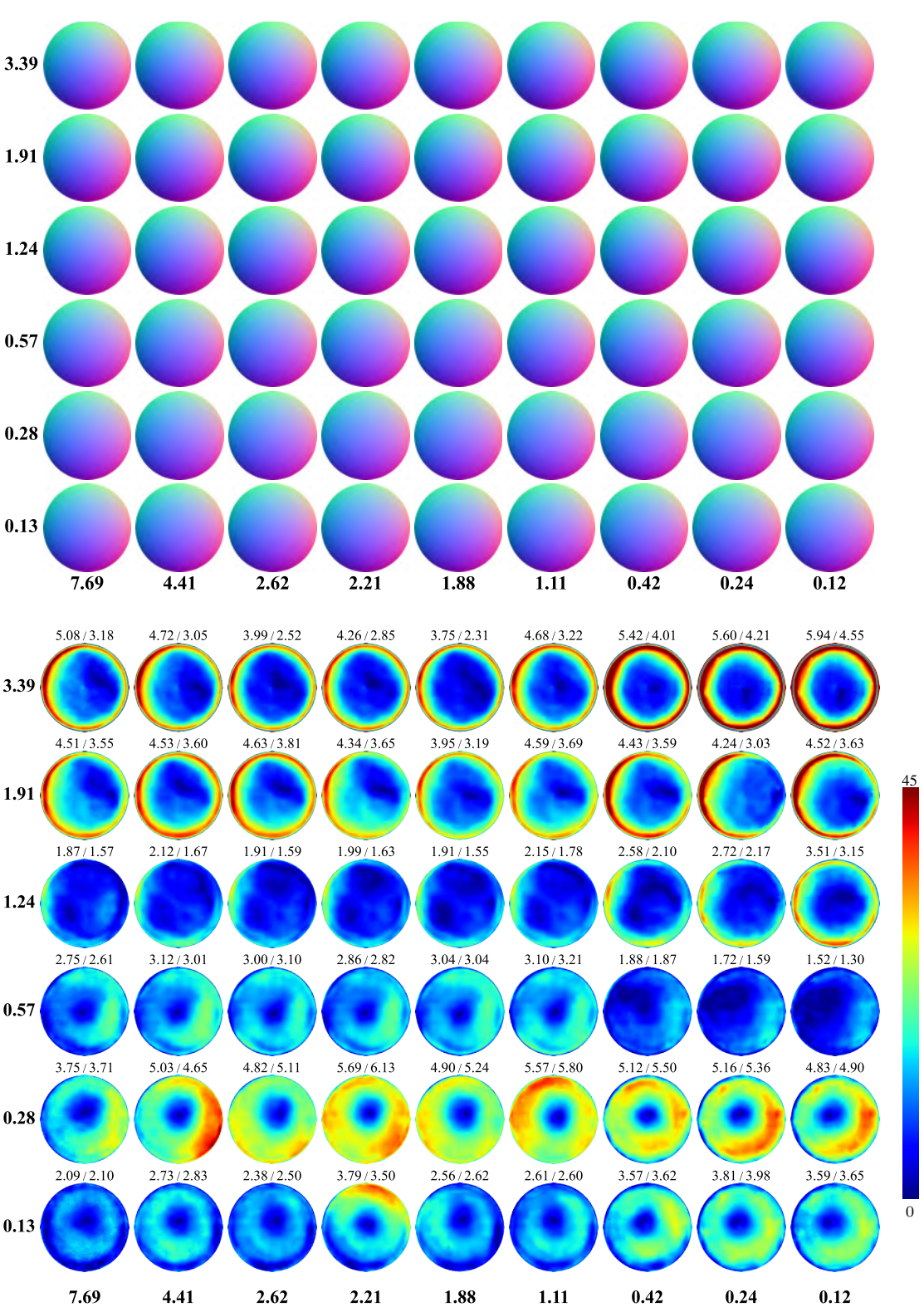



Figure S25. Estimated normal maps (top) and the corresponding angular error maps (bottom) of SDM-UniPS [7]. The mean and median errors for each material are displayed at the top of each error map.