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:

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

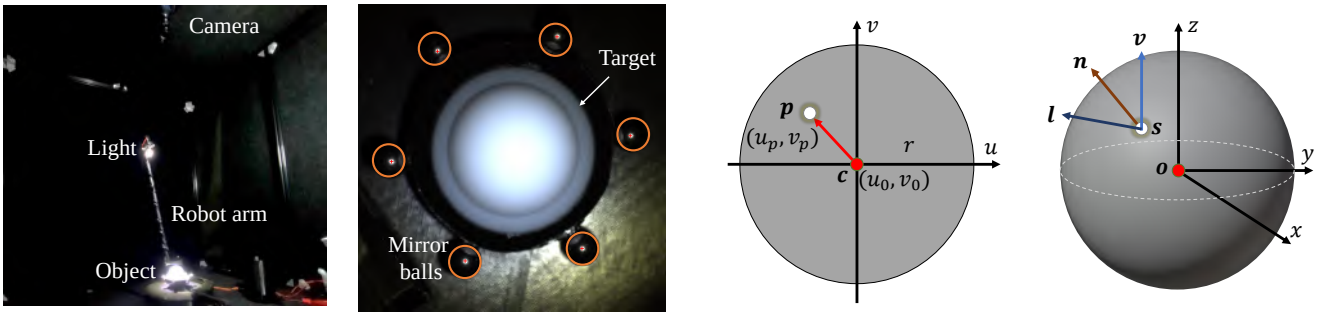
where ϵ_{xi} and ϵ_{yi} denote the Gaussian noise terms. Using the principle of maximum likelihood inference, we perform least squares optimization to obtain the optimal values of n_x and n_y:

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

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.

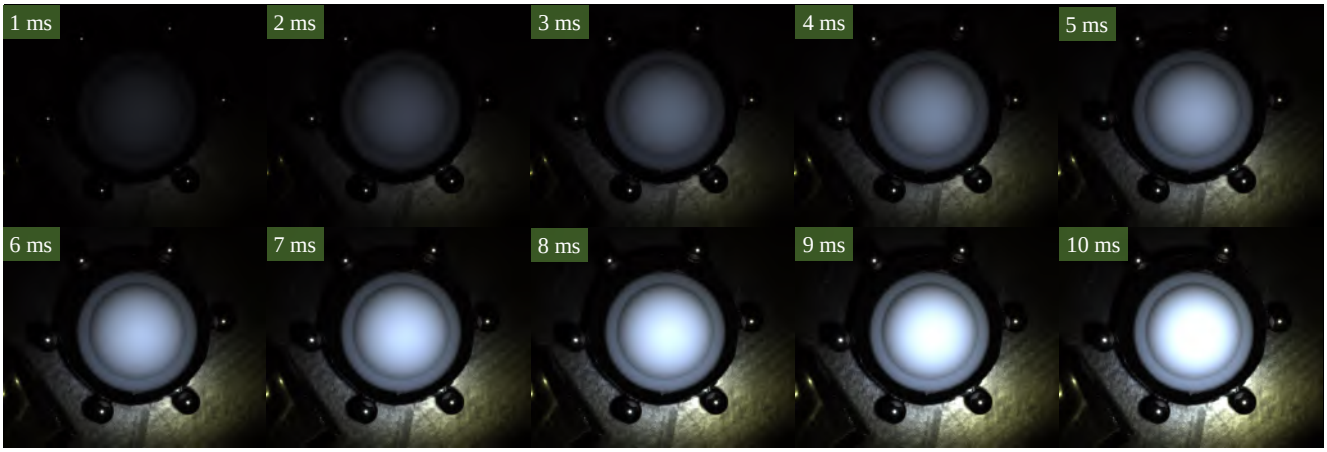


Figure S2. The target scene is captured under multiple exposure times from 1ms to 10ms to composite an HDR image.

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 2448×2048, as shown in Fig. S1. We crop the images to 960×960 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

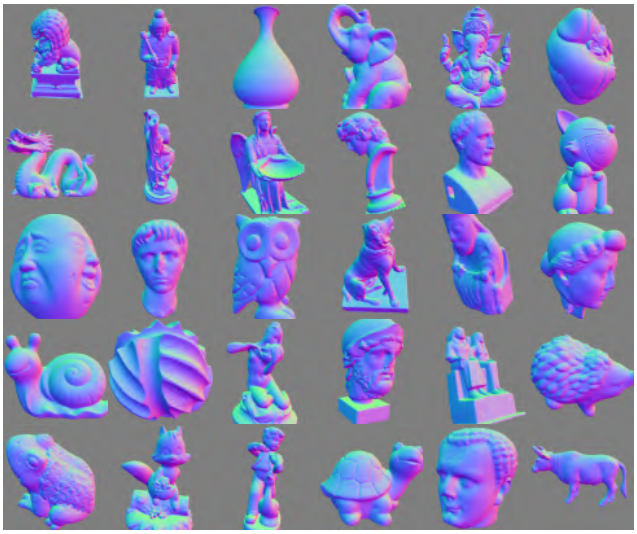


Figure S3. Subset of the 3000 surface normals of our synthetic dataset PS-SSS.

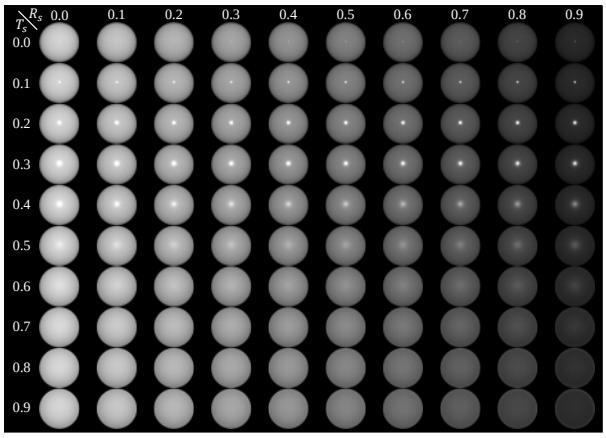


Figure S4. Synthetic dataset PS-Sphere indexed by the roughness R_s and transmittance T_s defined in the Principled BSDF model.

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.

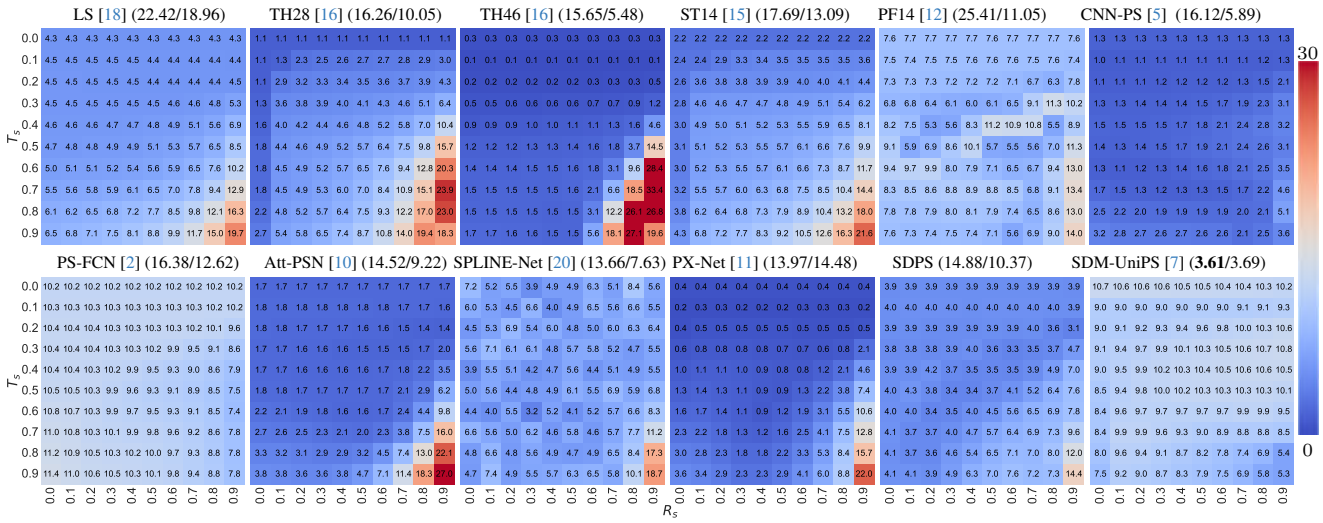


Figure S5. Roughness-translucency MAE matrices for 16 photometric stereo methods, where the ticks of row and column are σ_t and Sa. The mean and median of the MAE matrix are presented near the method name, showing method's performance profile under different level of reflectance properties. 'Att-PSN'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

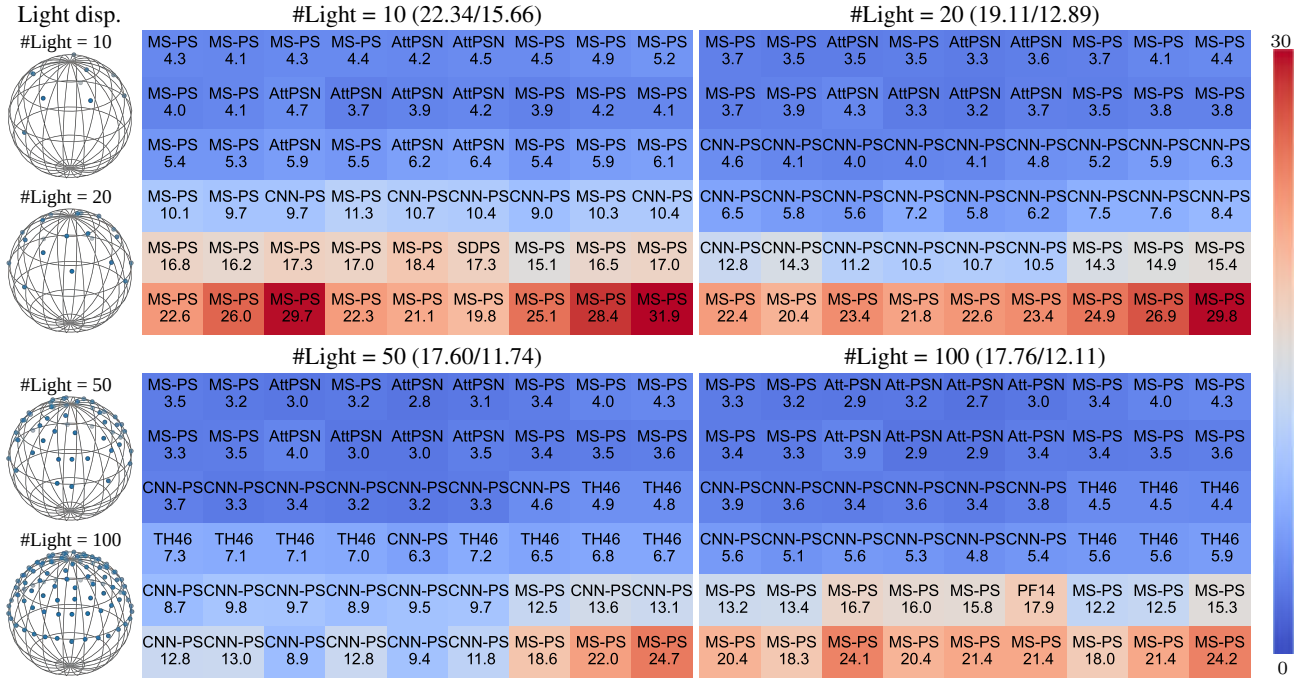


Figure S6. Benchmark evaluation on DiLiGenRT under different number of lights (#10 to #100) distributed uniformly, summarized by mean/median MAE values. Each cell records the best performing method for the material in that cell, along with the associated lowest MAE value.

achieves much smaller MAE on surfaces whose translucency measurement (σ_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

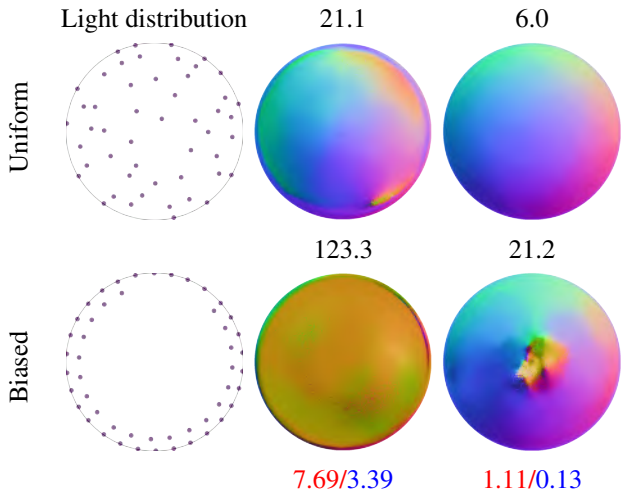


Figure S7. The influence of light distribution on surface normal estimation, tested on two spheres whose roughness and translucency measured by Sa/σ_t are shown at the bottom. The normal estimation error measured by MAE is displayed at the top of the surface normal estimates.

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 6×9 matrix of their estimated surface normal map alongside their corresponding angular error distribution map. For better visualization, the maximal MAE is truncated at 10° for UniPS [6] and SDM-UniPS [7] and 45° for other methods. The x and y axes in the matrix denote the translucency and roughness measurements, measured by σ_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

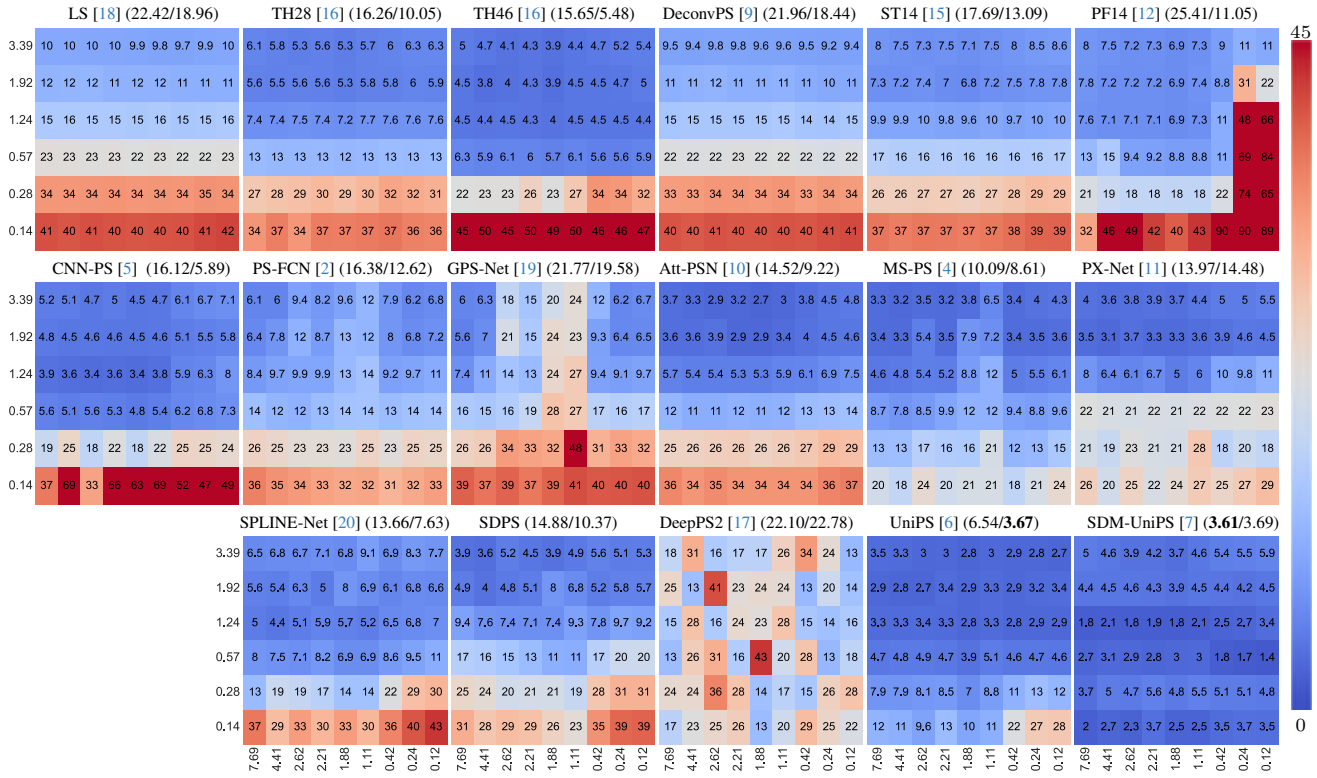


Figure S8. Roughness-translucency MAE matrices for 16 photometric stereo methods, where the ticks of row and column are σ_t and Sa. The mean and median of the MAE matrix are presented near the method name, showing method's performance profile under different level of reflectance properties. 'Att-PSN' 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

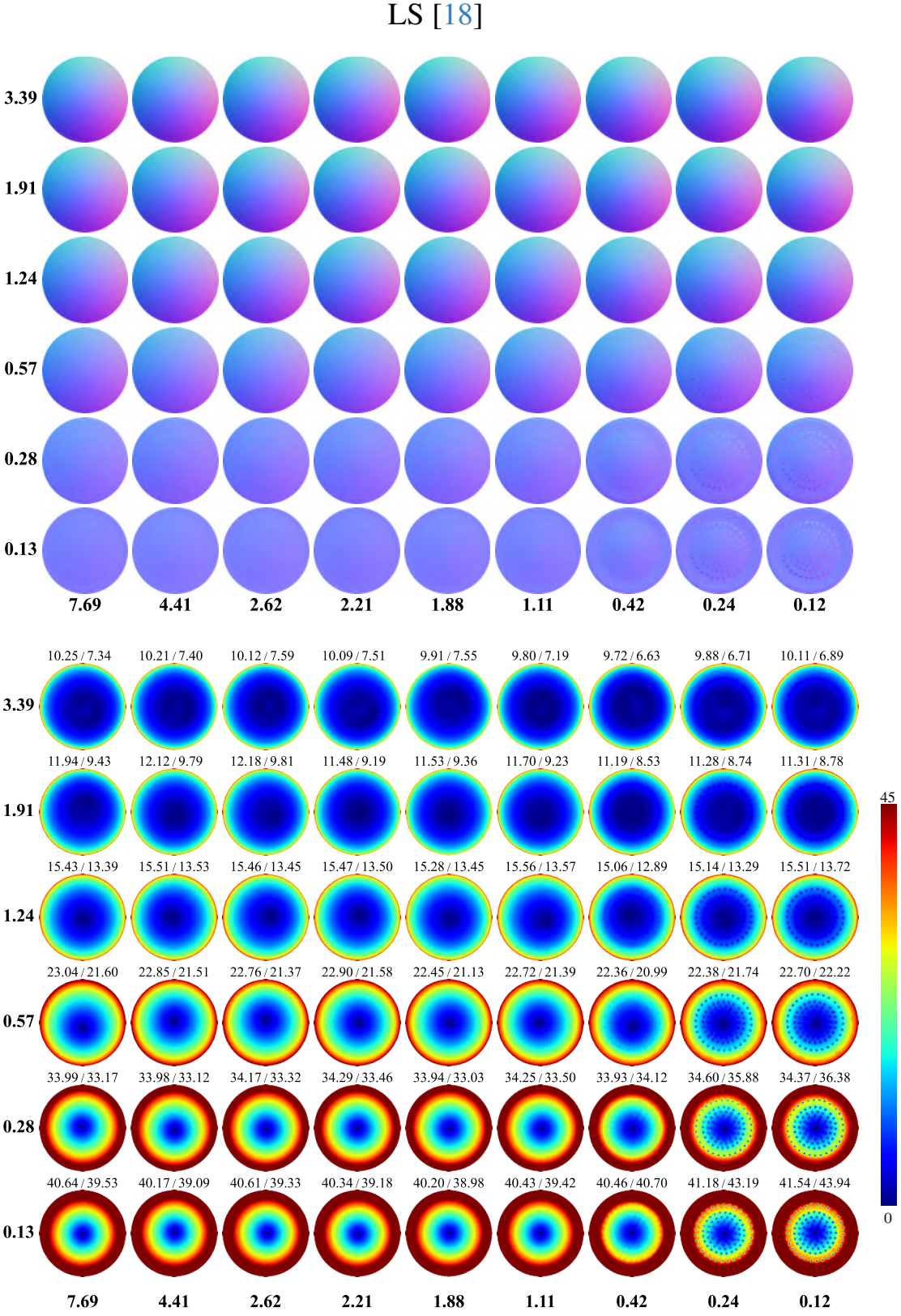


Figure S9. Estimated normal maps (top) and the corresponding angular error maps (bottom) of LS [18]. The mean and median errors for each material are displayed at the top of each error map.

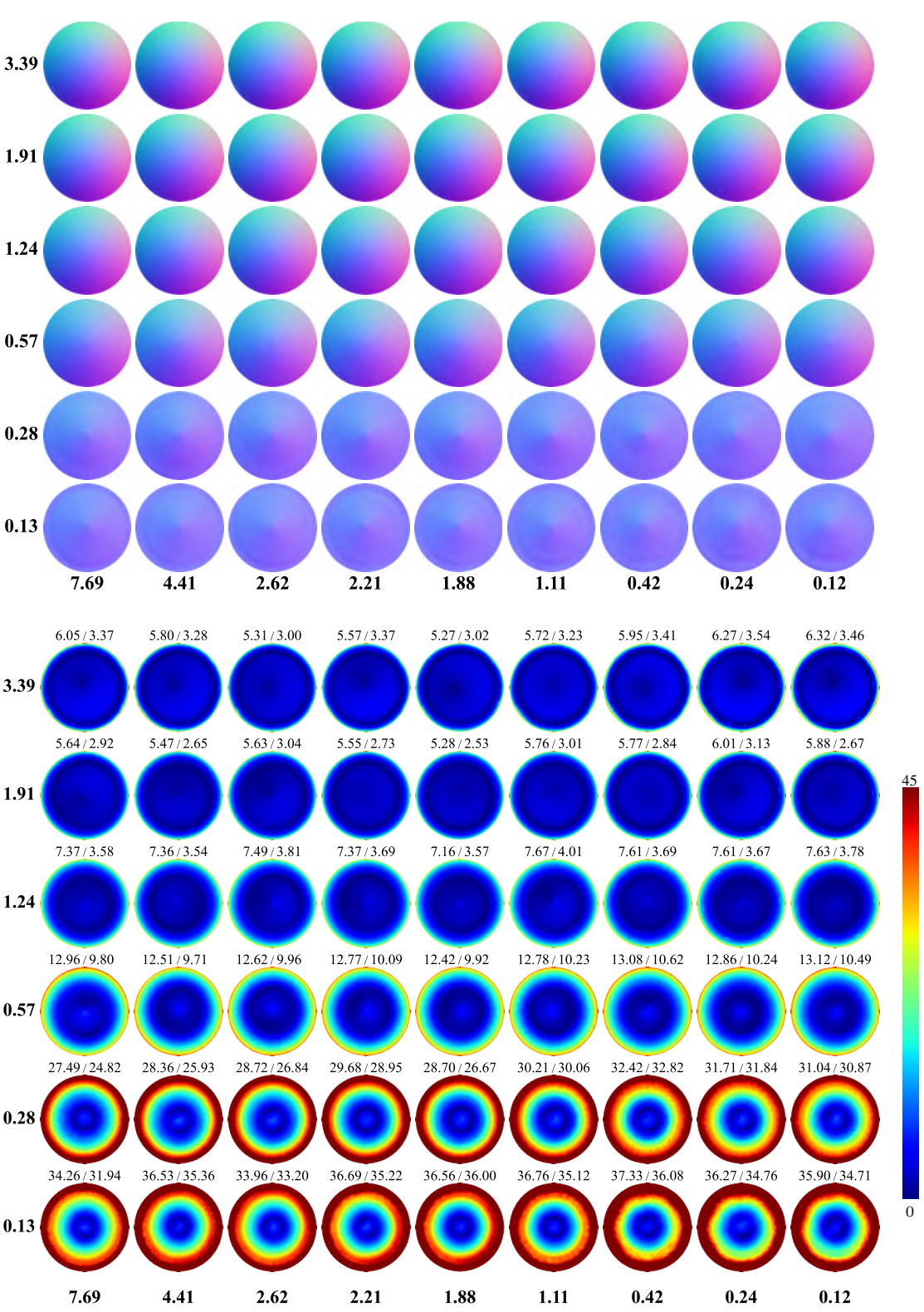


Figure S10. Estimated normal maps (top) and the corresponding angular error maps (bottom) of TH28 [16]. The mean and median errors for each material are displayed at the top of each error map.

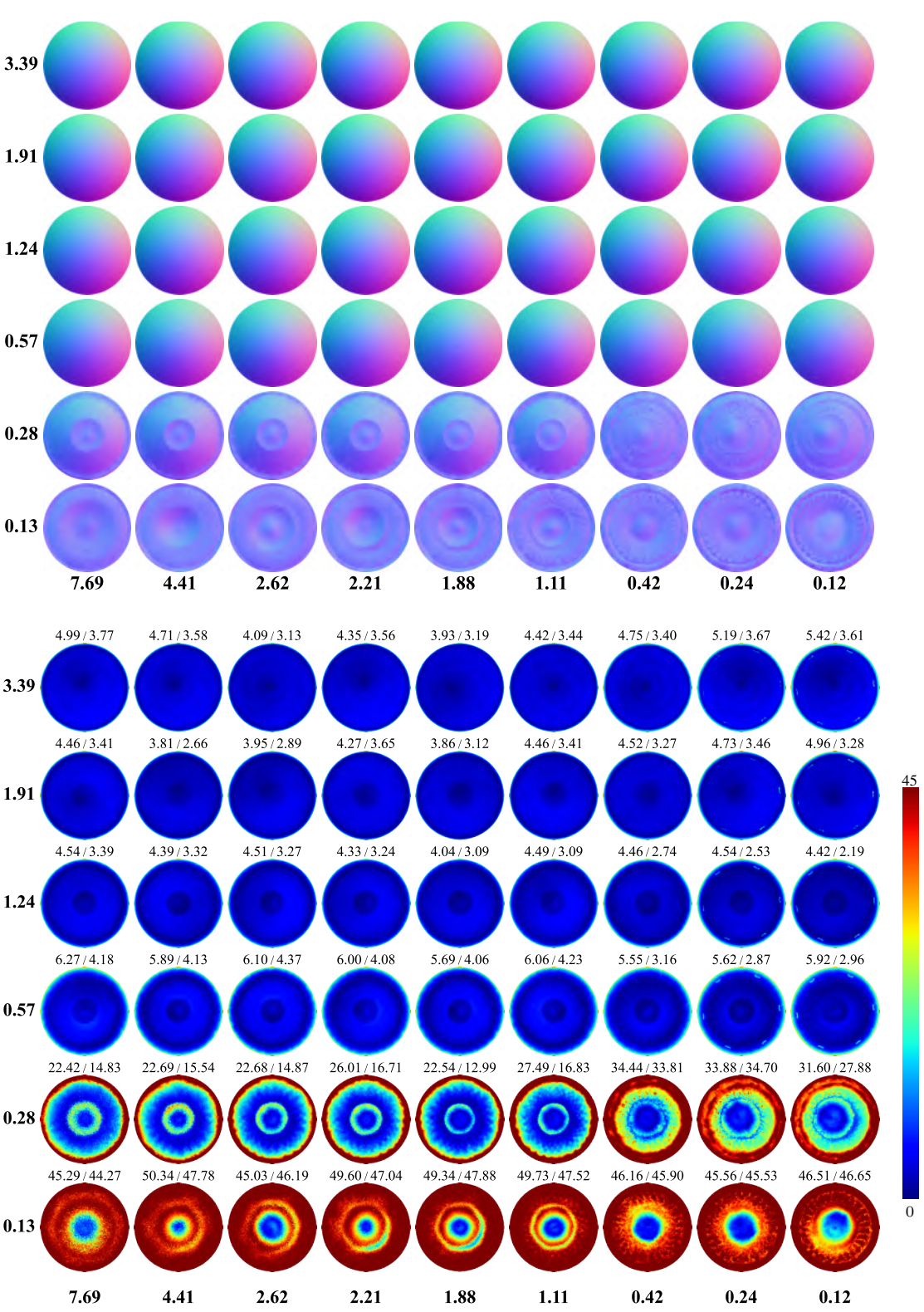


Figure S11. Estimated normal maps (top) and the corresponding angular error maps (bottom) of TH46 [16]. The mean and median errors for each material are displayed at the top of each error map.

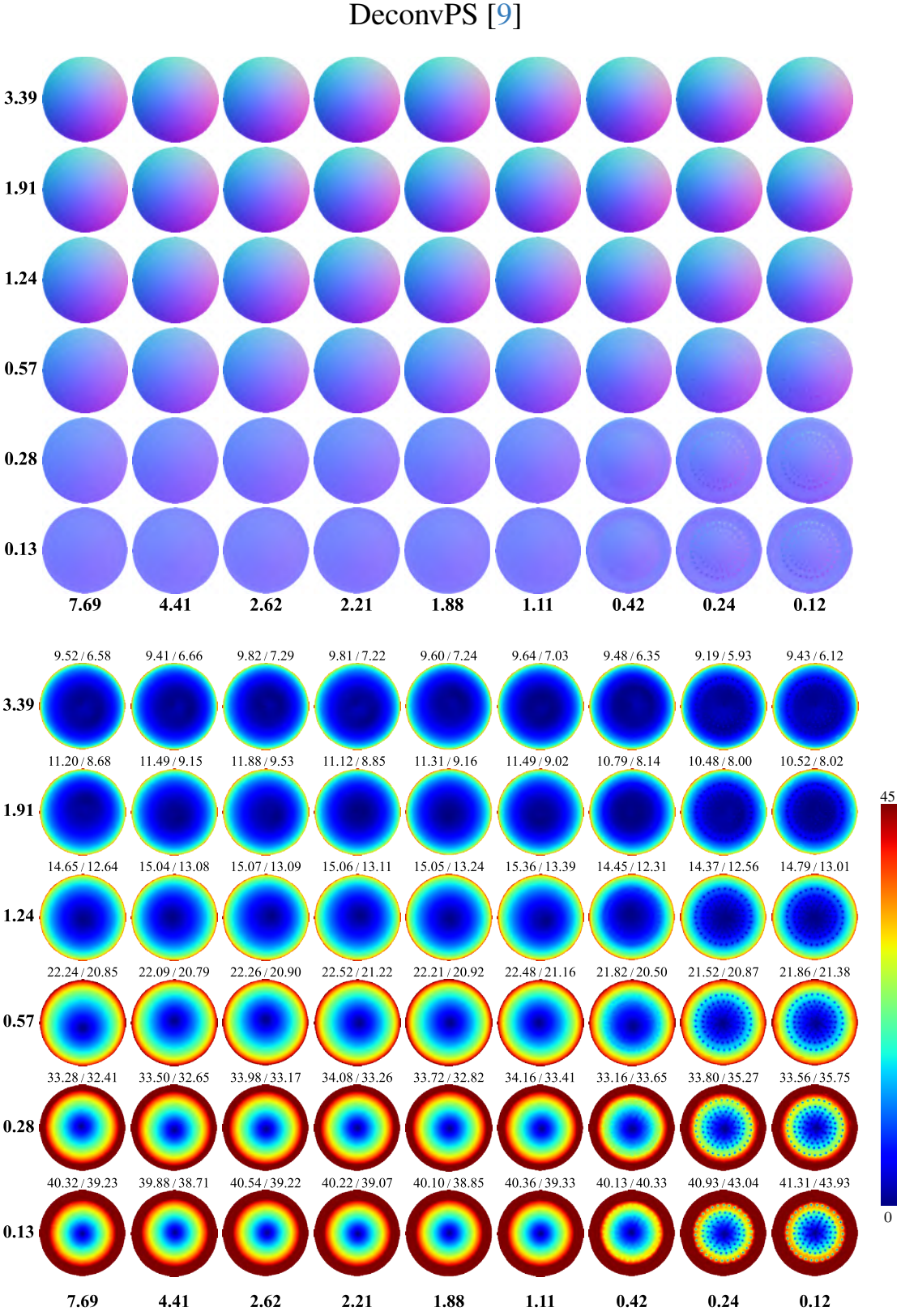


Figure S12. Estimated normal maps (top) and the corresponding angular error maps (bottom) of DeconvPS [9]. The mean and median errors for each material are displayed at the top of each error map.

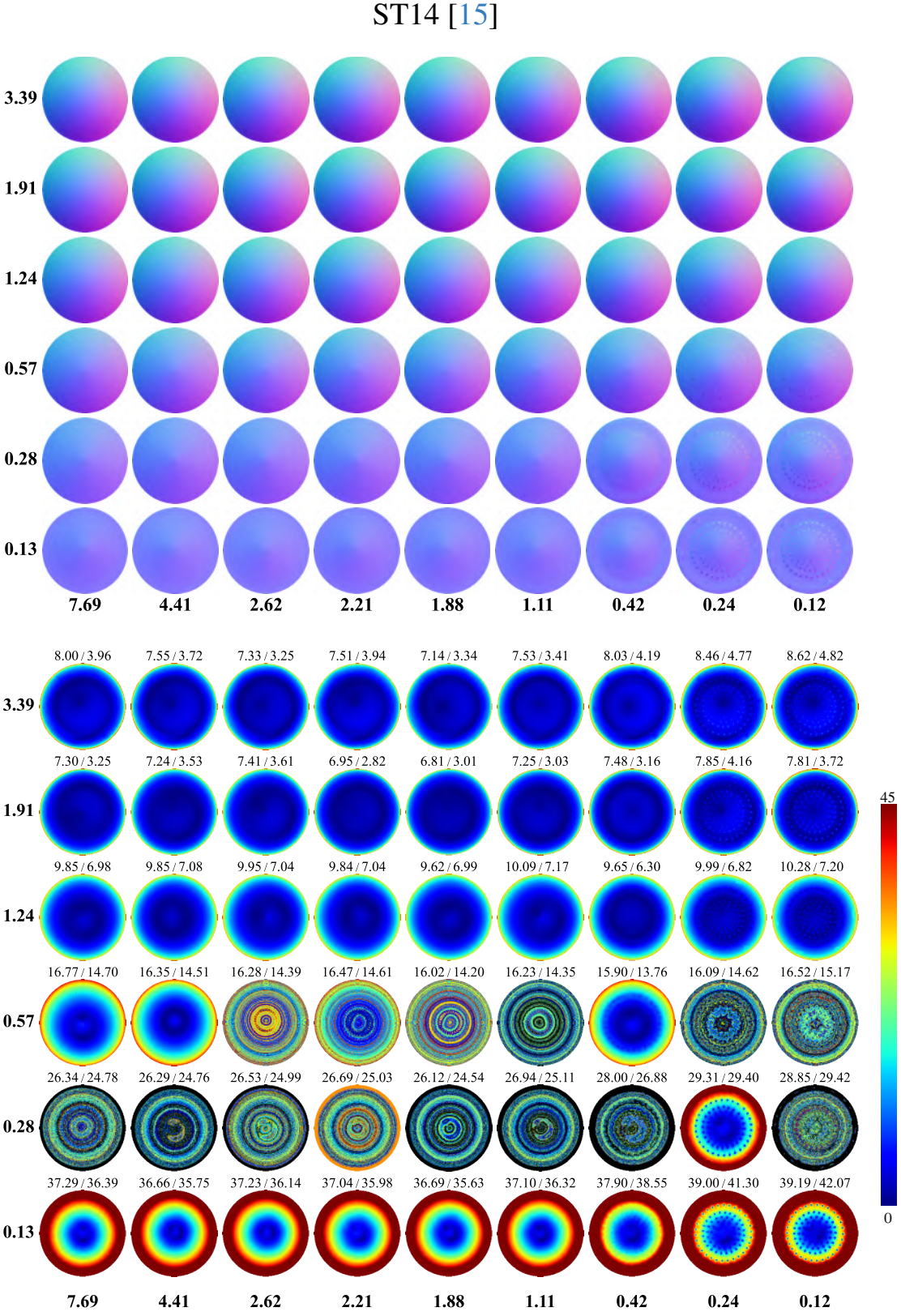


Figure S13. Estimated normal maps (top) and the corresponding angular error maps (bottom) of ST14 [15]. The mean and median errors for each material are displayed at the top of each error map.

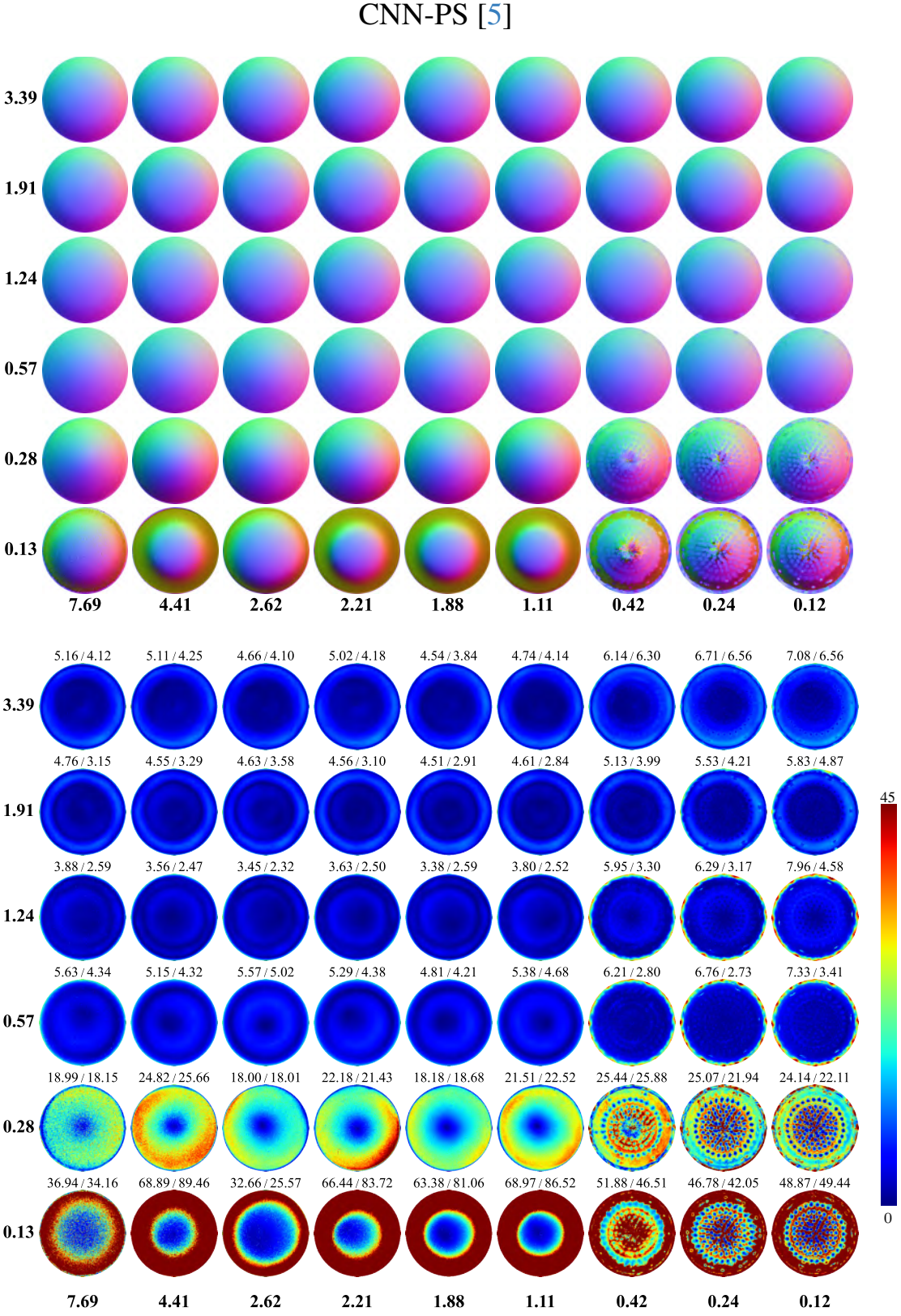


Figure S14. Estimated normal maps (top) and the corresponding angular error maps (bottom) of CNN-PS [5]. The mean and median errors for each material are displayed at the top of each error map.

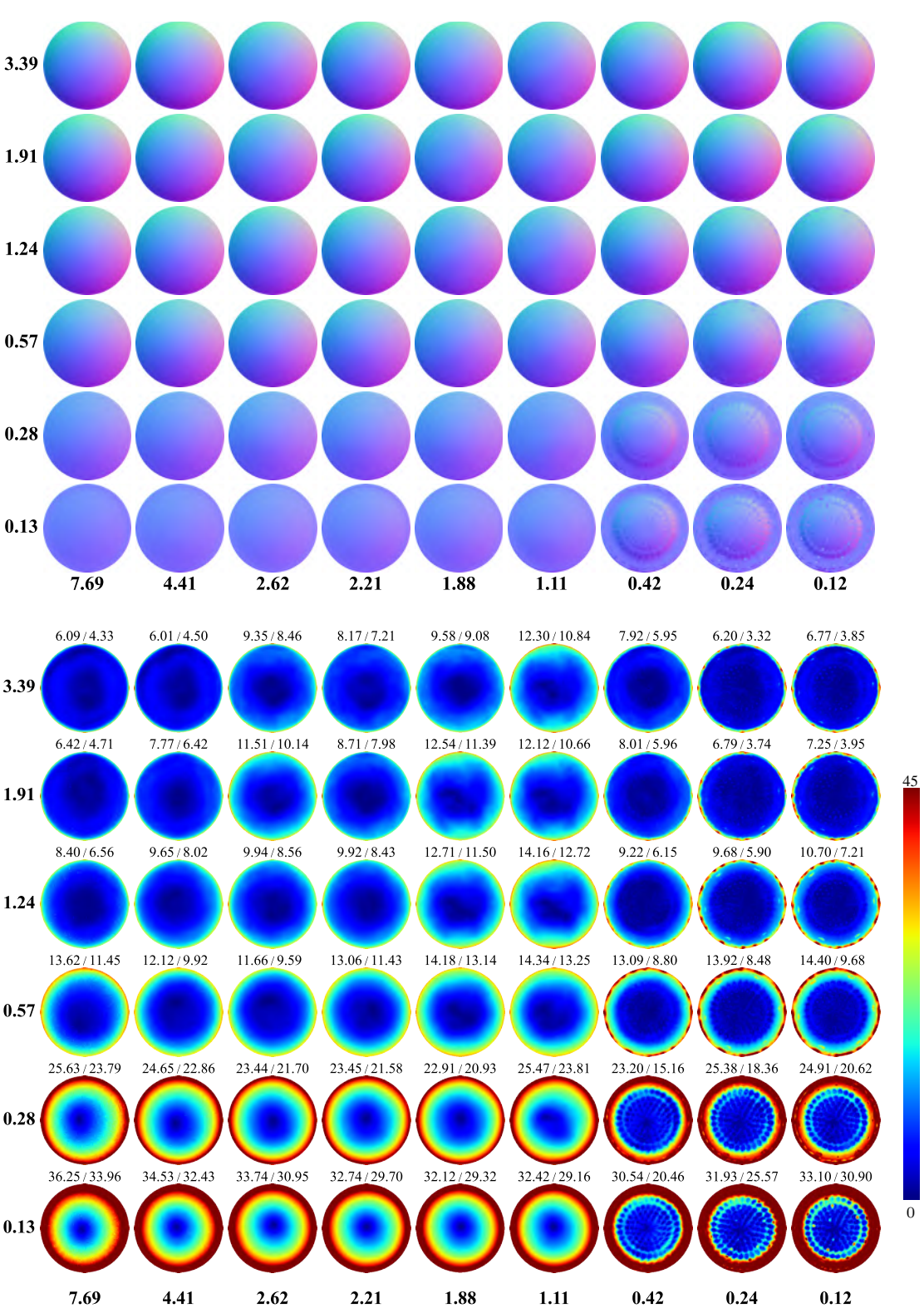


Figure S15. Estimated normal maps (top) and the corresponding angular error maps (bottom) of PS-FCN [2]. The mean and median errors for each material are displayed at the top of each error map.

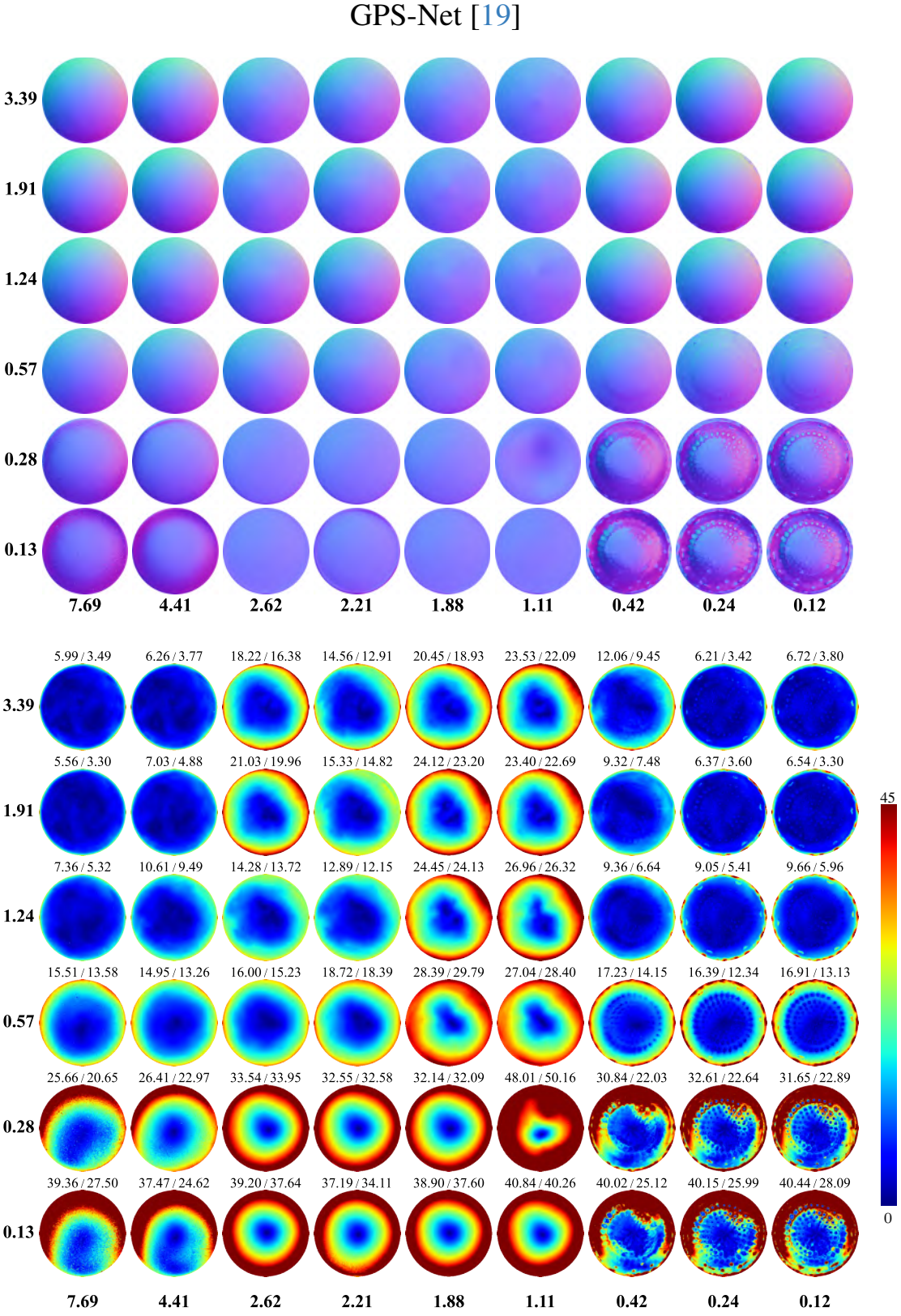


Figure S16. Estimated normal maps (top) and the corresponding angular error maps (bottom) of GPS-Net [19]. The mean and median errors for each material are displayed at the top of each error map.

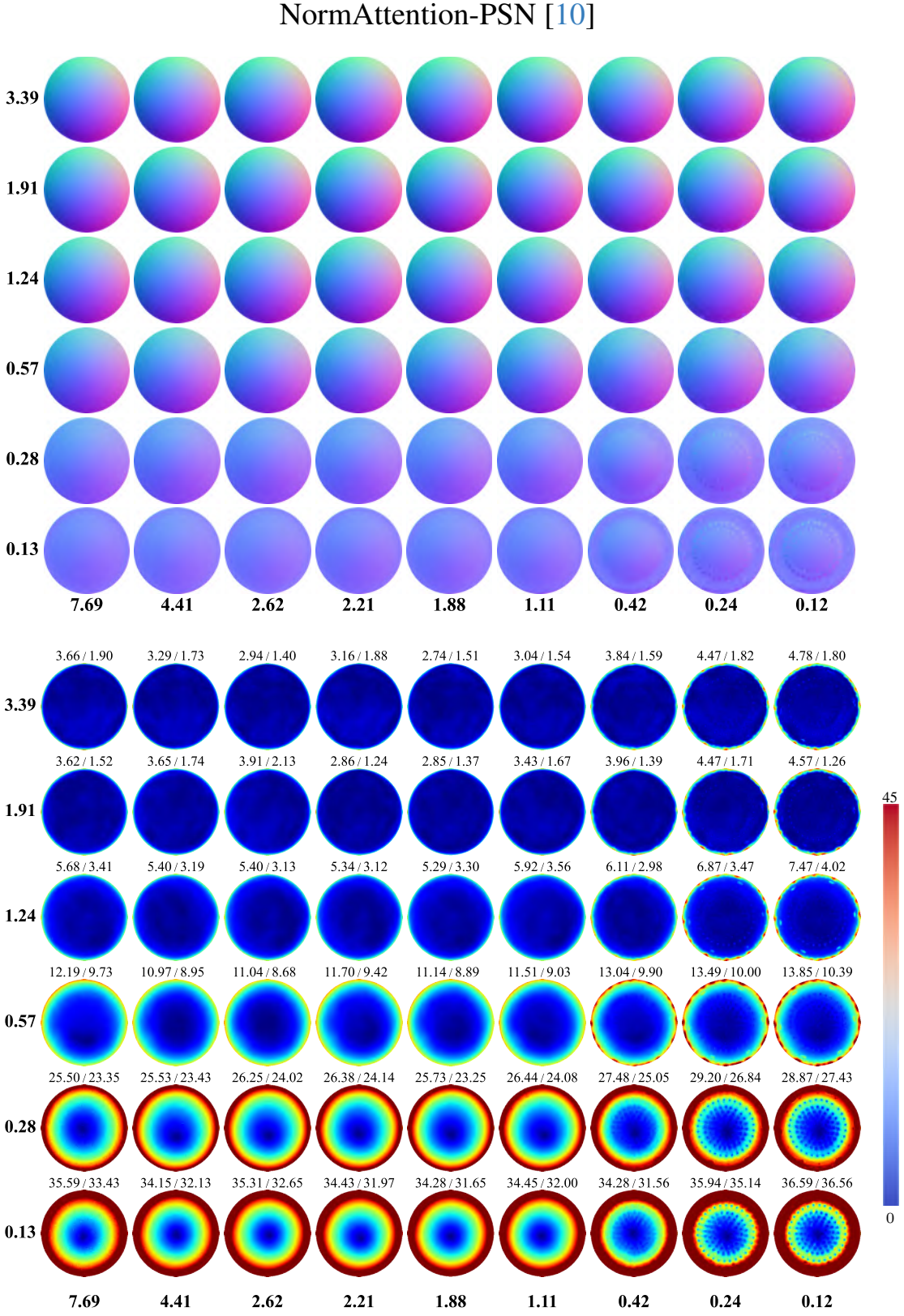


Figure S17. Estimated normal maps (top) and the corresponding angular error maps (bottom) of NormAttention-PSN [10]. The mean and median errors for each material are displayed at the top of each error map.

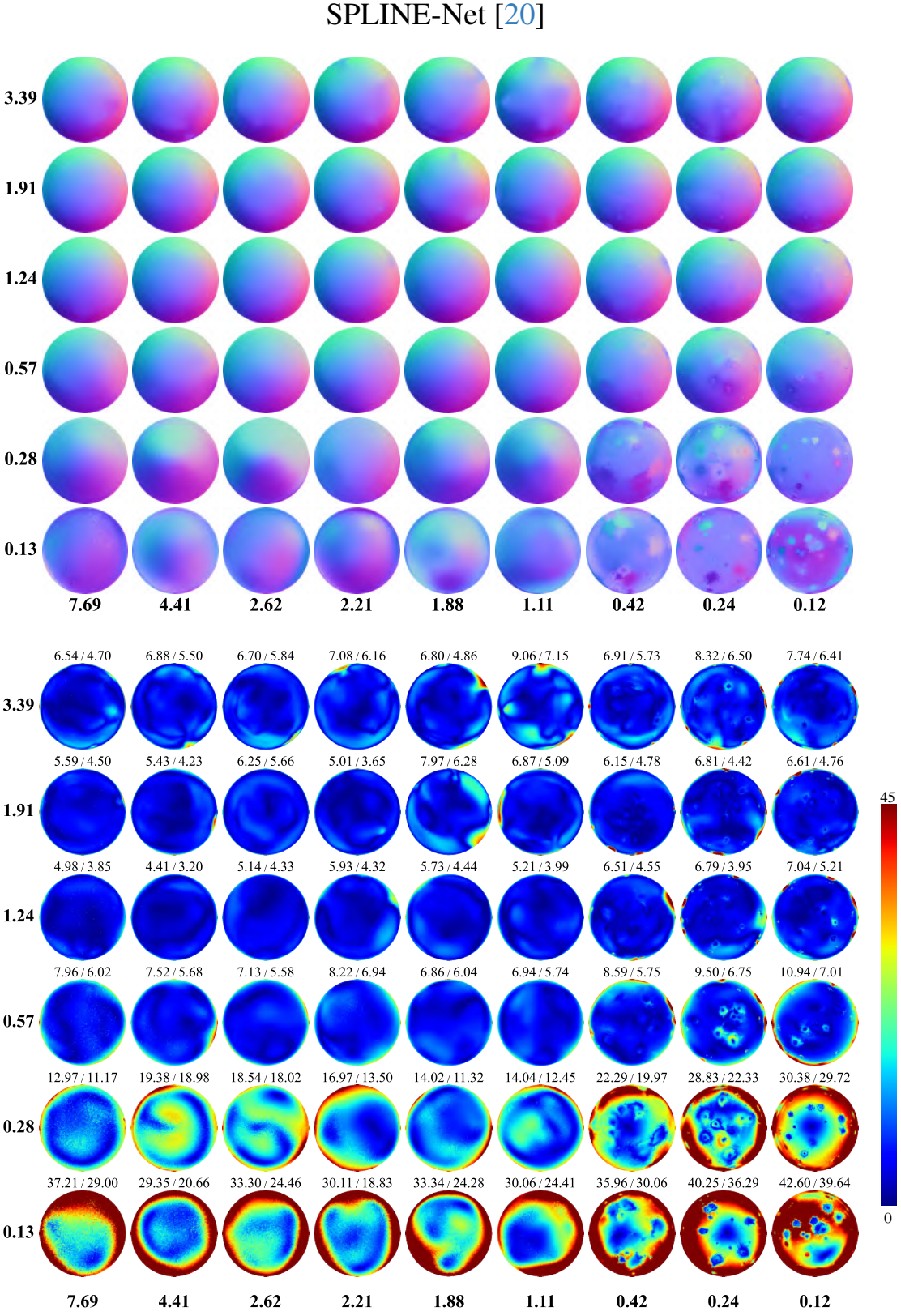


Figure S18. Estimated normal maps (top) and the corresponding angular error maps (bottom) of SPLINE-Net [20]. The mean and median errors for each material are displayed at the top of each error map.

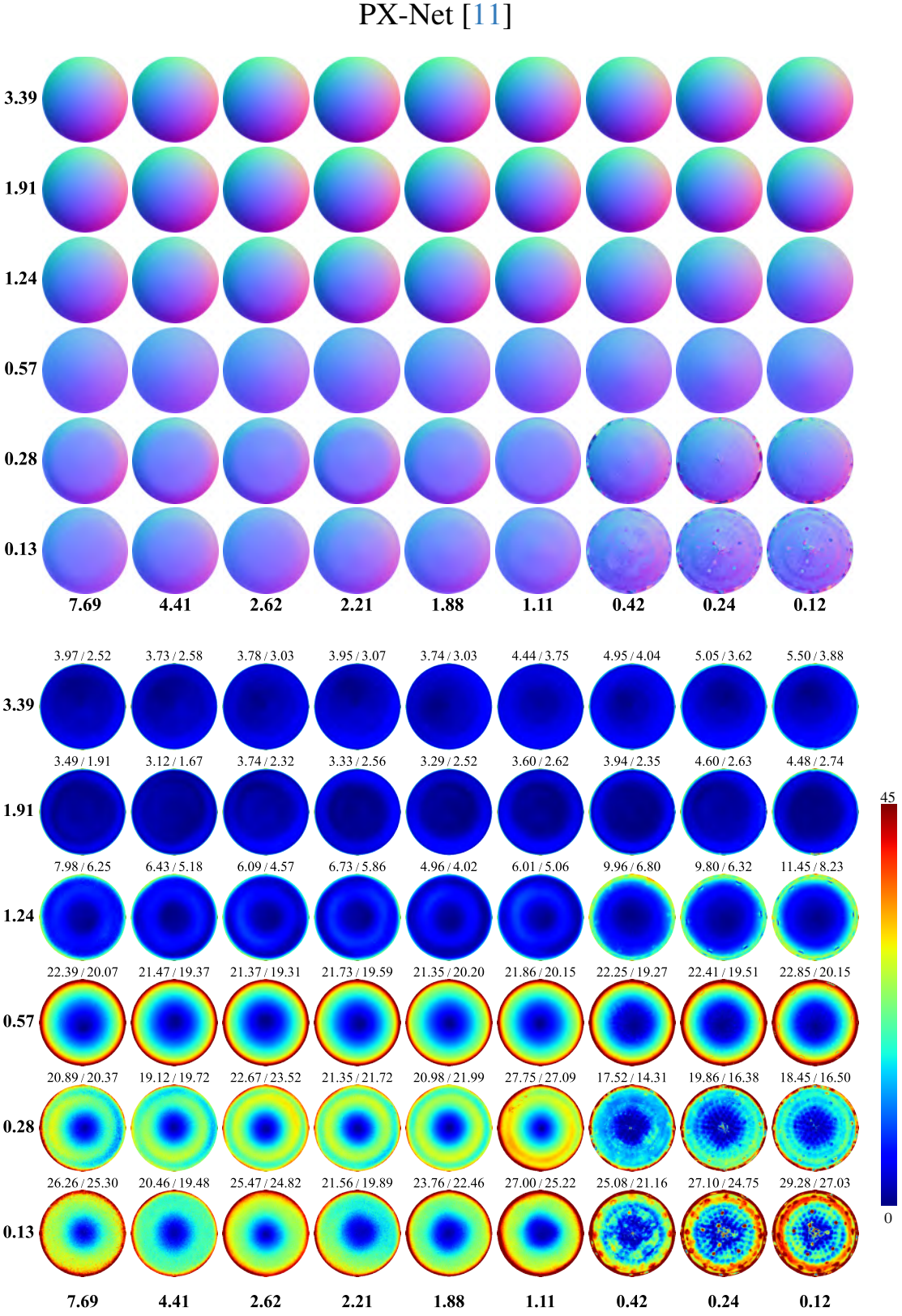


Figure S19. Estimated normal maps (top) and the corresponding angular error maps (bottom) of PX-Net [11]. The mean and median errors for each material are displayed at the top of each error map.

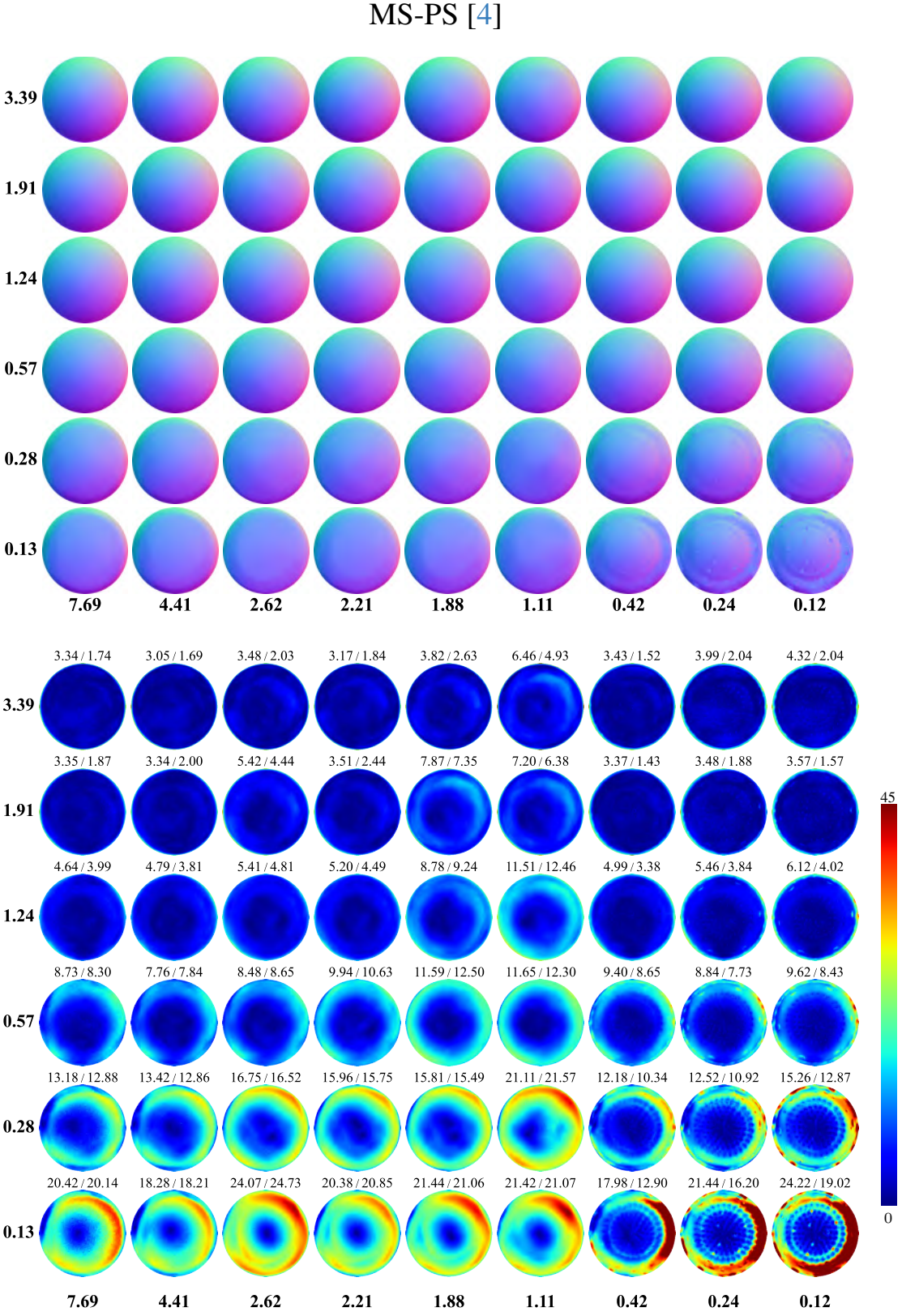


Figure S20. Estimated normal maps (top) and the corresponding angular error maps (bottom) of MS-PS [4]. The mean and median errors for each material are displayed at the top of each error map.

PF14 [12]

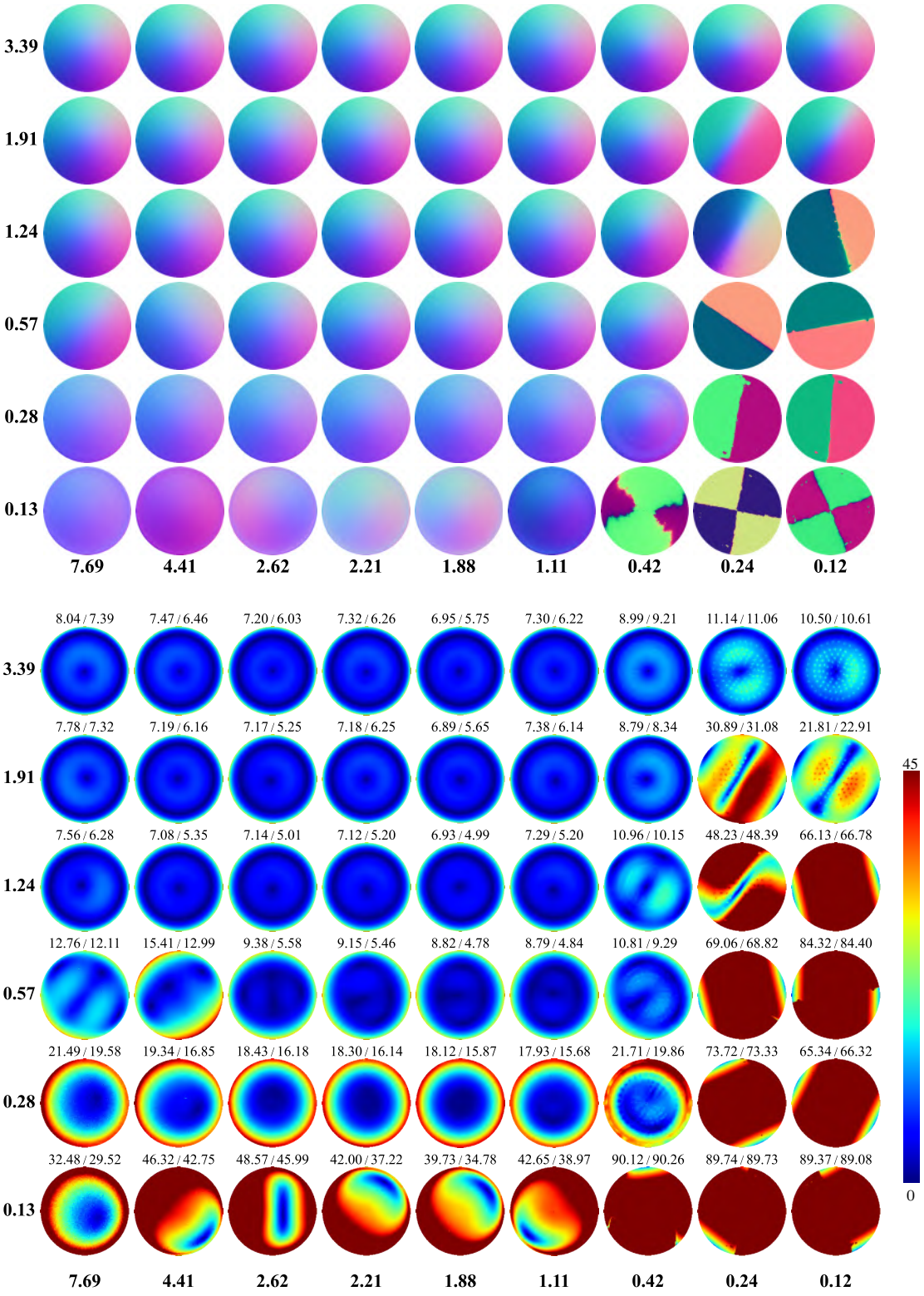


Figure S21. Estimated normal maps (top) and the corresponding angular error maps (bottom) of PF14 [12]. The mean and median errors for each material are displayed at the top of each error map.

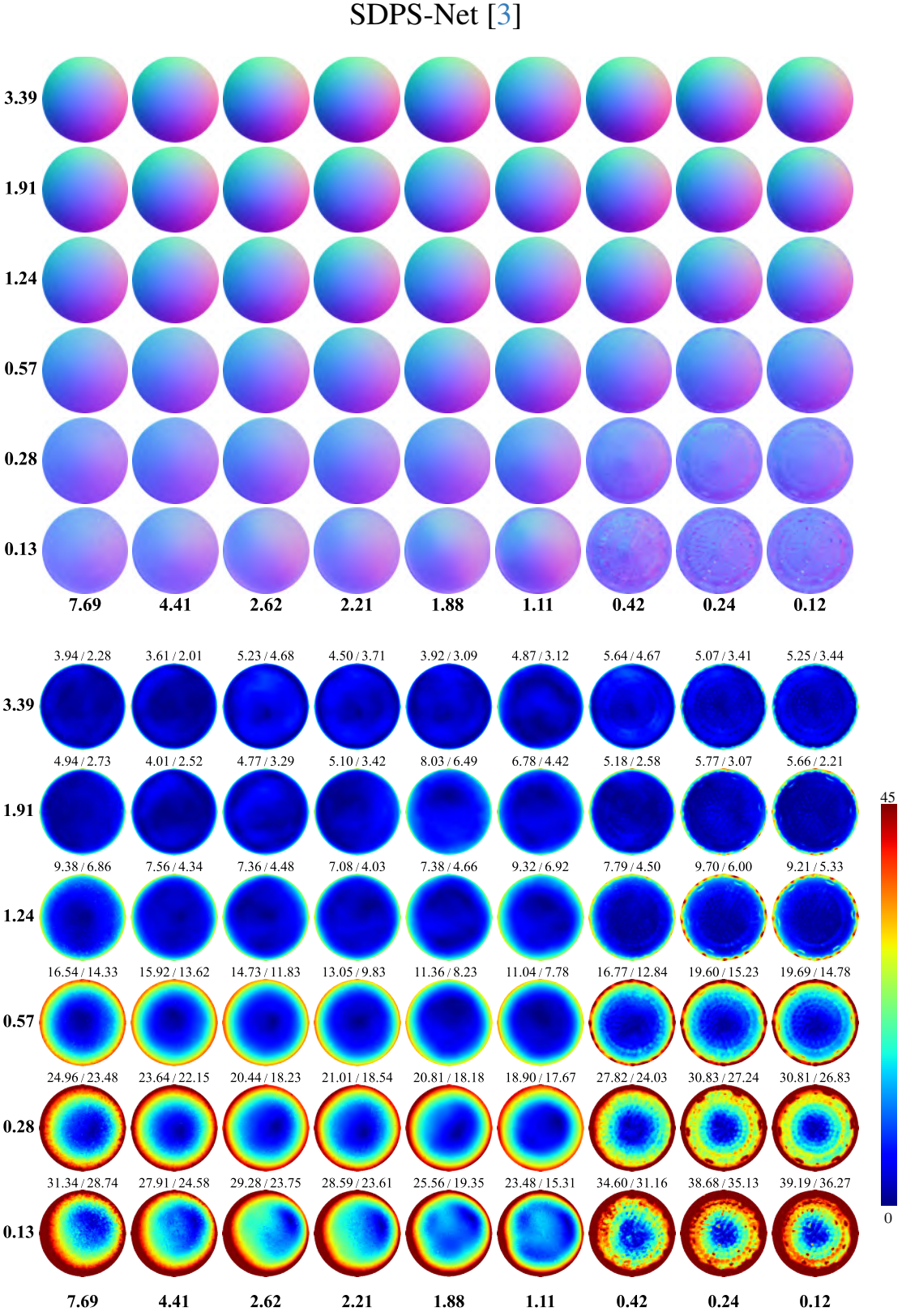


Figure S22. Estimated normal maps (top) and the corresponding angular error maps (bottom) of SDPS-Net [3]. The mean and median errors for each material are displayed at the top of each error map.

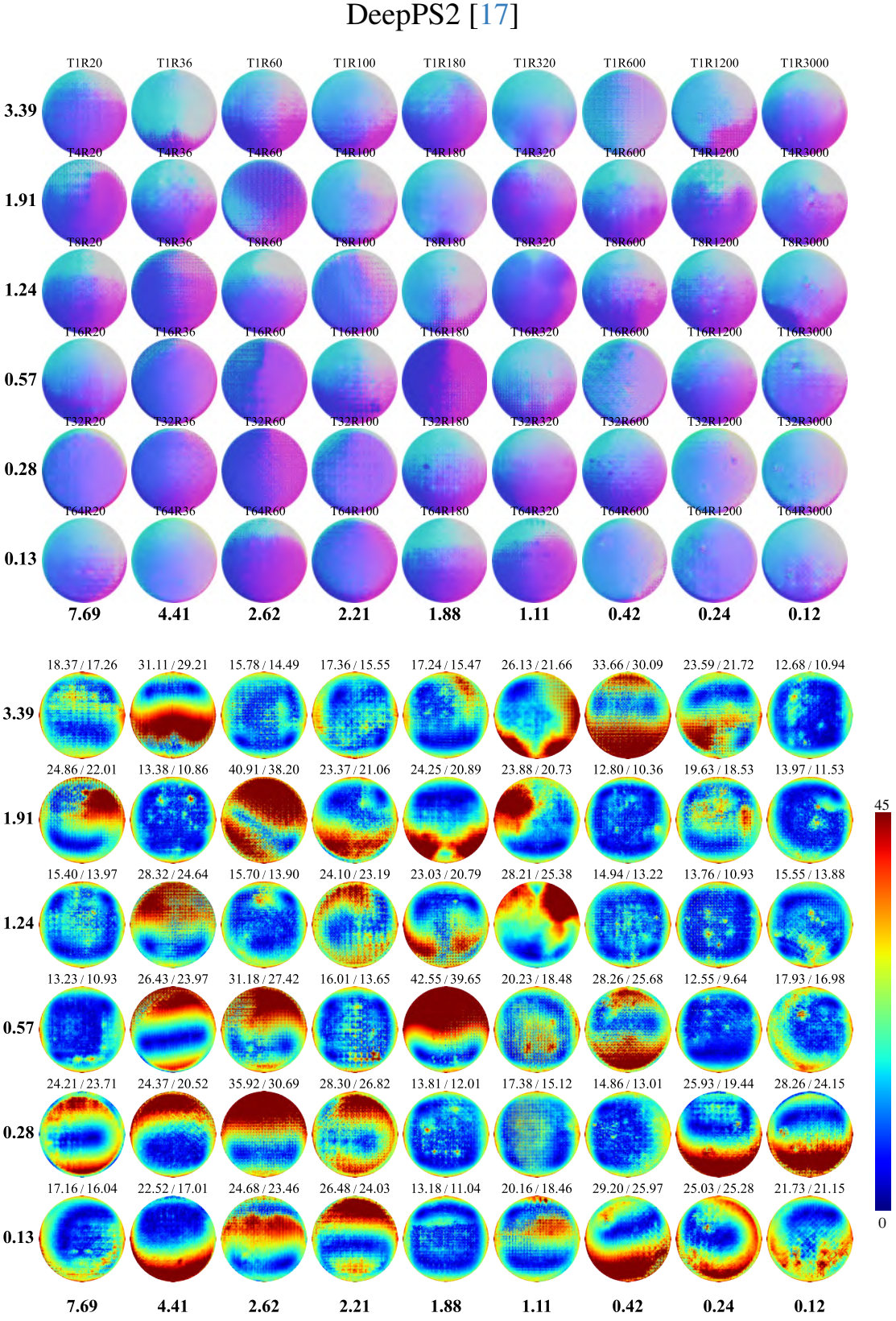


Figure S23. Estimated normal maps (top) and the corresponding angular error maps (bottom) of DeepPS2 [17]. The mean and median errors for each material are displayed at the top of each error map.

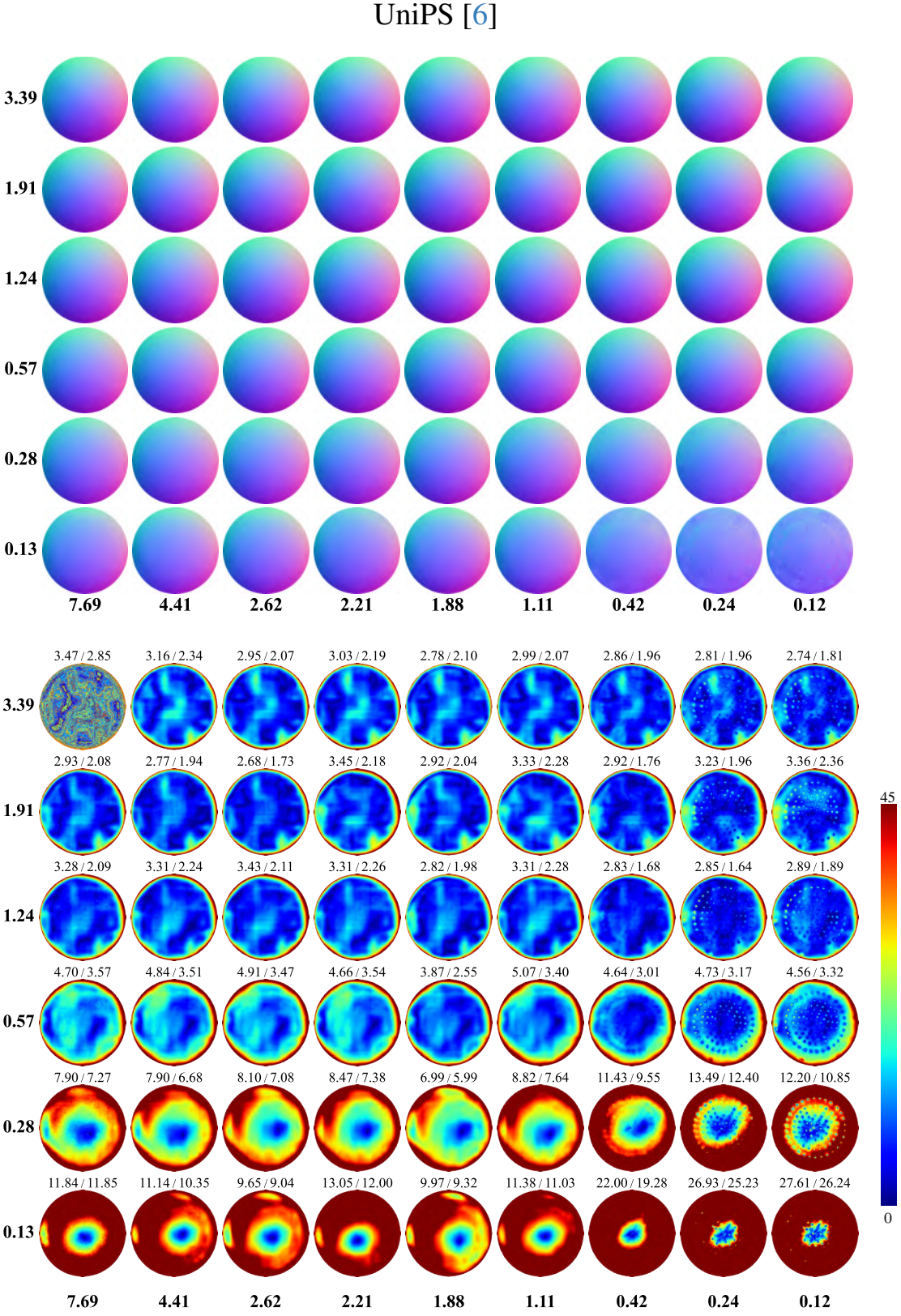


Figure S24. Estimated normal maps (top) and the corresponding angular error maps (bottom) of UniPS [6]. The mean and median errors for each material are displayed at the top of each error map.

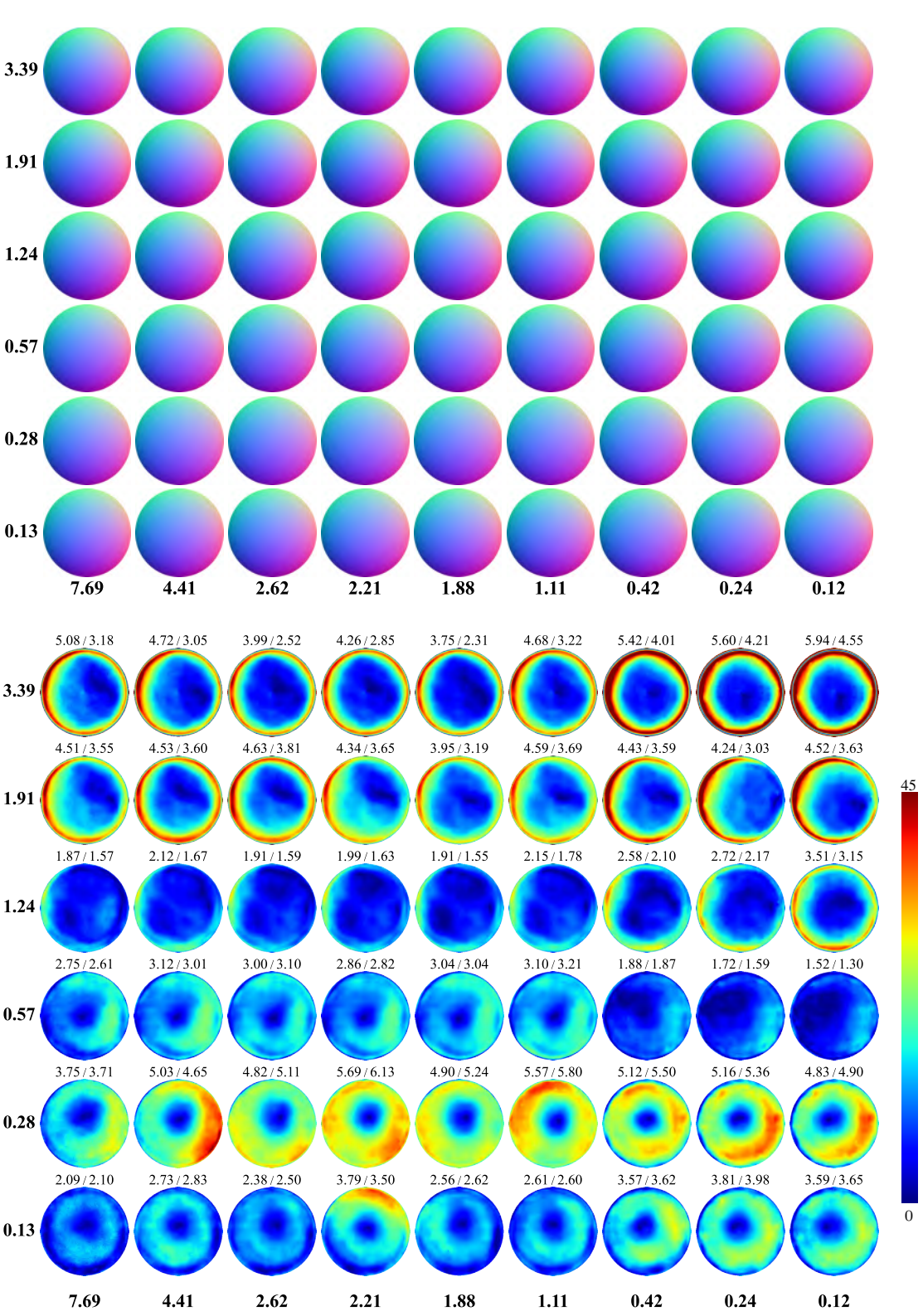


Figure S25. Estimated normal maps (top) and the corresponding angular error maps (bottom) of SDM-UniPS [7]. The mean and median errors for each material are displayed at the top of each error map.