DiLiGenT-Π: Photometric Stereo for Planar Surfaces with Rich Details – Benchmark Dataset and Beyond
Supplementary Material

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

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

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

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

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

[5] School of Artificial Intelligence, Beijing University of Posts and Telecommunications [6] Osaka University

{wangfeishi, shiboxin}@pku.edu.cn, {jiejiren, renmj}@sjtu.edu.cn, guoheng.bupt@gmail.com

This supplementary material provides additional technical details about the normal distribution of the objects in our dataset, the process of calibrating lighting direction, and complete benchmark results on all objects.

A. Distribution of Surface Normal Directions

In contrast to the objects in DiLiGenT [11] that have diverse surface normal directions, the near-planar objects in DiLiGenT-Π exhibit a strongly unbalanced distribution of surface normals, with most of the normals being concentrated around the Z direction. We present histograms of the surface normal directions for both datasets in Fig. S1 and Fig. S2 respectively, which clearly demonstrate the pronounced concentration of surface normals around the z direction in DiLiGenT-Π.

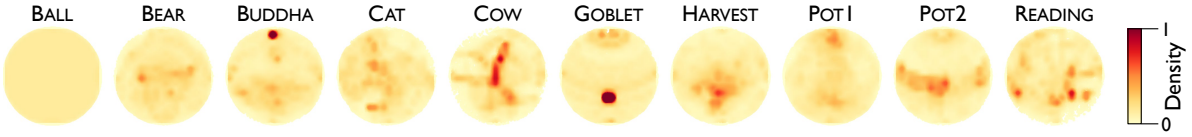


Figure S1: Surface normal distributions of objects in DiLiGenT [11].

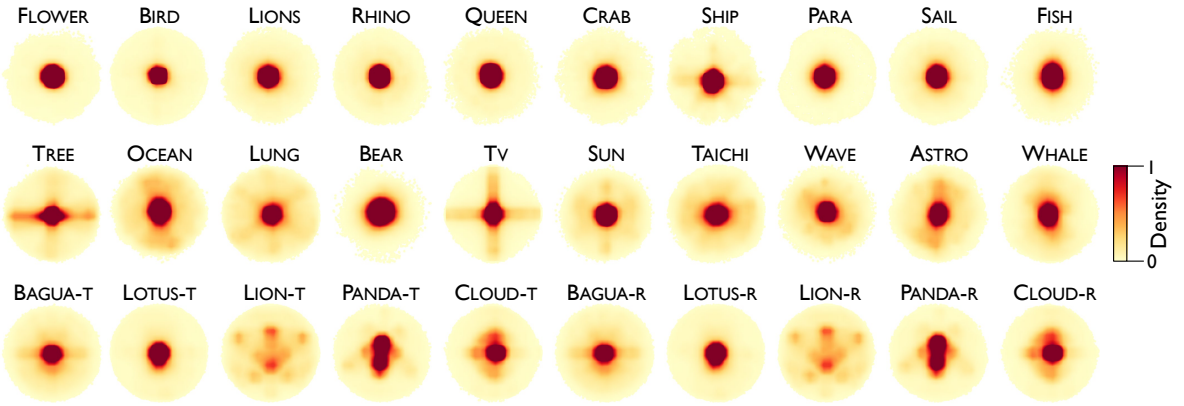


Figure S2: Surface normal distributions of objects in DiLiGenT-Π.

† Equally contributed authors ‡ Corresponding authors.

B. Lighting Direction Calibration

As discussed in the main paper, we employ a robot arm to achieve diverse light directions in our capture setup. In DiLiGenT-Π, we design a uniform light distribution by placing 100 light positions at uniform latitude and longitude intervals.

Although light direction distribution is designed beforehand, the rotating device only ensures limited accuracy, which may lead to minor deviations between the intended light directions and the actual ones in a real scene. To obtain the light directions precisely, we follow DiLiGenT10^2 [10] and employ an "on the fly" calibration strategy that calculates the light direction of each image using six surrounding mirror spheres.

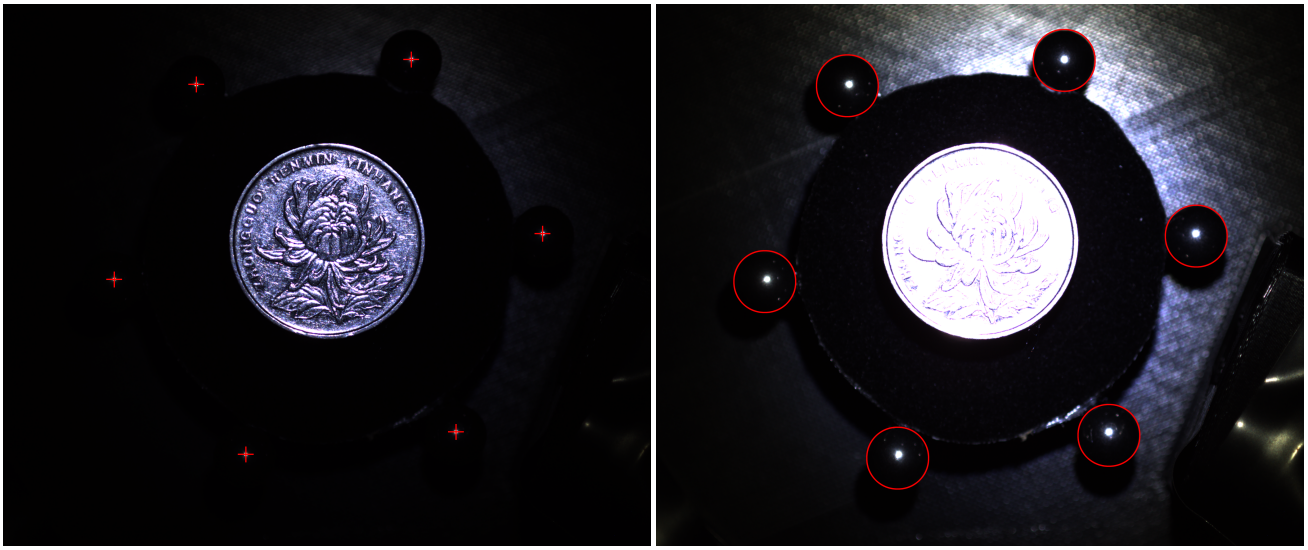


Figure S3: Images of FLOWER captured in low (left) and high (right) exposure time for detecting the specular spots and contour of the 6 mirror spheres for light calibration.

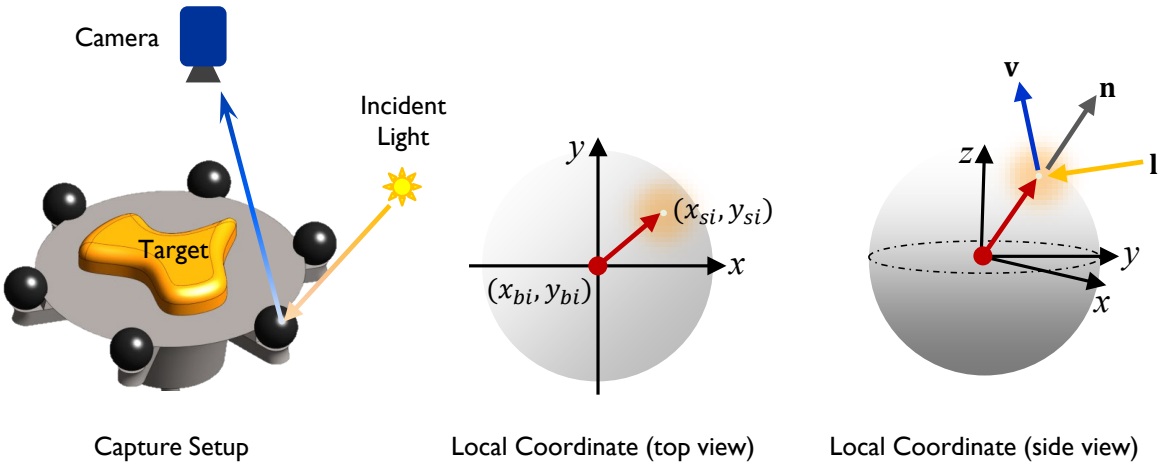


Figure S4: The geometrical relationship between the captured object and a mirror sphere in our light calibration process, showing the setup (left), the sphere in camera coordinate (middle), and the geometrical relationship on the sphere in 3D coordinate (right).

As shown in Fig. S3, in the low-exposure image, the specular spots of each ball are small enough to be viewed as a point, and the vector from the ball center to this point indicates the light direction. In the high-exposure image, the contour of mirror

spheres is more prominent, revealing their radius and center location. For the i-th of the six spheres, we denote the sphere radius as r_i, sphere center and specular spot location in the image plane as (x_{bi}, y_{bi}) and (x_{si}, y_{si}). Additionally, we define half vector $\mathbf{h}_i$ as the direction that lies halfway between the light source $\mathbf{l}$ and the viewer direction $\mathbf{v}$. Based on the middle part of Fig. S4, we have:

$$\mathbf{h}_i = \left[\frac{x_{si} - x_{bi}}{r_i}, \frac{y_{si} - y_{bi}}{r_i}, \frac{z_{bi}}{r_i}\right]^\top, \tag{1}$$

where $z_{bi} = \sqrt{r_i^2 - (x_{si} - x_{bi})^2 - (y_{si} - y_{bi})^2}$. Thus we have:

$$\begin{aligned} h_{ix} r_i &= x_{si} - x_{bi}, \\ h_{iy} r_i &= y_{si} - y_{bi}. \end{aligned} \tag{2}$$

By assuming distant illumination on all of the six mirror spheres, the half vector $\mathbf{h_i}$ for each sphere should be equal. Consequently, we denote an optimized half vector as $\mathbf{h} = [h_x, h_y, h_z]^\top$. Its coordinates under noise can be written as:

$$\begin{aligned} h_x r_i &= x_{si} - x_{bi} + \epsilon_{xi}, \\ h_y r_i &= y_{si} - y_{bi} + \epsilon_{yi}, \end{aligned} \tag{3}$$

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 h_x and h_y:

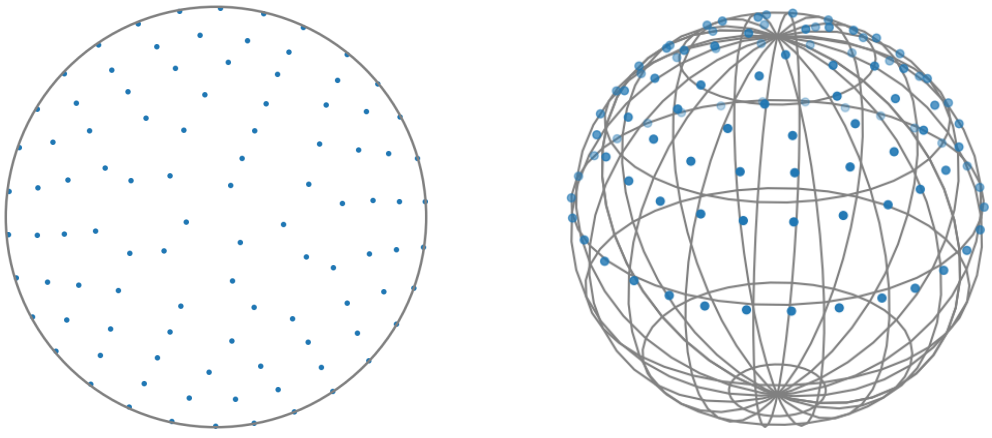
$$\begin{aligned} h_x &= \frac{\sum_{i=1}^{6}(x_{si} - x_{bi}) r_i}{\sum_{i=1}^{6} r_i^2}, \\ h_y &= \frac{\sum_{i=1}^{6}(y_{si} - y_{bi}) r_i}{\sum_{i=1}^{6} r_i^2}, \end{aligned} \tag{4}$$

and calculate h_z simply by $h_z = \sqrt{1 - h_x^2 - h_y^2}$. Given the optimized half vector $\mathbf{h}$, we acquire light direction according to the law of reflection illustrated in the right side of Fig. S4:

$$\mathbf{l} = 2(\mathbf{h}^\top \mathbf{v})\mathbf{h} - \mathbf{v}, \tag{5}$$

where $\mathbf{v} = [0, 0, 1]^\top$ represents the view direction.

The light distribution in different scenes could have slight differences due to the mechanical moving residue of the robot arm, therefore we provide calibrated light directions for each scene. Based on the above calibration process, we present the light distribution of one example scene in DiLiGenT-Π in Fig. S5, where the lights are uniformly distributed on the upper hemisphere.



(a) 2D Top View (b) 3D View

Figure S5: Light distribution of the FLOWER scene in our DiLiGenT-Π dataset.

C. Evaluation of Light Calibration

Our dataset adopts the assumption of uniform light intensity and a viewing direction of $[0, 0, 1]$. For the light intensity, our capture setup employs a single LED mounted on a robot arm instead of a multi-LED system, and the LED moves while maintaining a uniform distance to the object, ensuring that the light intensities are nearly uniform in practice. Regarding the viewing direction, the error it introduces to the light direction is negligible. To validate this point, we conduct an experiment. We put an Aruco marker at the target object position in our setup, and calibrate the view direction using the 3D marker positions and the camera origin. The angular difference between the calibrated view direction and the vector $[0, 0, 1]$ is $0.32°$. Based on this calibrated view direction, we further re-calibrate the light directions in DiLiGenT-Π dataset. The mean angular error (MAE) with the original one is $0.36°$, which can be disregarded.

D. Surface Measurement Process

As shown in Fig. S6 (a), to obtain the 'GT' normal map, we take the Alicona Infinity Focuse[1] to measure the surface details of each object and output precise point cloud (accuracy up to 10 nm). In the measurement process, we choose a $5\times$ objective lens (Fig. S6 (b)) to acquire the measurement in a relatively large local region. By traversing all the sub-regions step by step and registering the measurement at each region with the IF-MeasureSuit software of Alicona, we obtain the point cloud and the corresponding mesh file of the complete target surface. Compared with the scanned mesh from DiLiGenT which measures around 10^4 points per cm^2, the scanned point cloud of DiLiGenT-Π contains around 4M points per cm^2, which is essential to guarantee that our 'GT' normal maps contain tiny detailed structures of the near-planar surfaces.

E. Evaluation under Sparse Inputs

We uniformly select subsets of lights with sizes 5, 10, and 20, respectively, and evaluate PS baselines on DiLiGenT-Π dataset. The average MAEs are shown in Table S1. For 20 lights, there is no significant performance drop. As the light number reduces to 10 or 5, the MAE increases significantly. As expected, TH46 and CNN-PS are more sensitive to the number of lights due to the usage of thresholding strategy and the observation map.

Table S1: Benchmark results on DiLiGenT-Π with 5, 10, and 20 lights. The difference compared with 100 lights is shown in the even rows. We denote 'NA-PSN' as the abbreviation of NormAttention-PSN [6].

# Light	Plane	LSPS [13]	TH28 [11]	TH46 [11]	WG10 [14]	ST14 [12]	CNN-PS [4]	PS-FCN [2]	GPS-Net [15]	NA-PSN [6]	PF14 [9]	UPS-FCN [2]	SDPS-Net [1]
5	14.5	15.9	13.9	37.9	83.0	16.5	65.5	17.7	14.3	20.0	45.4	25.0	30.7
Diff.	0	5.32	2.92	27.4	70.3	4.93	56.4	7.88	4.18	10.8	5.49	3.49	4.74
10	14.5	13.16	12.9	17.5	17.1	13.8	18.3	13.6	12.2	13.2	41.7	24.3	29.2
Diff.	0	2.62	1.90	6.95	4.38	2.29	9.19	3.78	2.09	3.95	1.72	2.76	3.26
20	14.5	11.3	11.6	12.1	13.2	12.3	11.5	11.4	11.2	10.2	38.7	22.9	28.9
Diff.	0	0.72	0.62	1.55	0.5	0.78	2.32	1.52	1.04	0.97	-1.3	1.35	2.97
100	14.5	10.5	11.0	10.6	12.7	11.5	9.15	9.84	10.1	9.2	39.9	21.5	25.9

[1] Website: https://www.alicona.com/products/. Retrieved August 16, 2023.

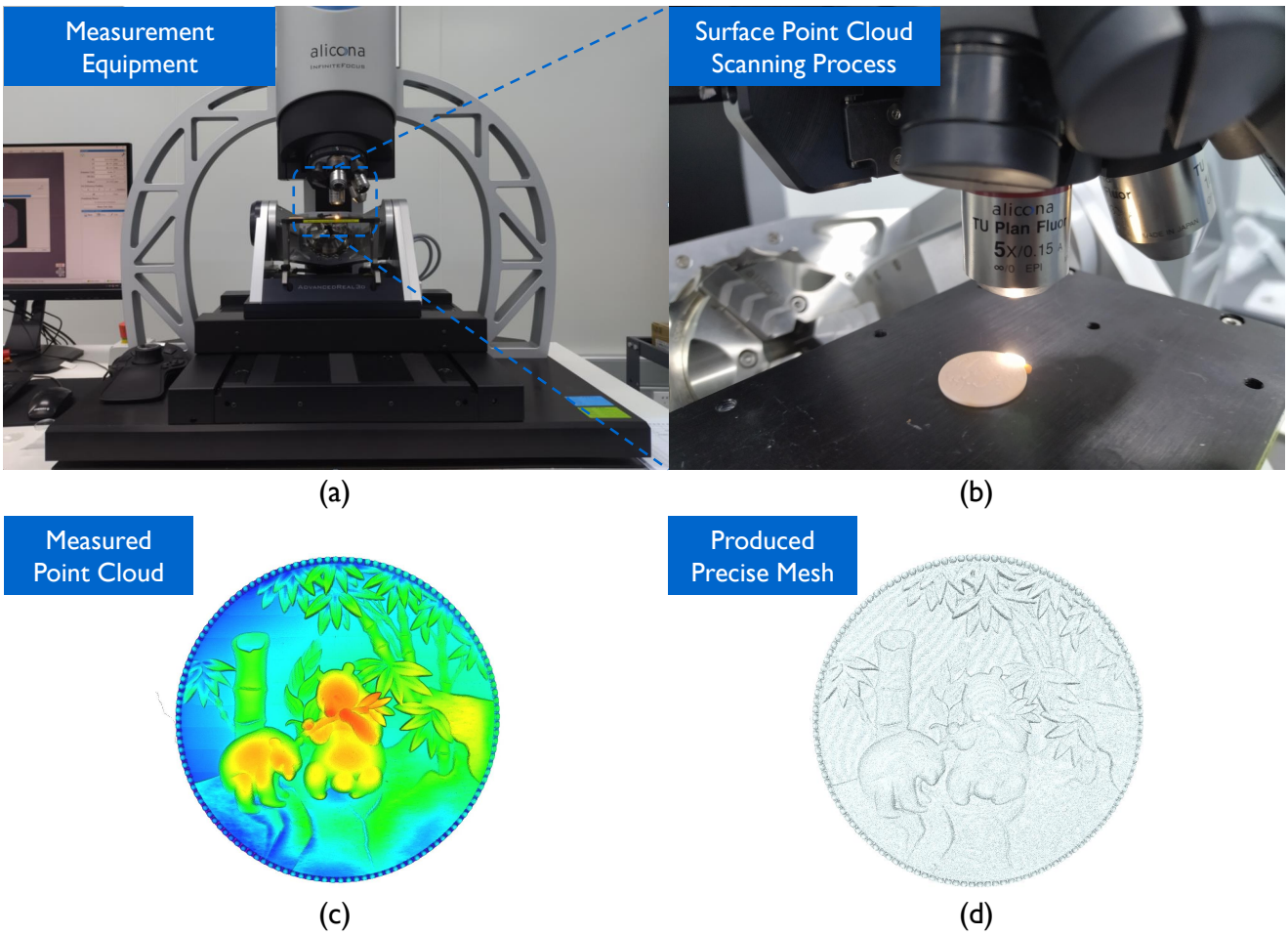


Figure S6: Surface measurement process of the proposed dataset. A precise mesh is obtained via Alicona Infinity Focuse to generate the 'GT' surface normal map.

F. Complete Benchmark Results on DiLiGenT-Π

From Figs S7 to S36, we provide complete benchmark results including estimated normal maps and their corresponding angular error maps. Each page focuses on a particular object and displays the performance comparison of 15 photometric stereo methods. Here, we include a range of calibrated photometric stereo methods (LSPS [13], TH28, TH46 [11], WG10 [14], ST14 [12], PX-Net [8], CNN-PS [4], PS-FCN [2], GPS-Net [15], and NA-PSN [6]), various uncalibrated photometric stereo methods (SDPS-Net [1], PF14 [9], UPS-FCN [2], and UPS-GCNet [3]), as well as methods dealing with natural lights (LW21 [7] and SDM-UniPS [5]). The plane shape defined by the $[0, 0, 1]$ vector is also included for reference. The mean and median of angular errors are at the top of each error map, displayed as "mean / median", with the minimum values highlighted in red.

FLOWER

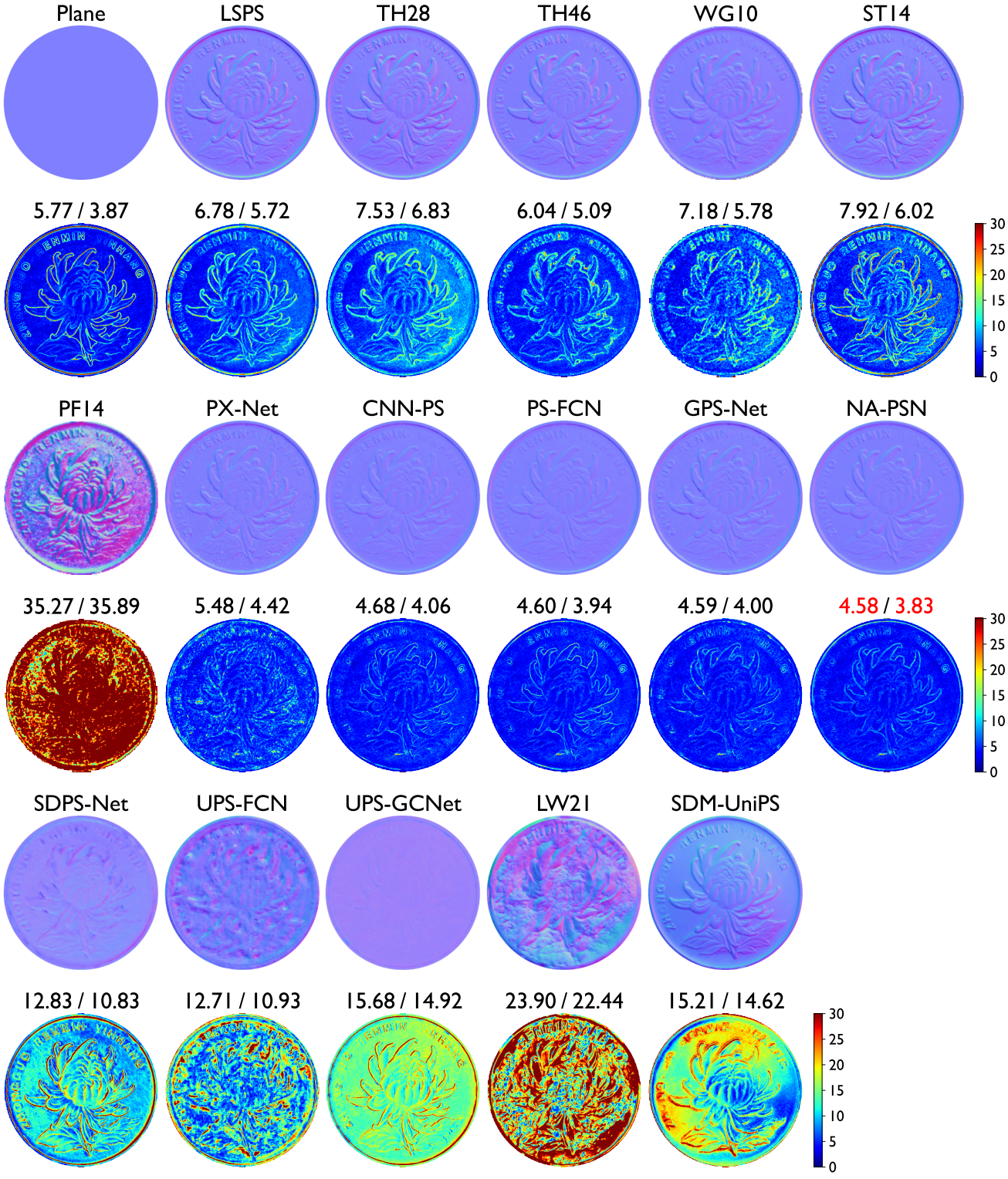


Figure S7: Estimated surface normal maps and angular error maps of FLOWER.

BIRD

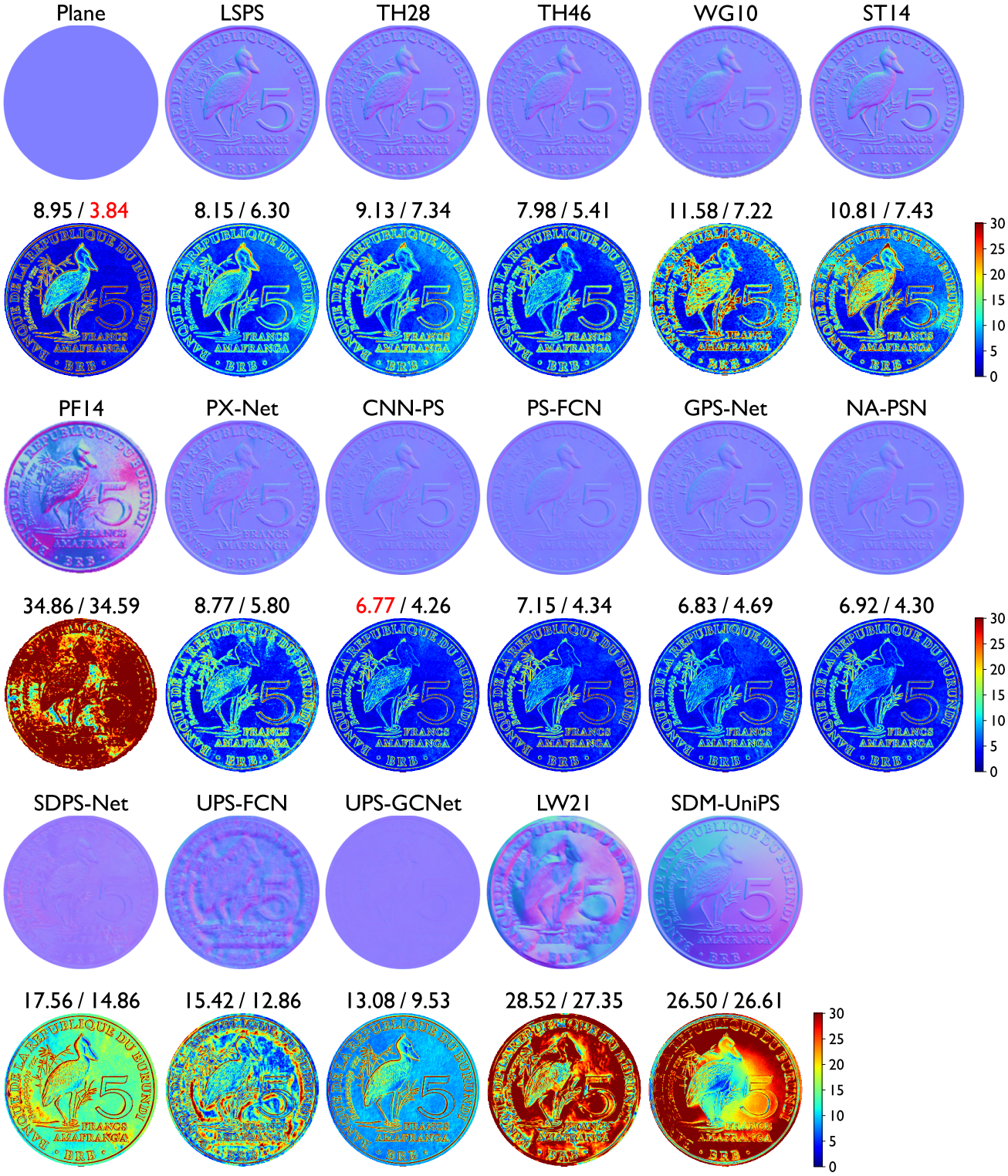


Figure S8: Estimated surface normal maps and angular error maps of BIRD.

LIONS

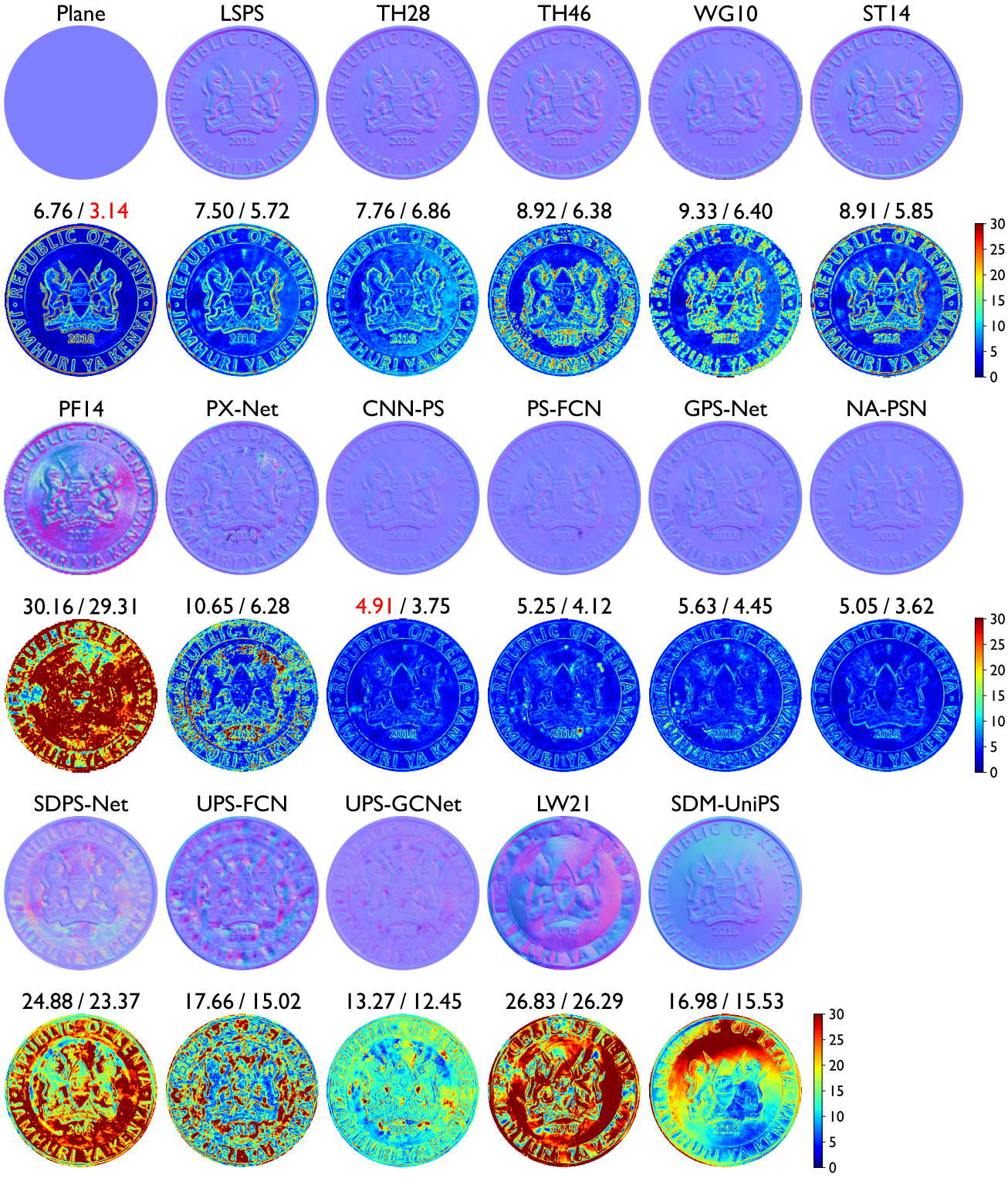


Figure S9: Estimated surface normal maps and angular error maps of LIONS.

RHINO

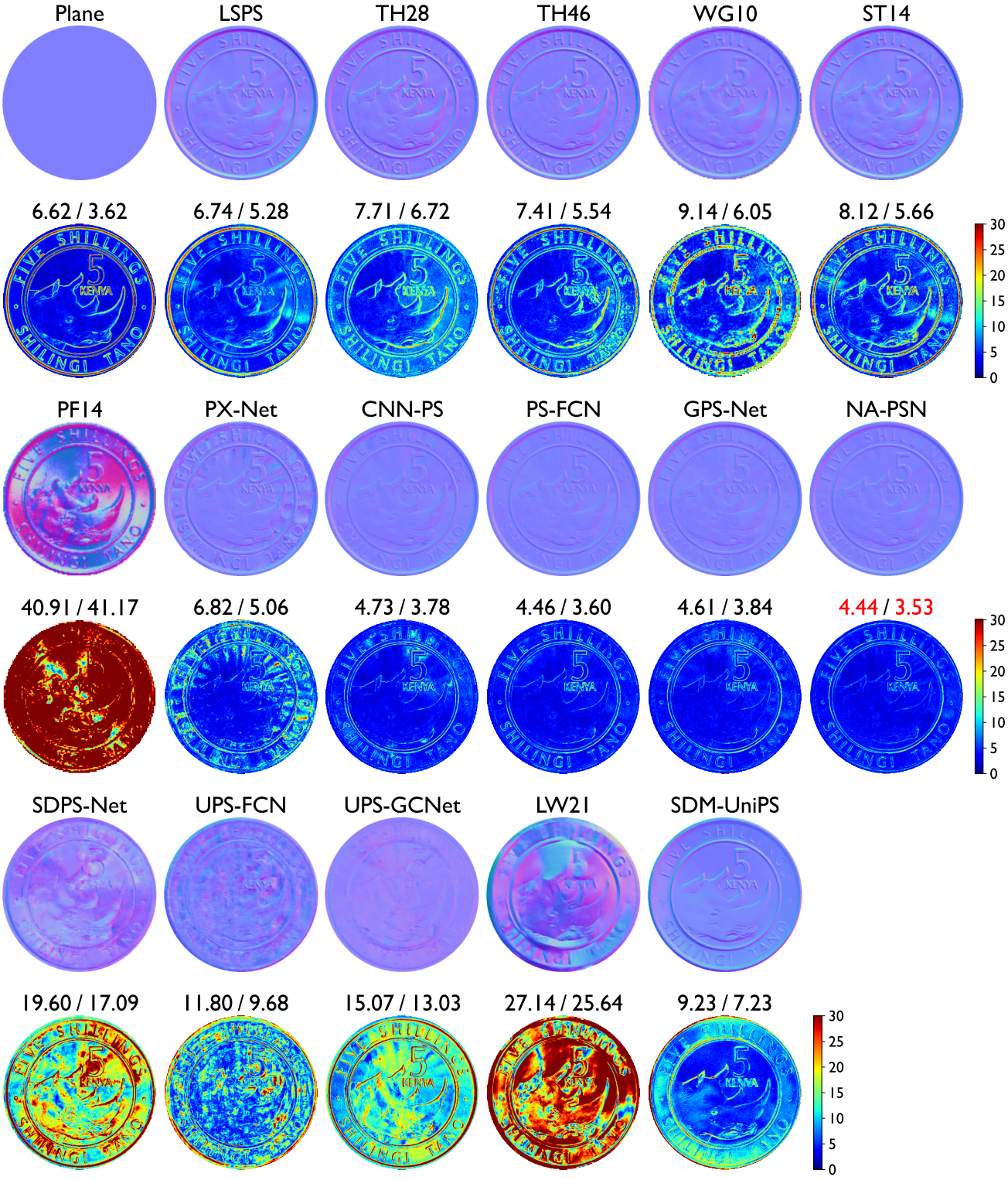


Figure S10: Estimated surface normal maps and angular error maps of RHINO.

QUEEN

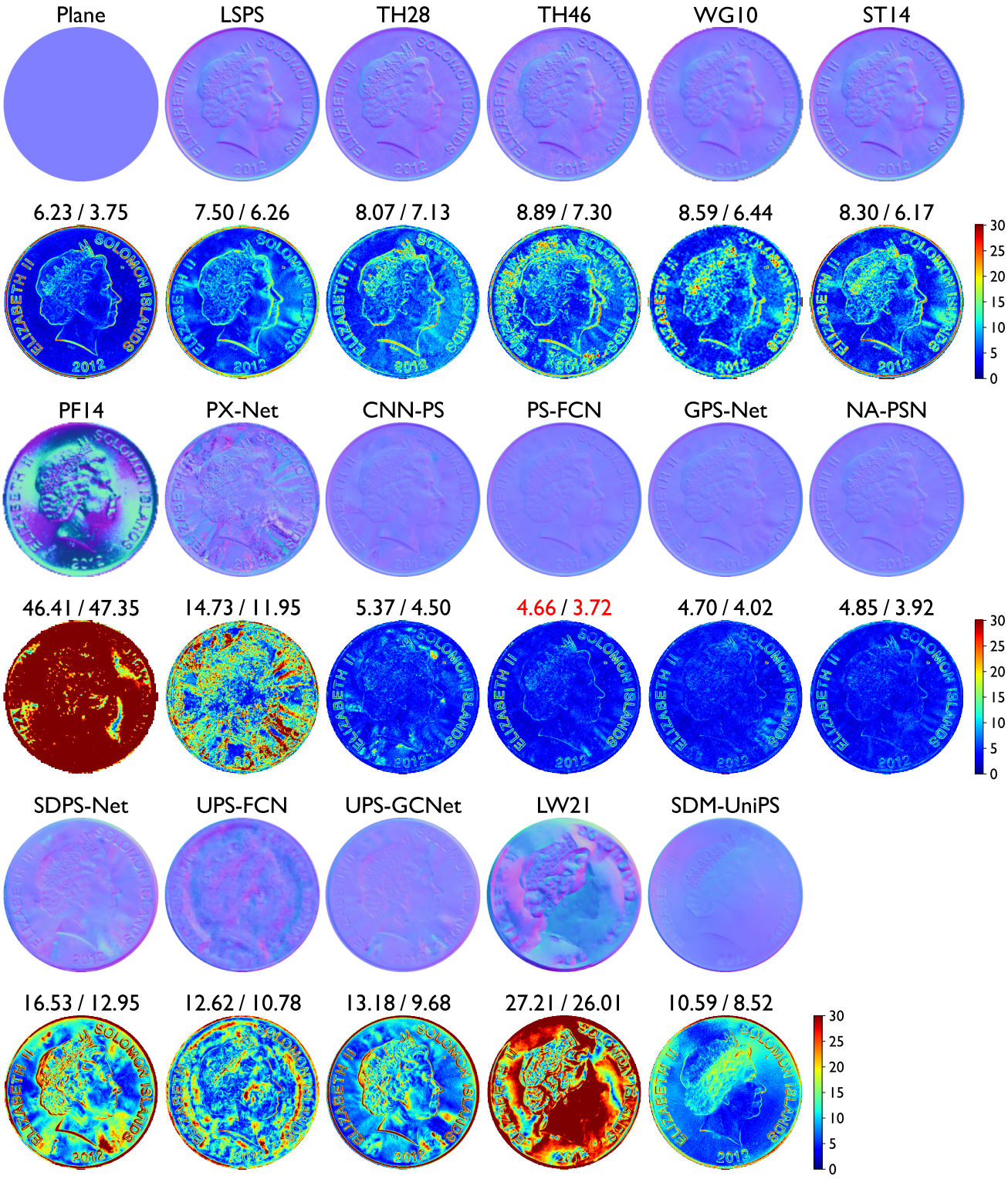


Figure S11: Estimated surface normal maps and angular error maps of QUEEN.

CRAB

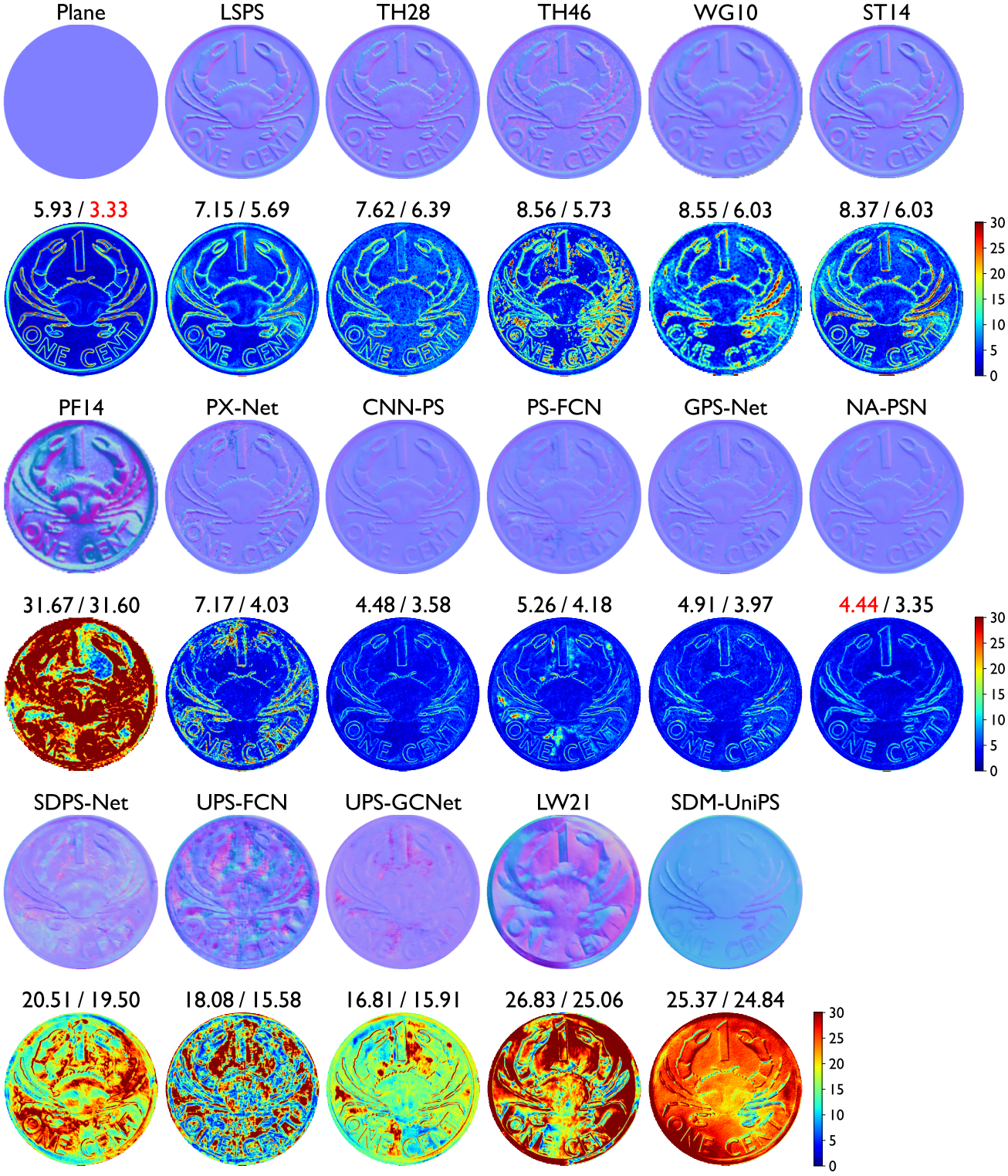


Figure S12: Estimated surface normal maps and angular error maps of CRAB.

SHIP

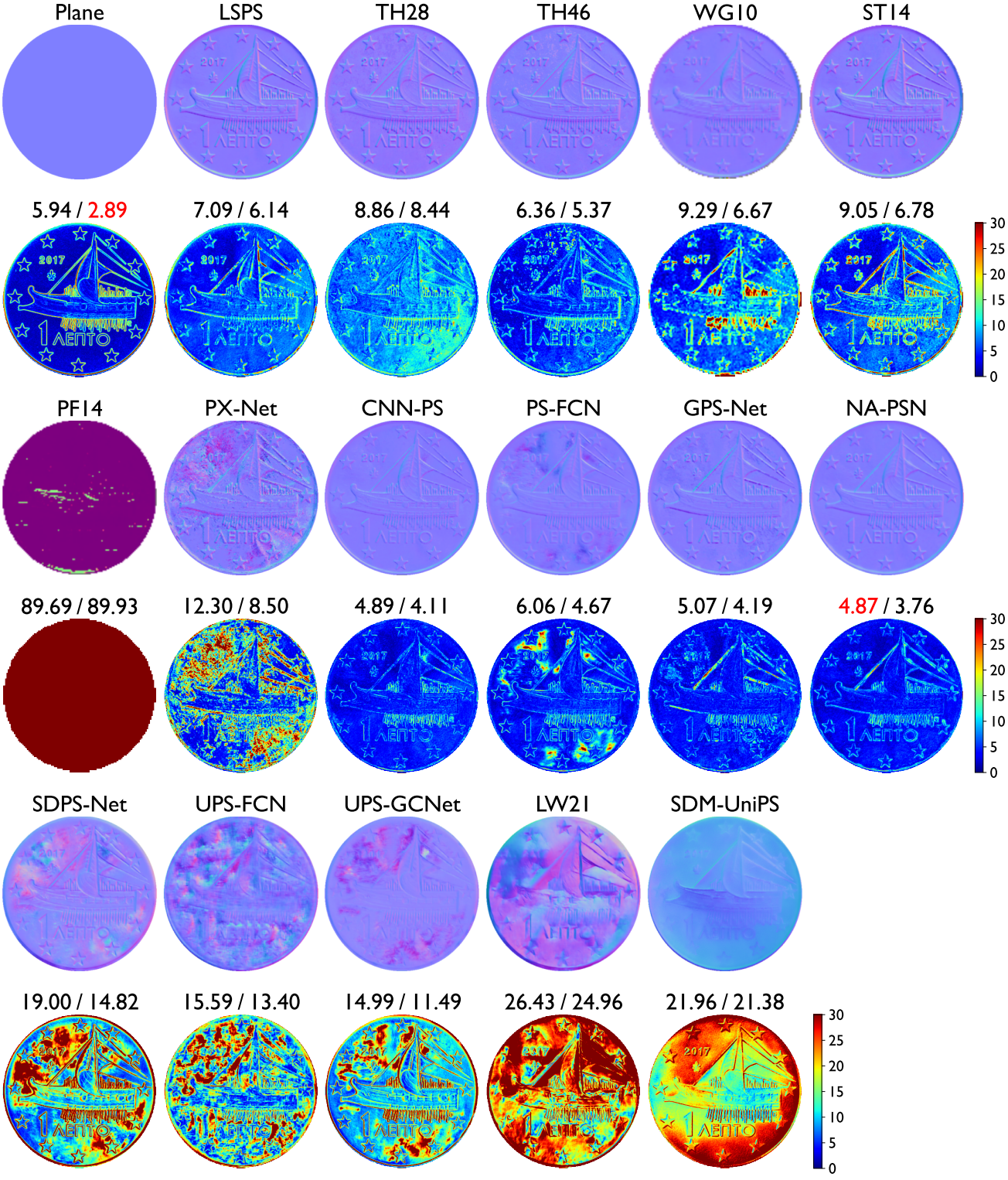


Figure S13: Estimated surface normal maps and angular error maps of SHIP.

PARA

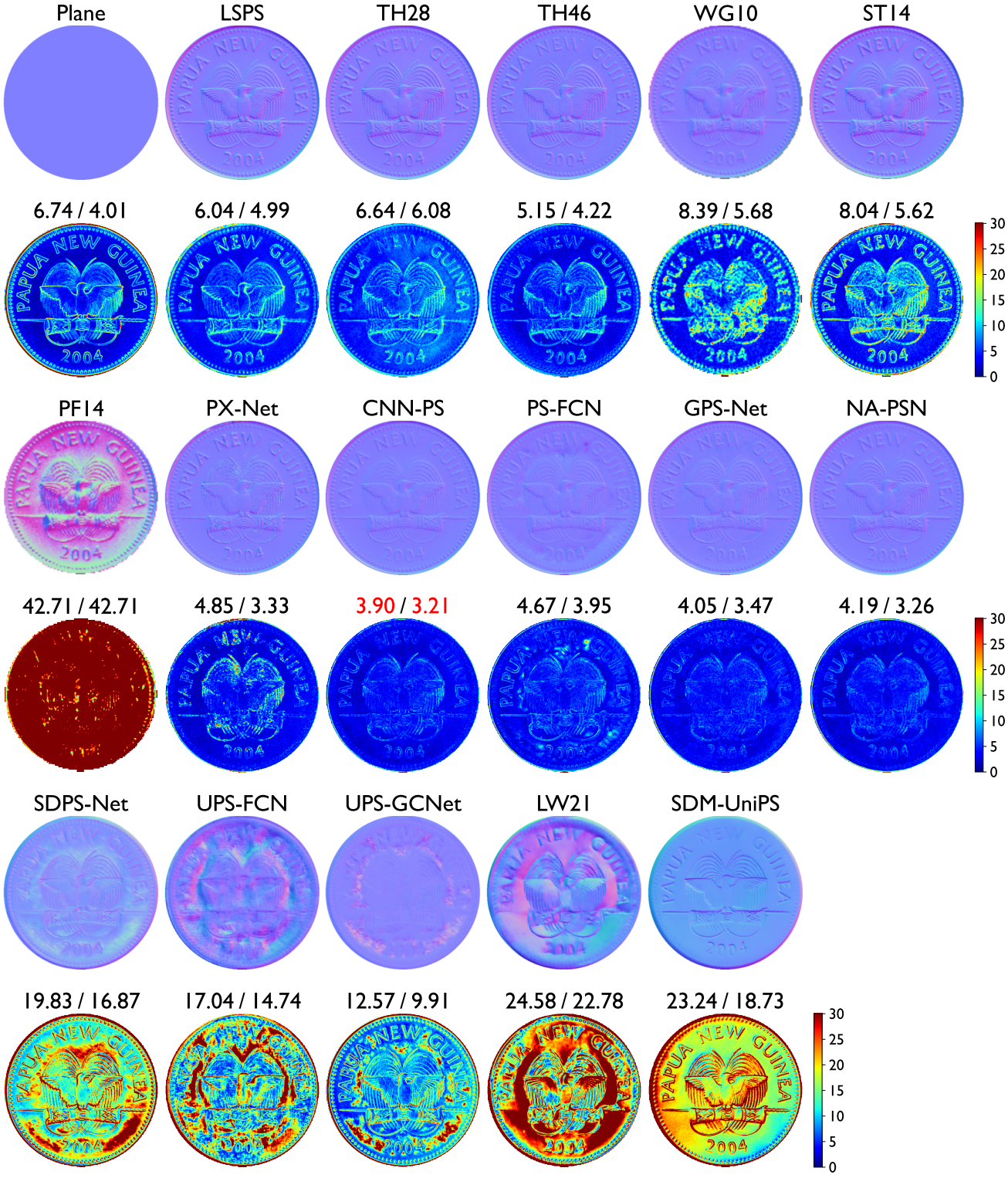


Figure S14: Estimated surface normal maps and angular error maps of PARA.

SAIL

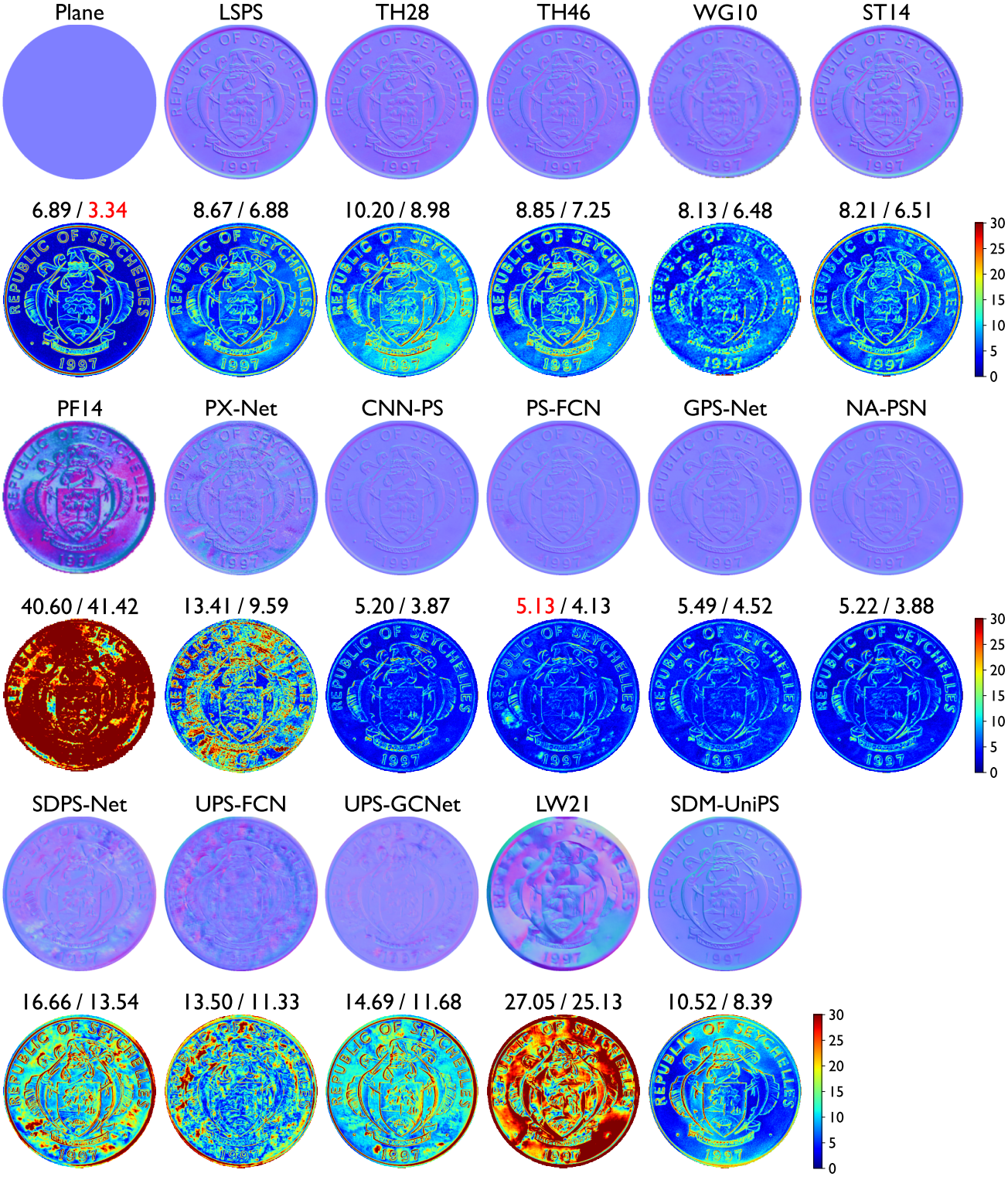


Figure S15: Estimated surface normal maps and angular error maps of SAIL.

FISH

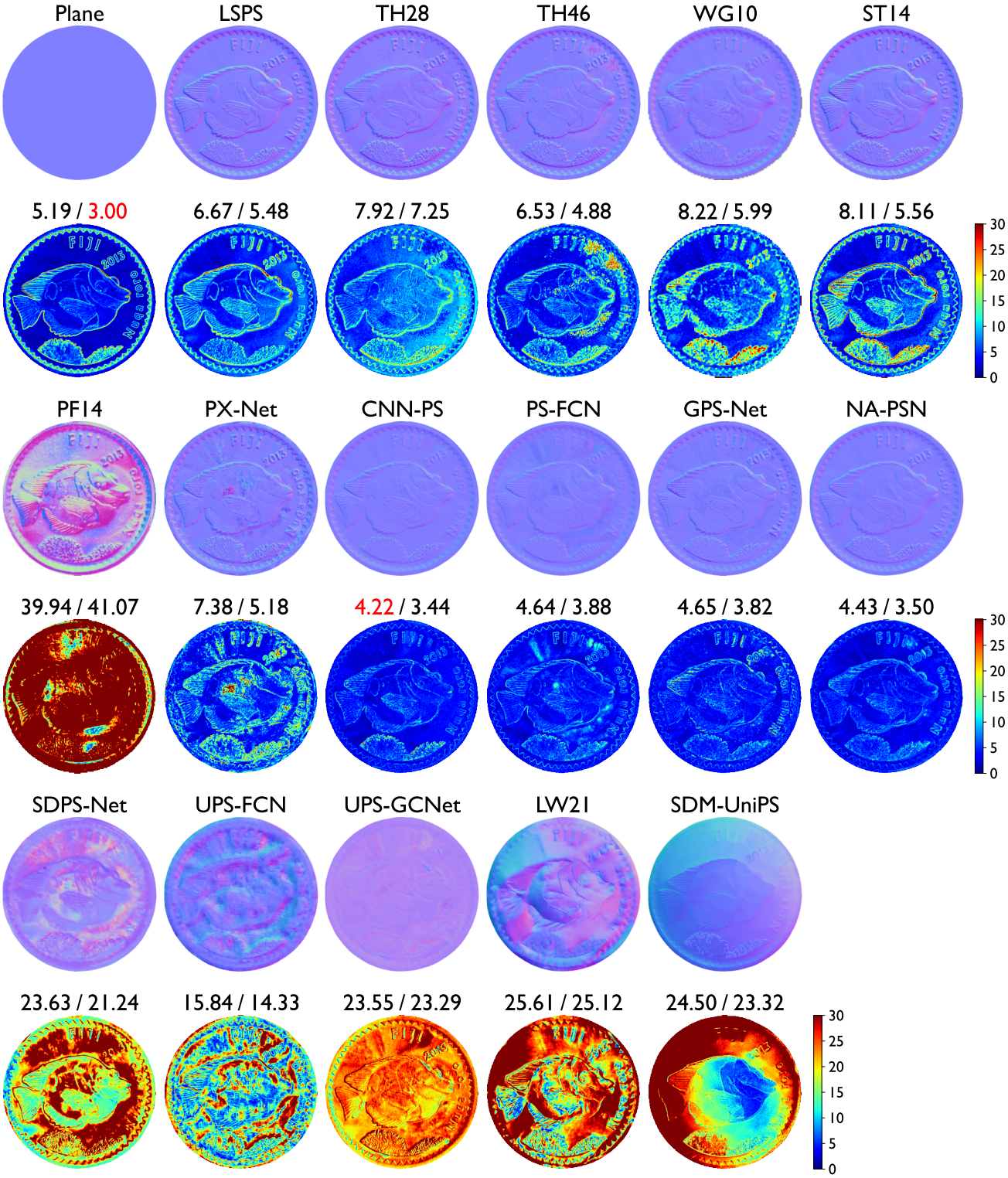


Figure S16: Estimated surface normal maps and angular error maps of FISH.

TREE

Figure S17: Estimated surface normal maps and angular error maps of TREE.

OCEAN

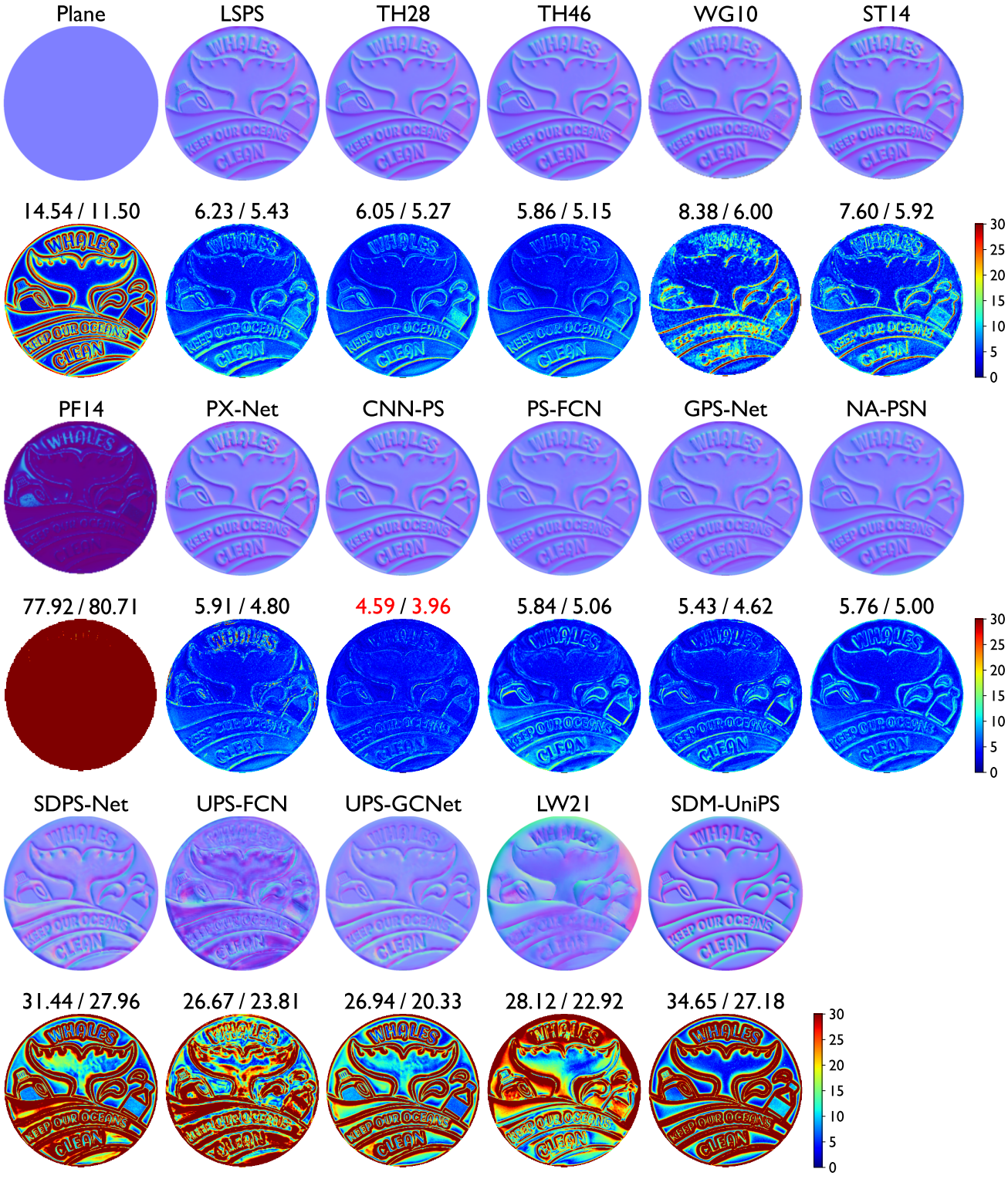


Figure S18: Estimated surface normal maps and angular error maps of OCEAN.

LUNG

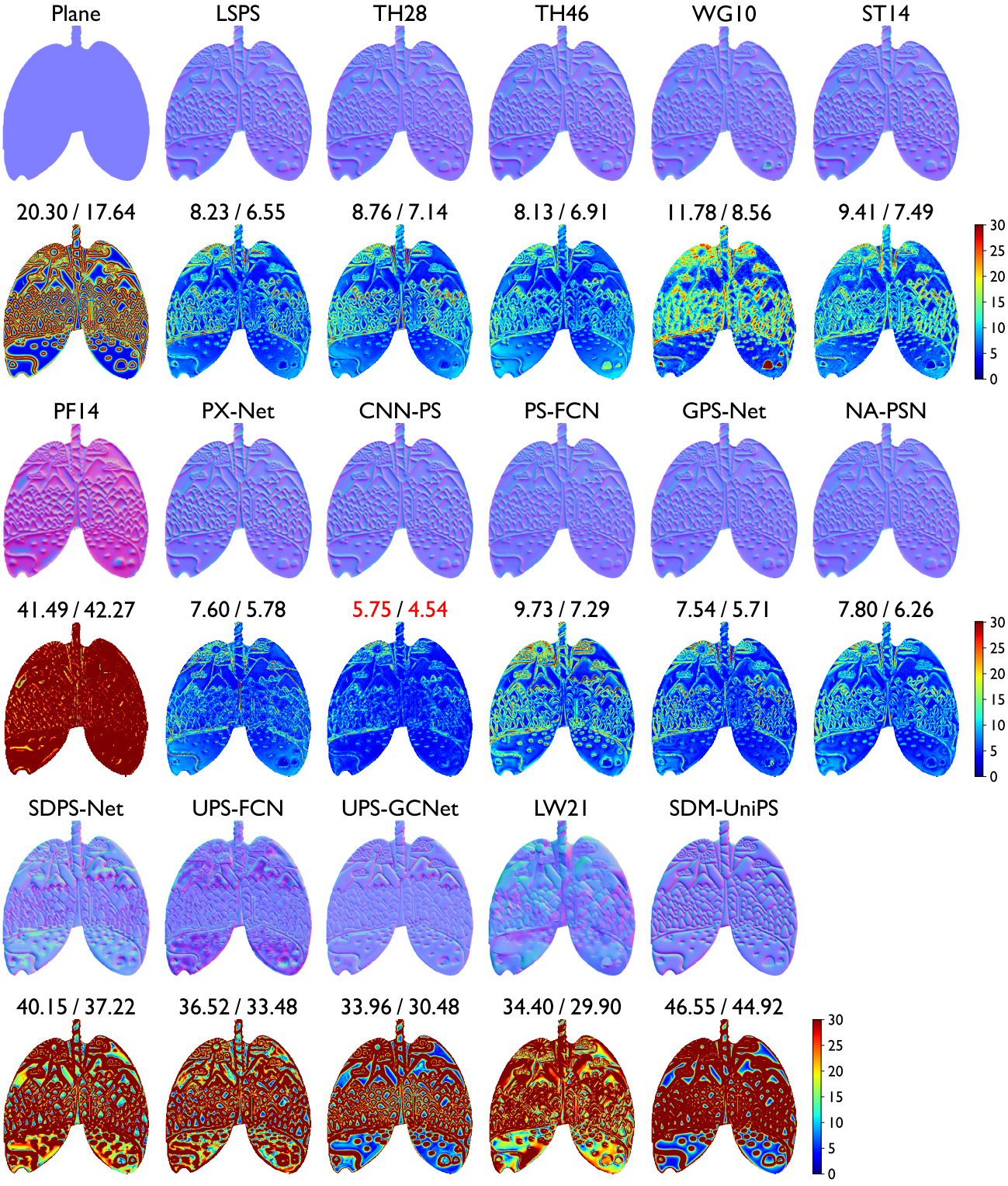


Figure S19: Estimated surface normal maps and angular error maps of LUNG.

BEAR

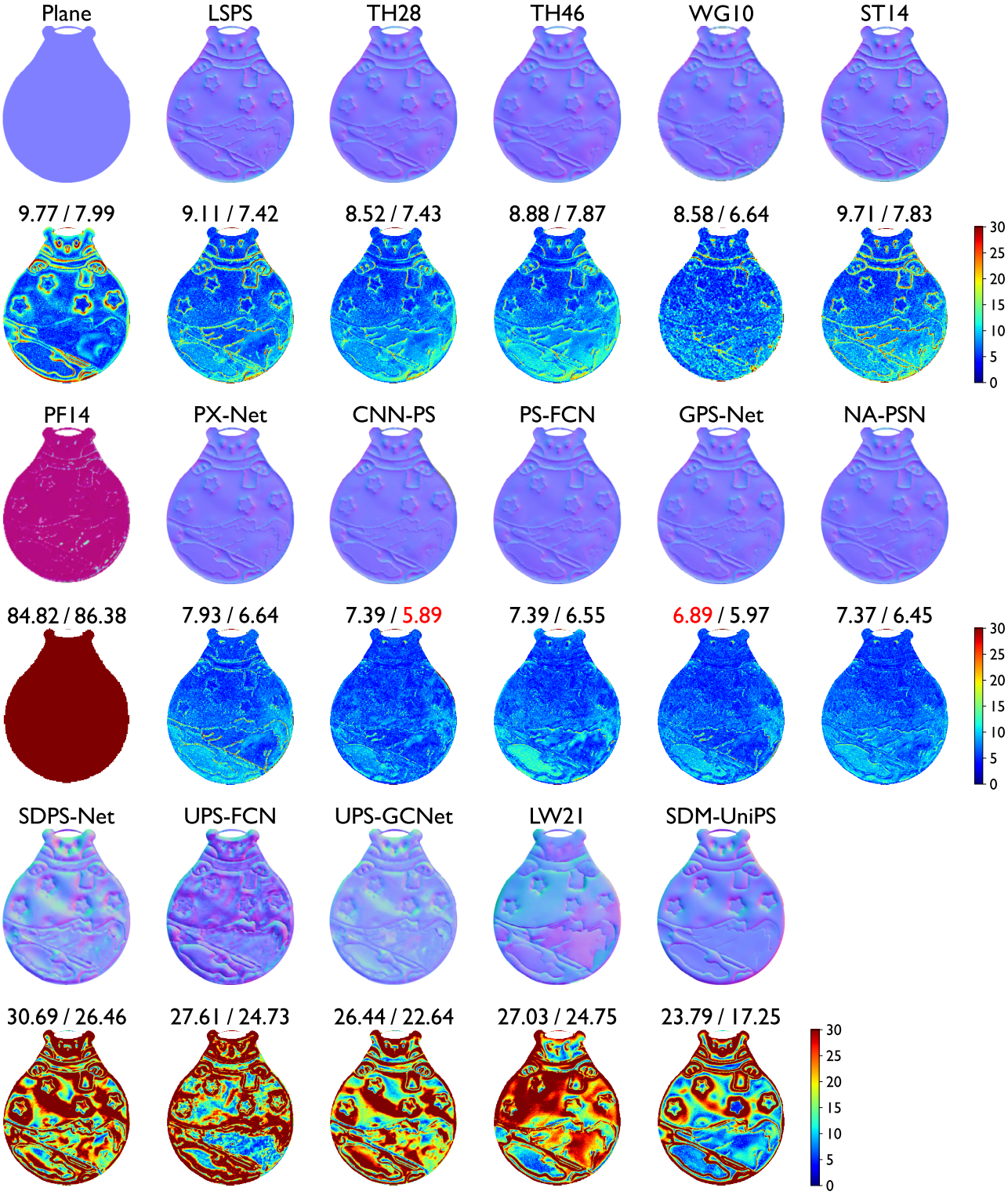


Figure S20: Estimated surface normal maps and angular error maps of BEAR.

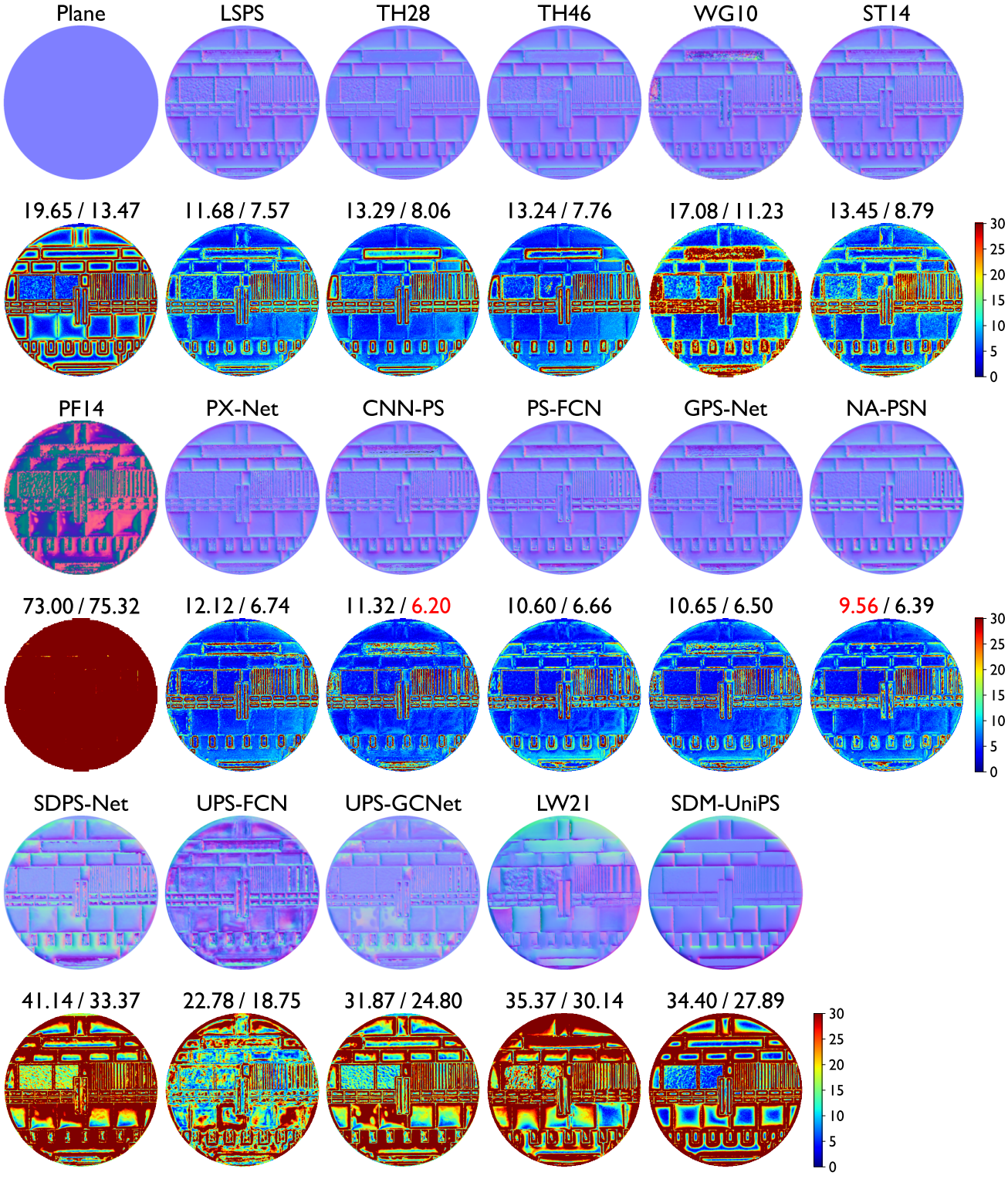


Figure S21: Estimated surface normal maps and angular error maps of Tv.

SUN

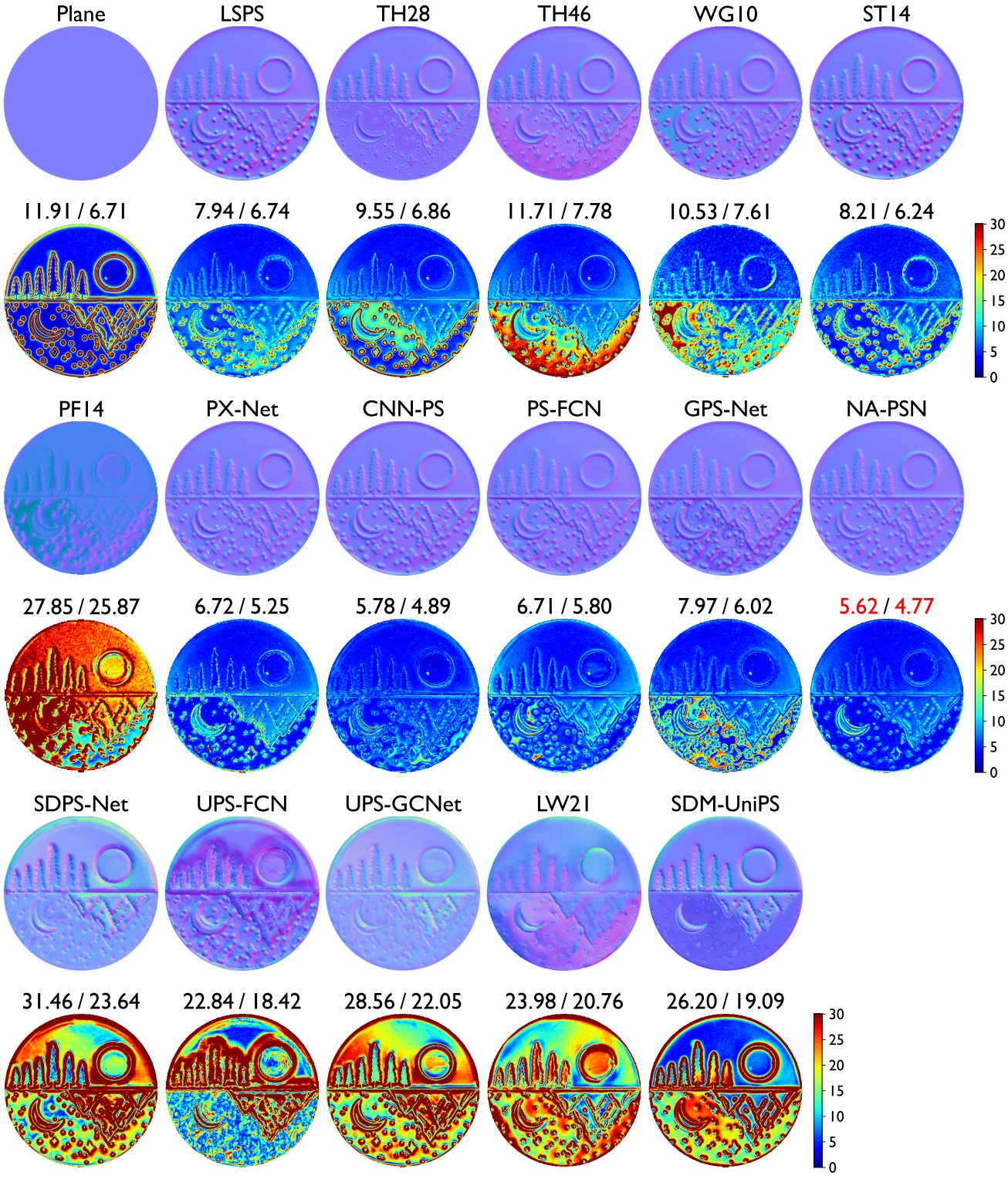


Figure S22: Estimated surface normal maps and angular error maps of SUN.

TAICHI

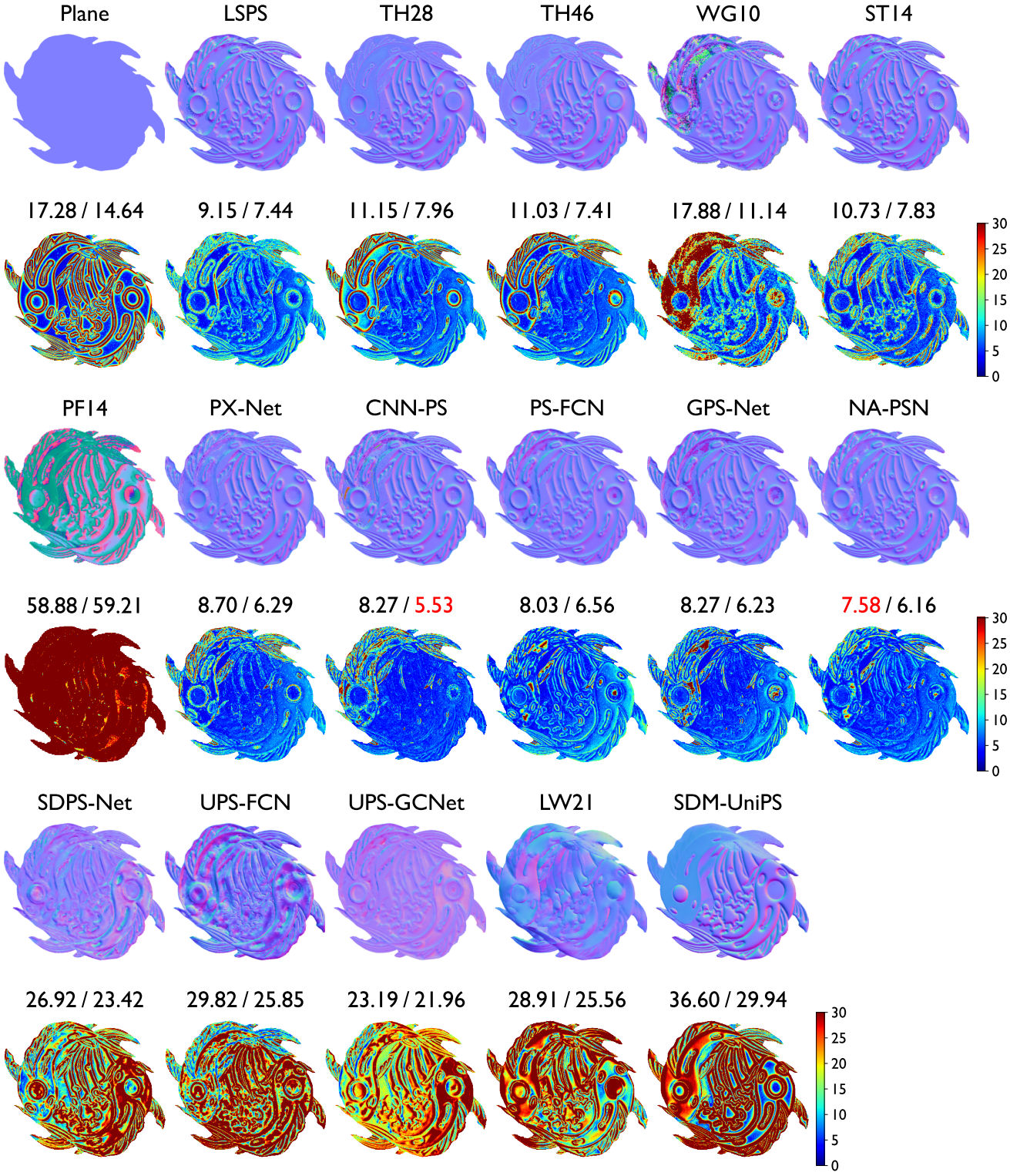


Figure S23: Estimated surface normal maps and angular error maps of TAICHI.

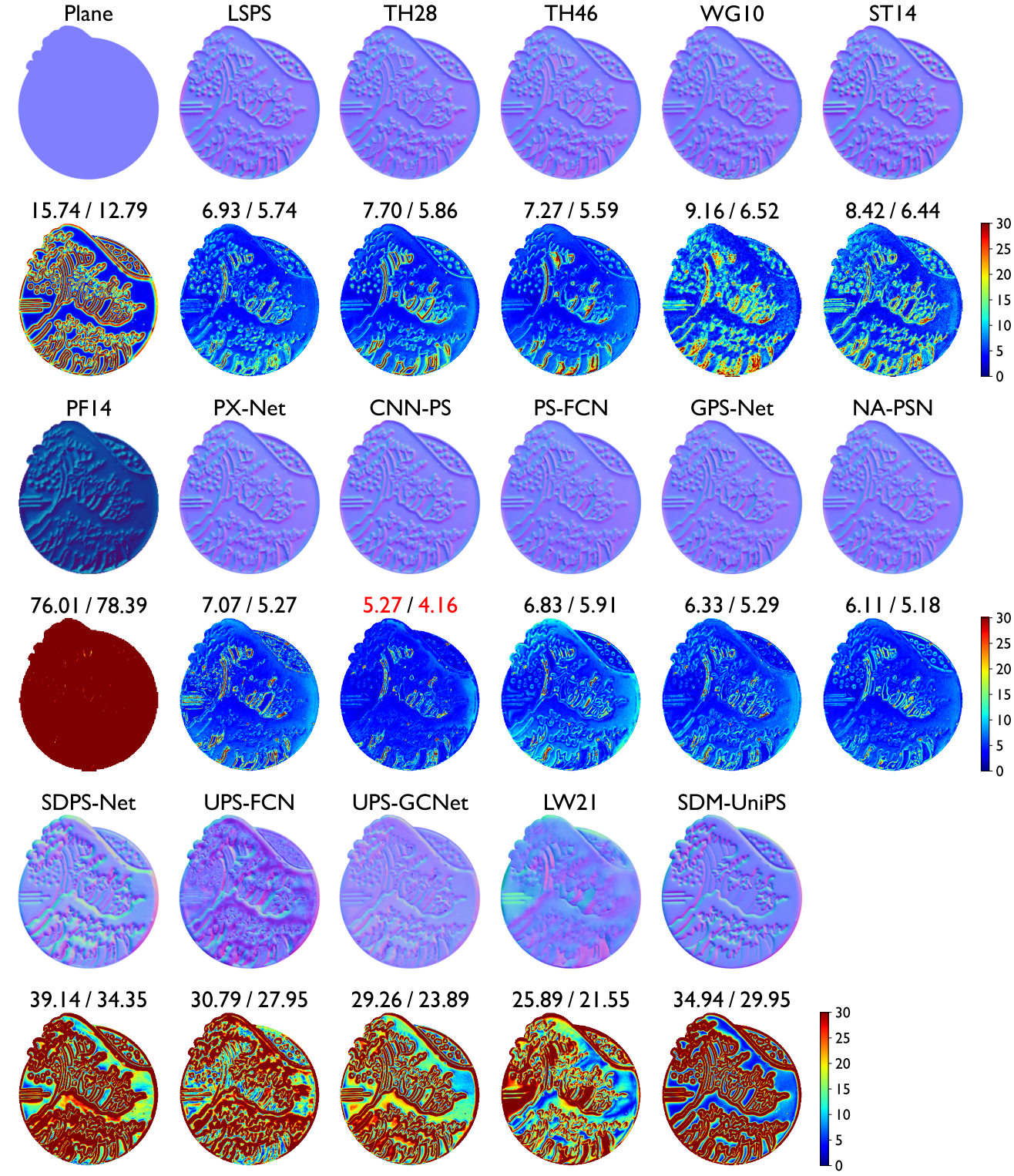


Figure S24: Estimated surface normal maps and angular error maps of WAVE.

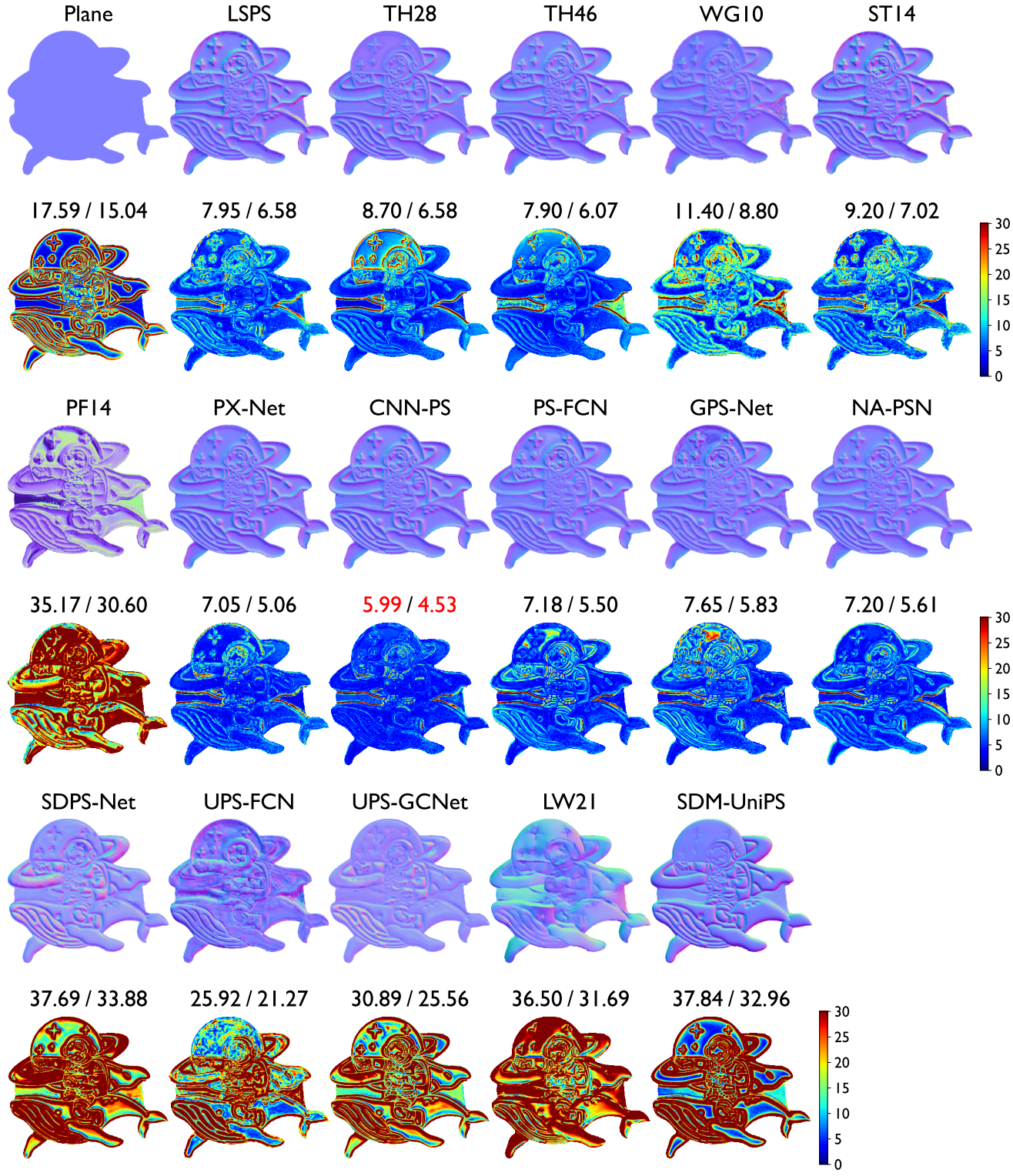


Figure S25: Estimated surface normal maps and angular error maps of ASTRO.

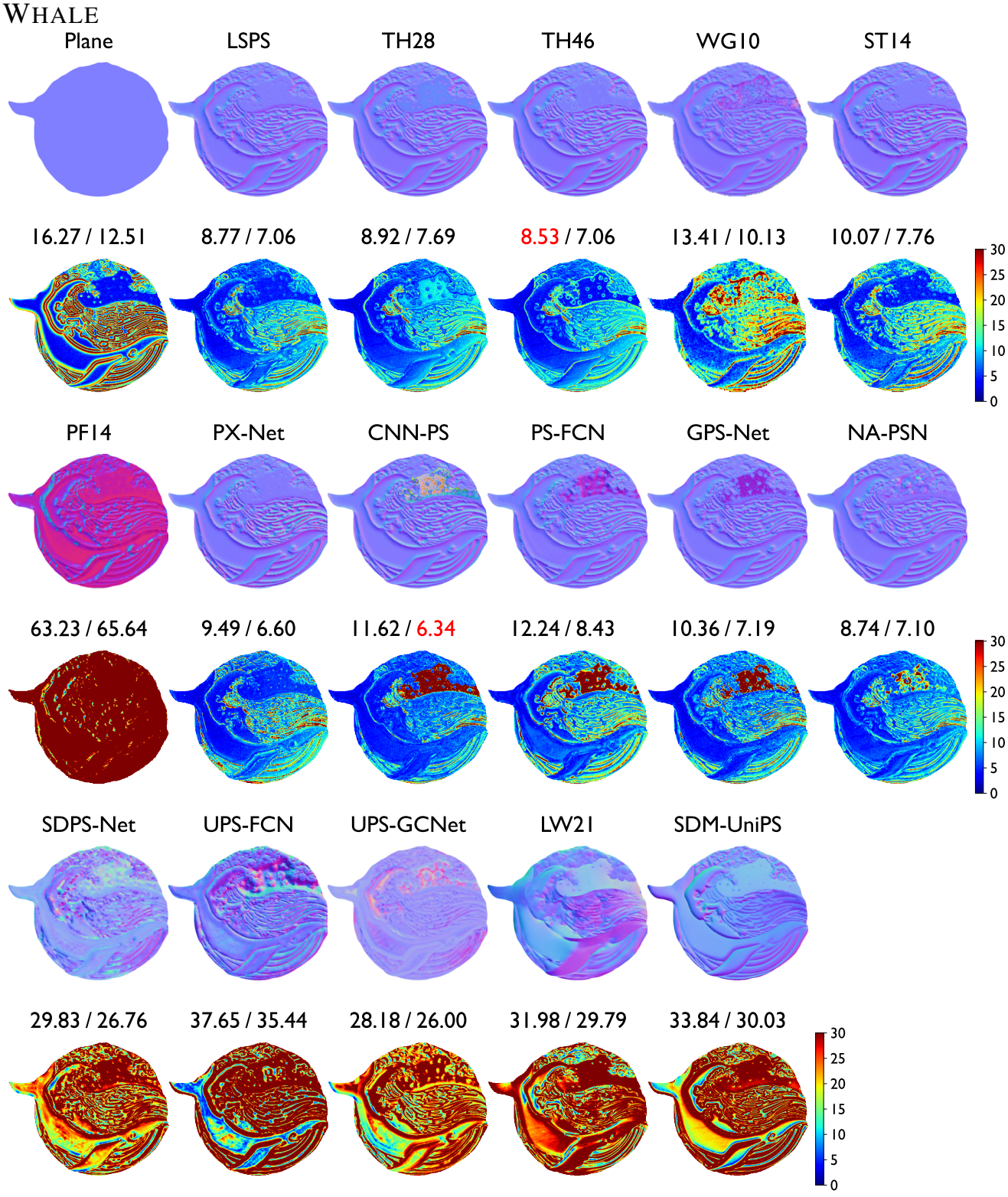


Figure S26: Estimated surface normal maps and angular error maps of WHALE.

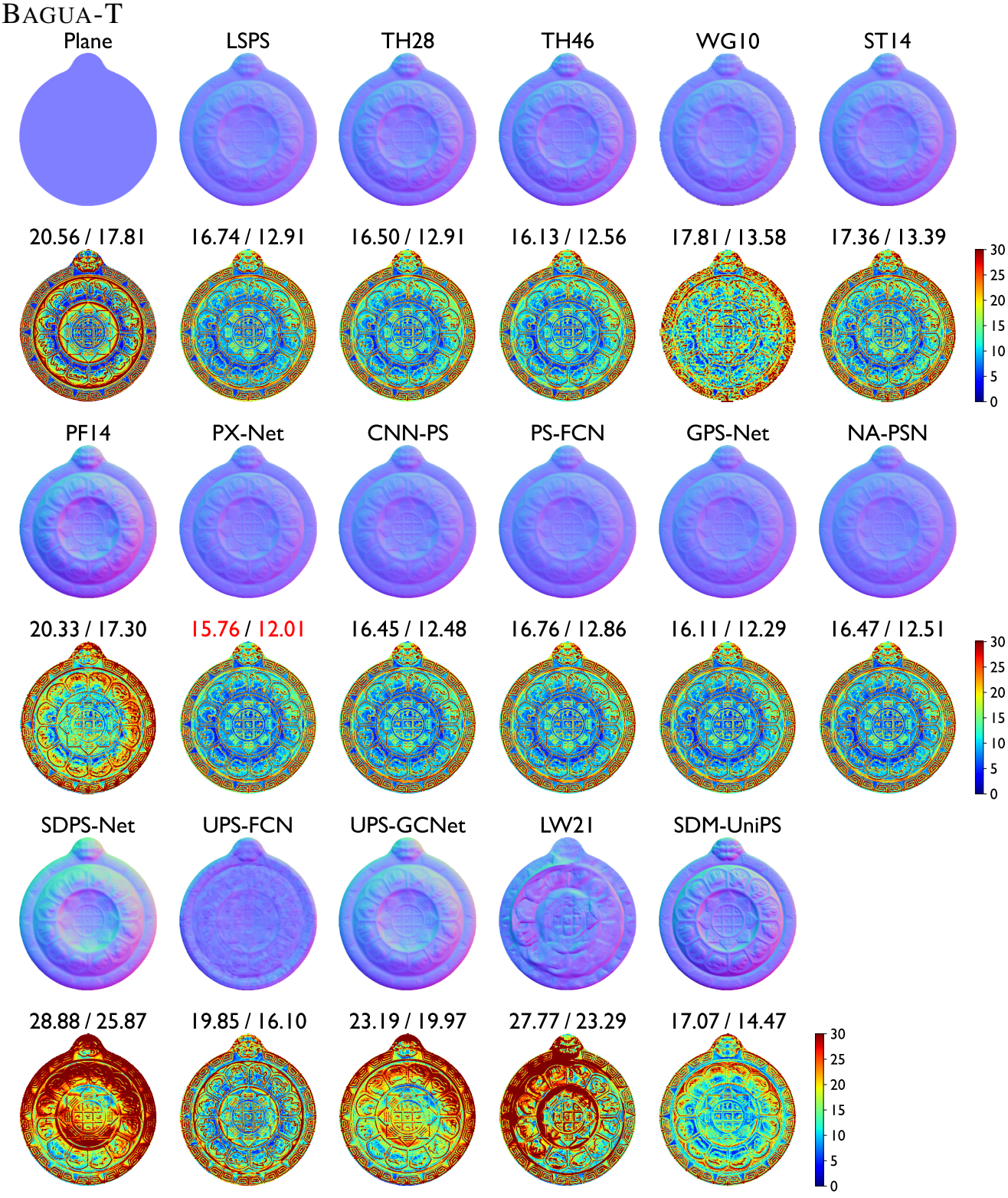


Figure S27: Estimated surface normal maps and angular error maps of BAGUA-T.

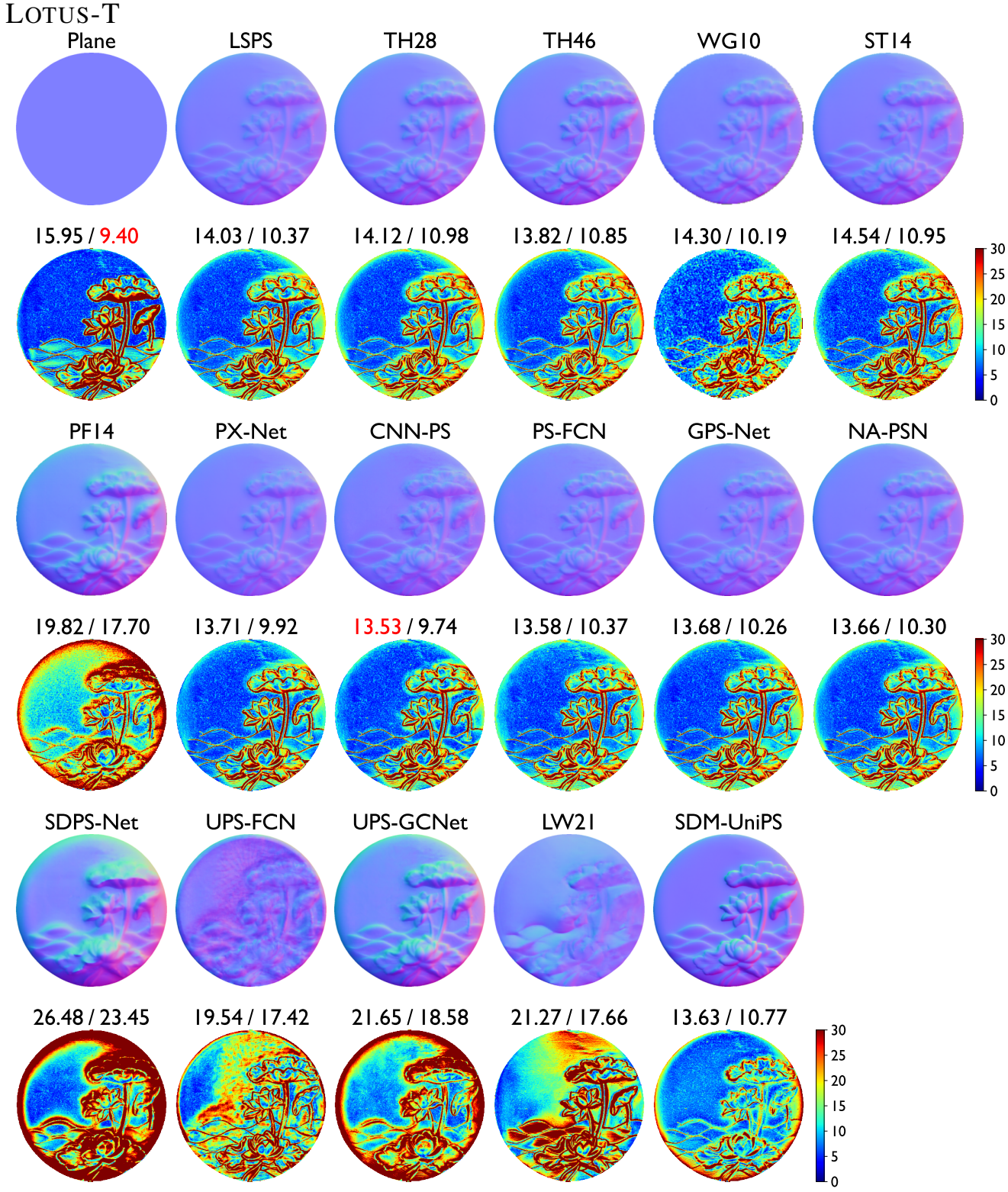


Figure S28: Estimated surface normal maps and angular error maps of LOTUS-T.

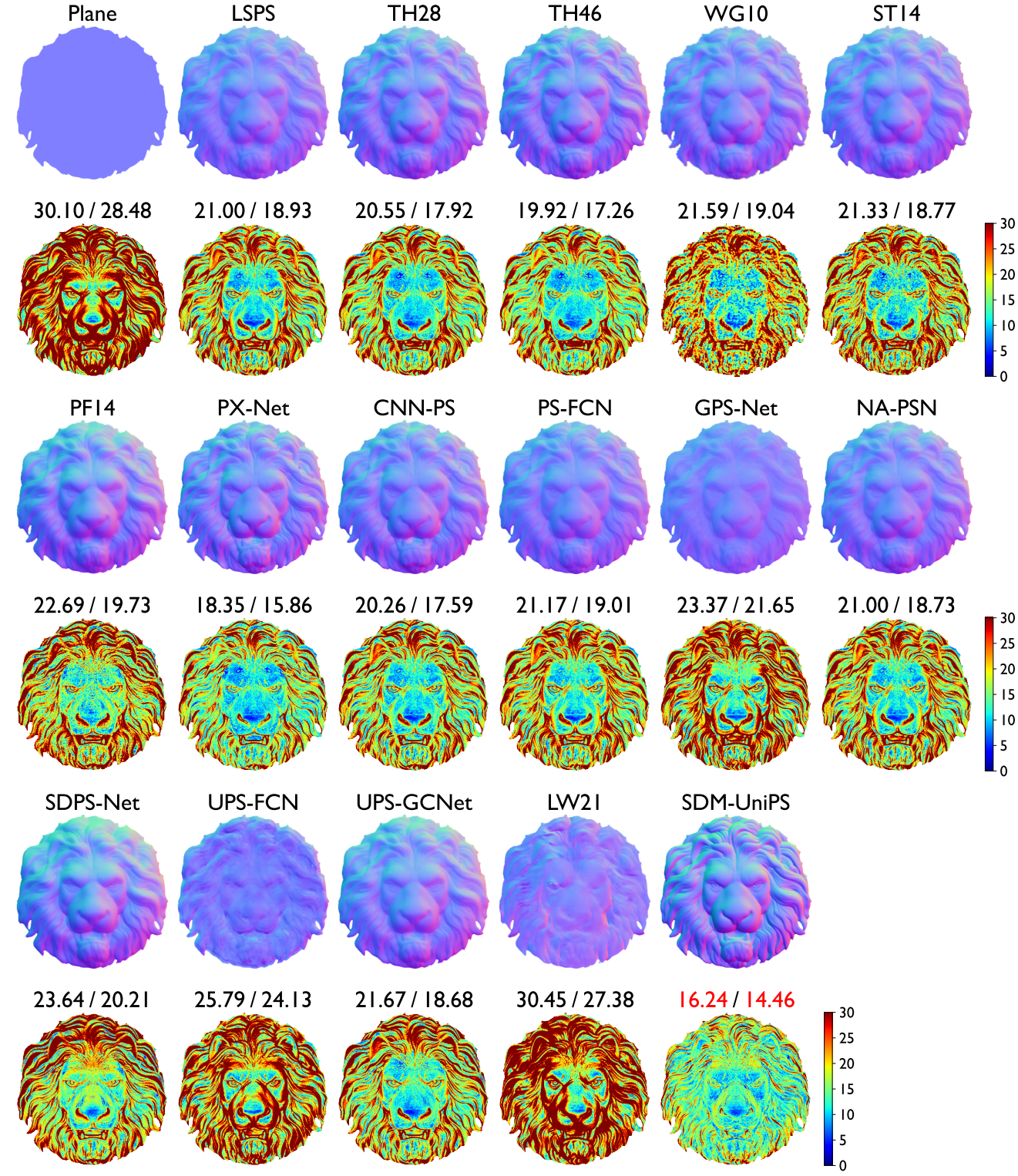


Figure S29: Estimated surface normal maps and angular error maps of LION-T.

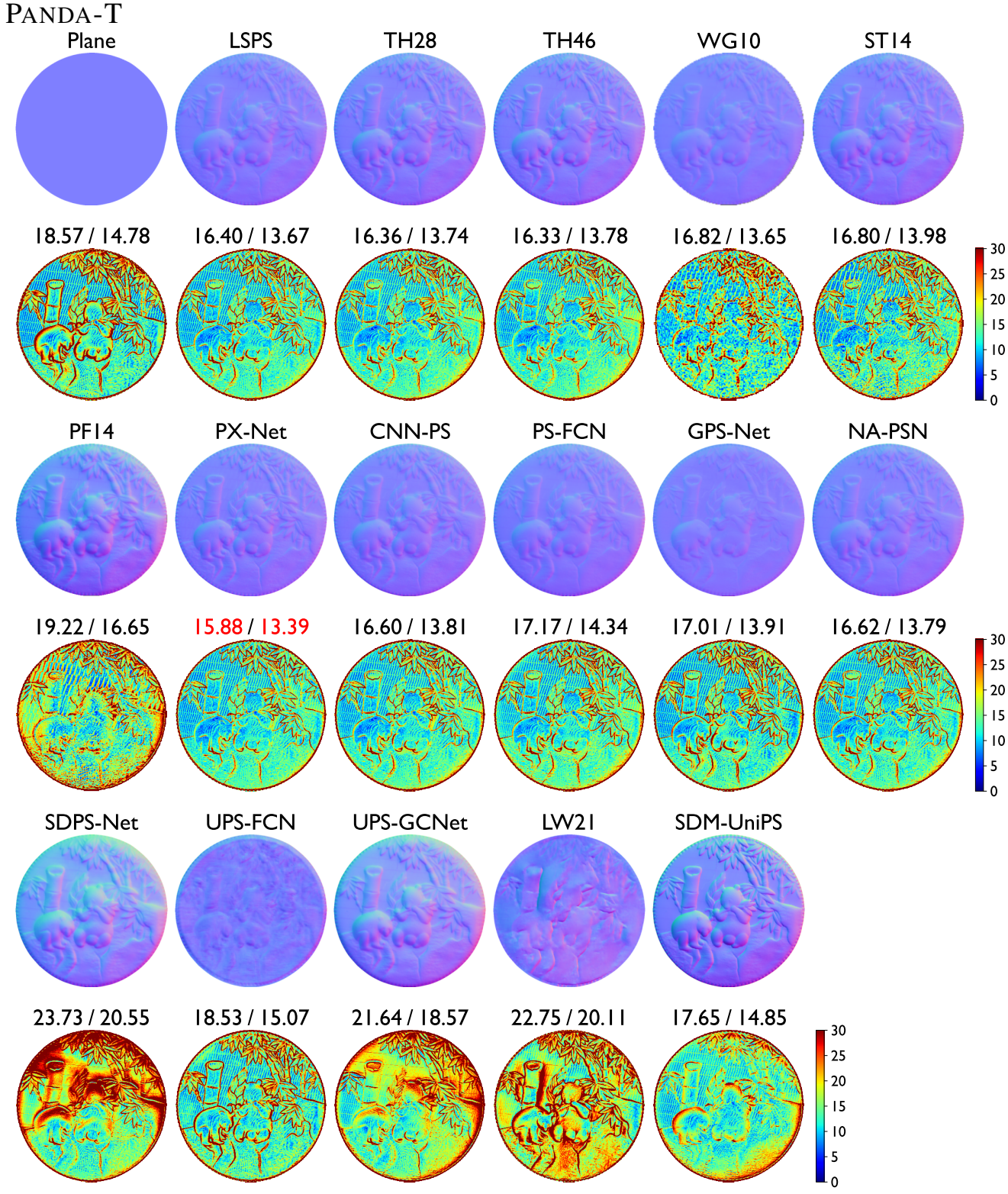


Figure S30: Estimated surface normal maps and angular error maps of PANDA-T.

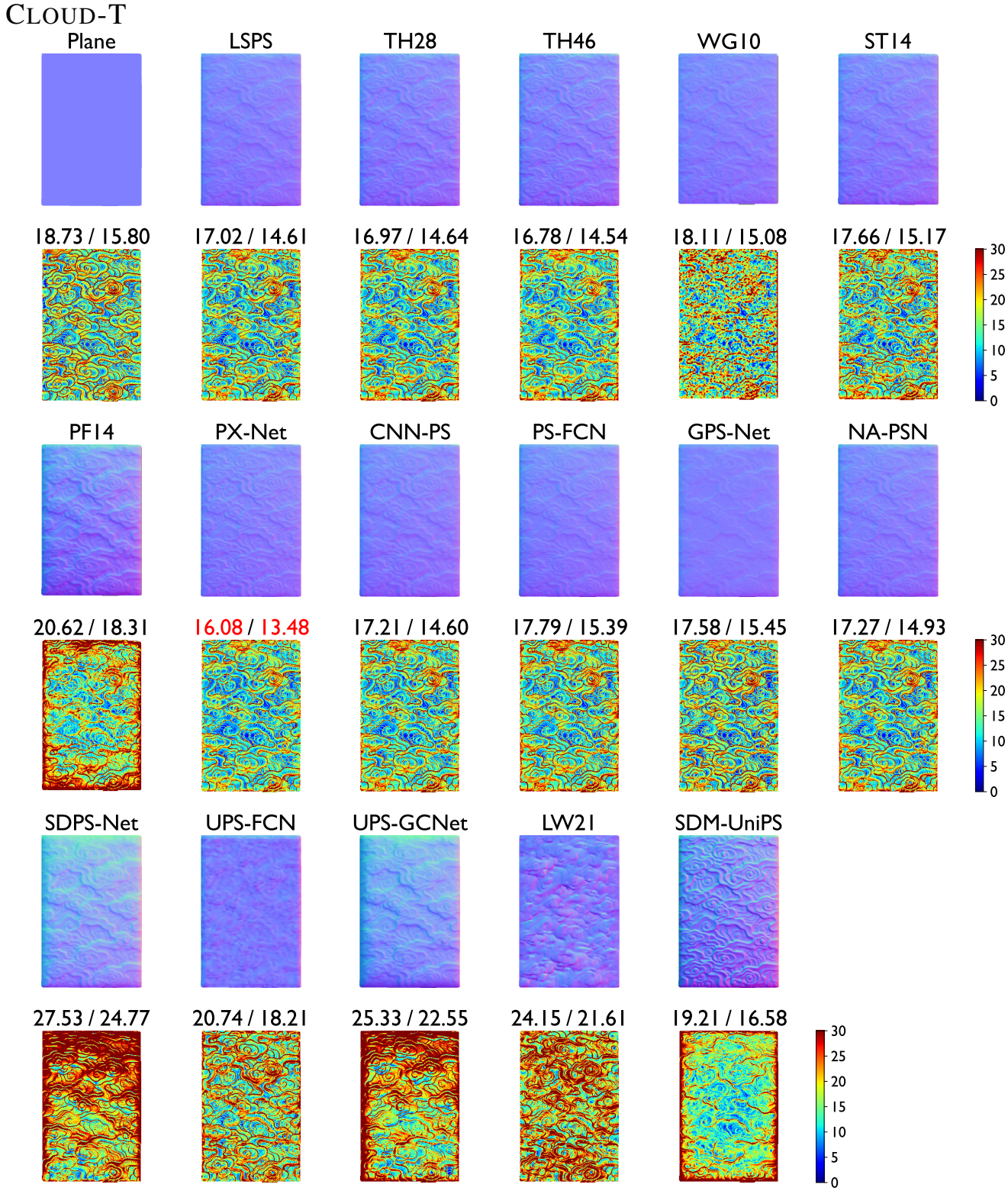


Figure S31: Estimated surface normal maps and angular error maps of CLOUD-T.

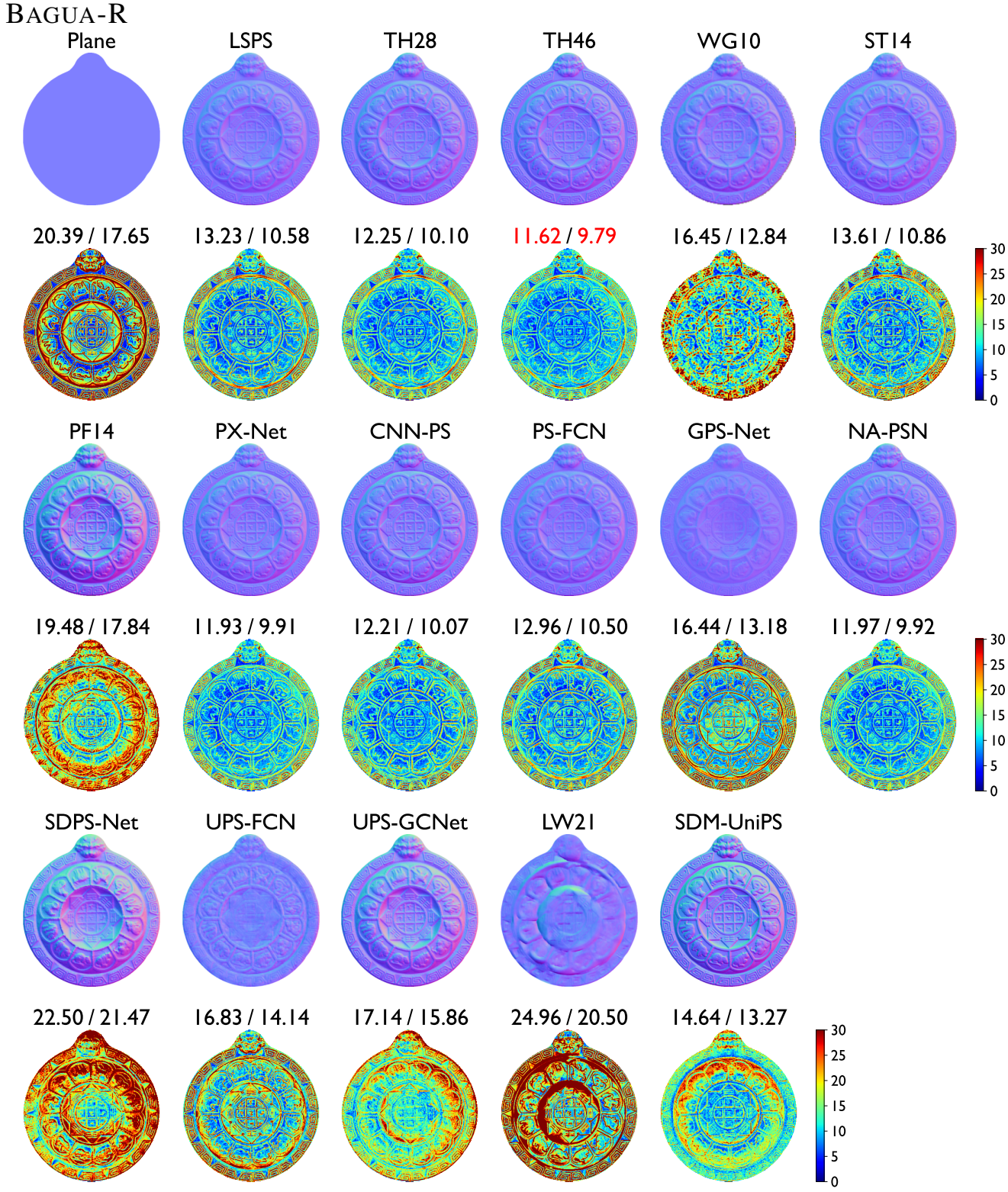


Figure S32: Estimated surface normal maps and angular error maps of BAGUA-R.

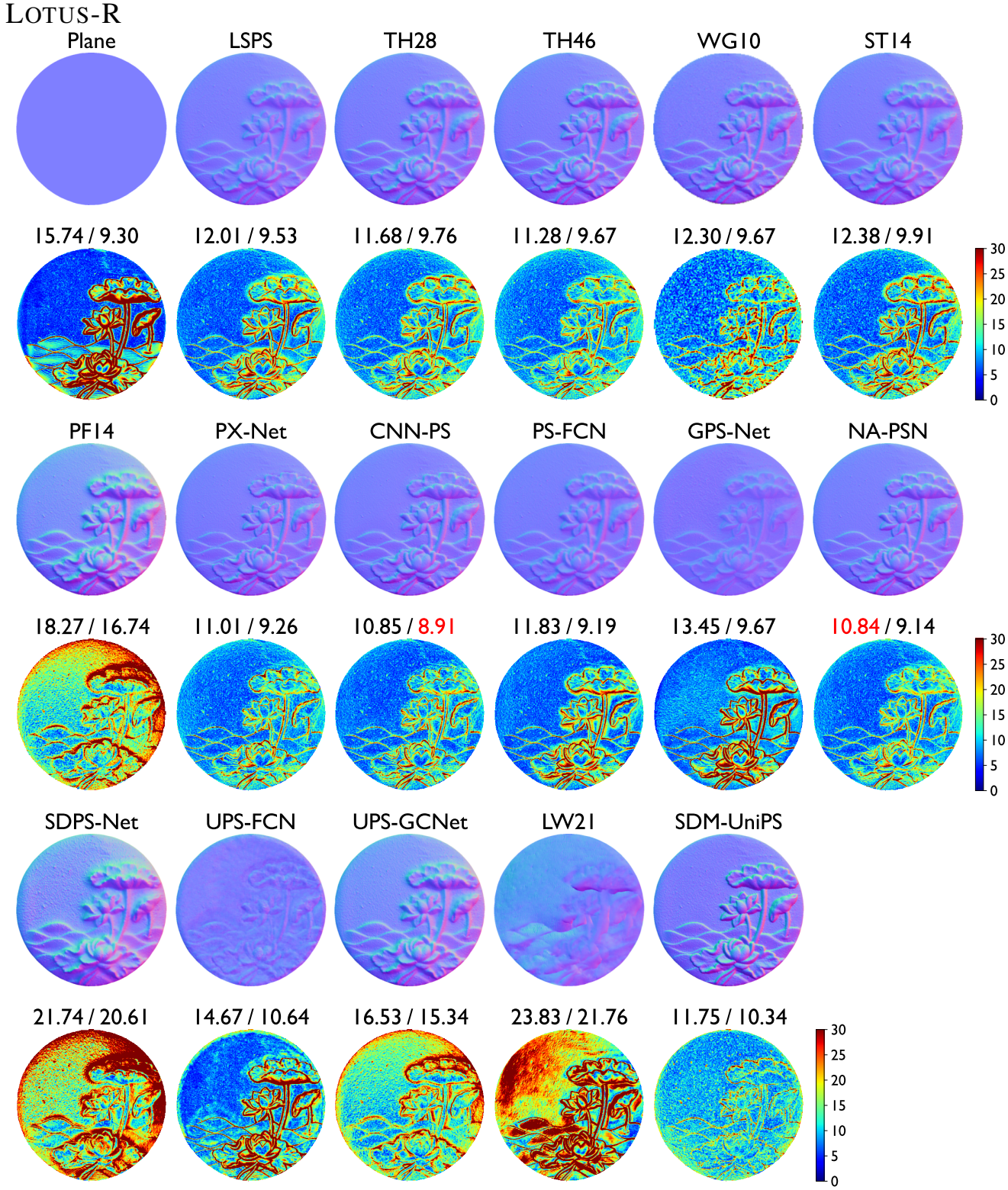


Figure S33: Estimated surface normal maps and angular error maps of LOTUS-R.

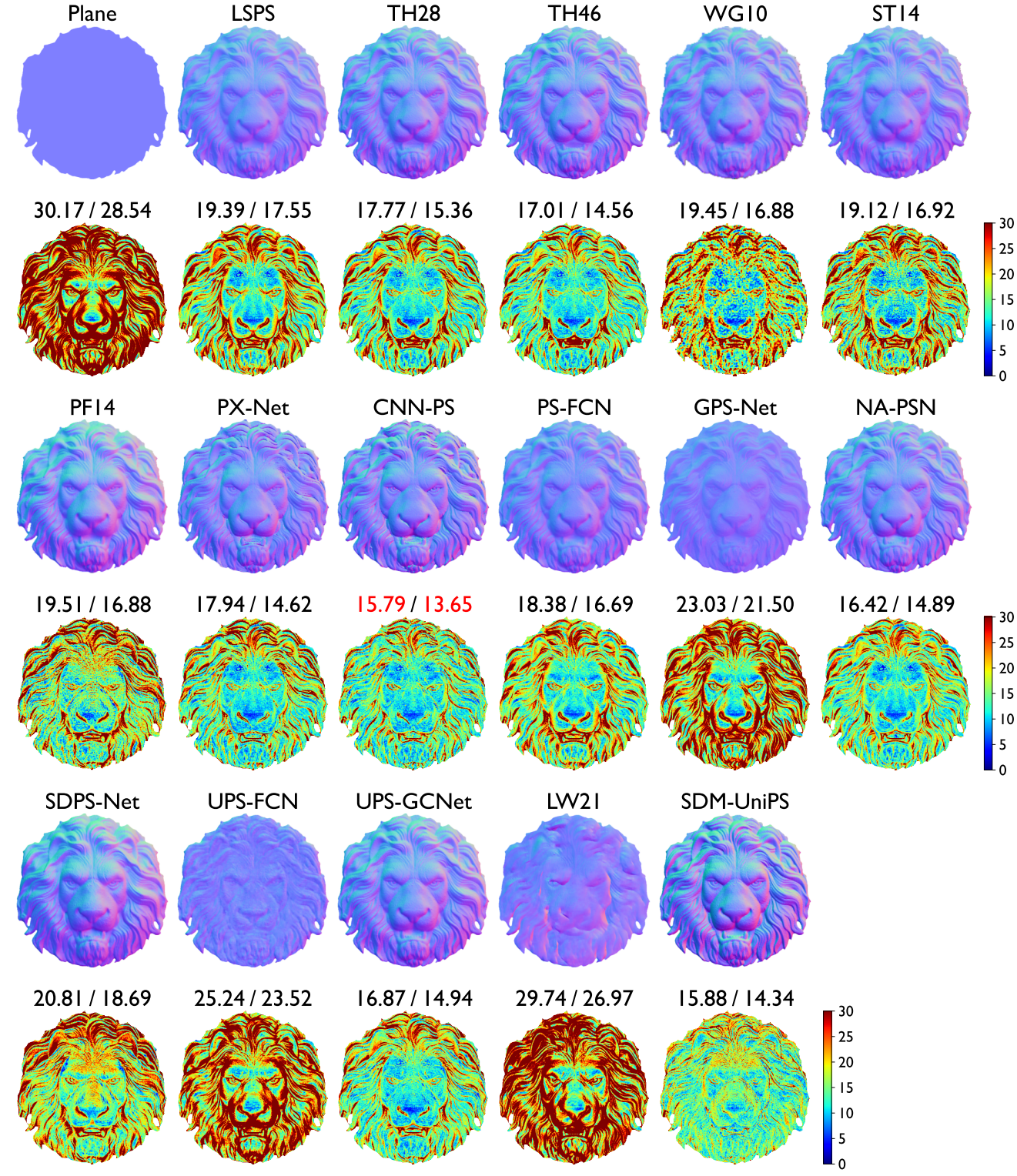


Figure S34: Estimated surface normal maps and angular error maps of LION-R.

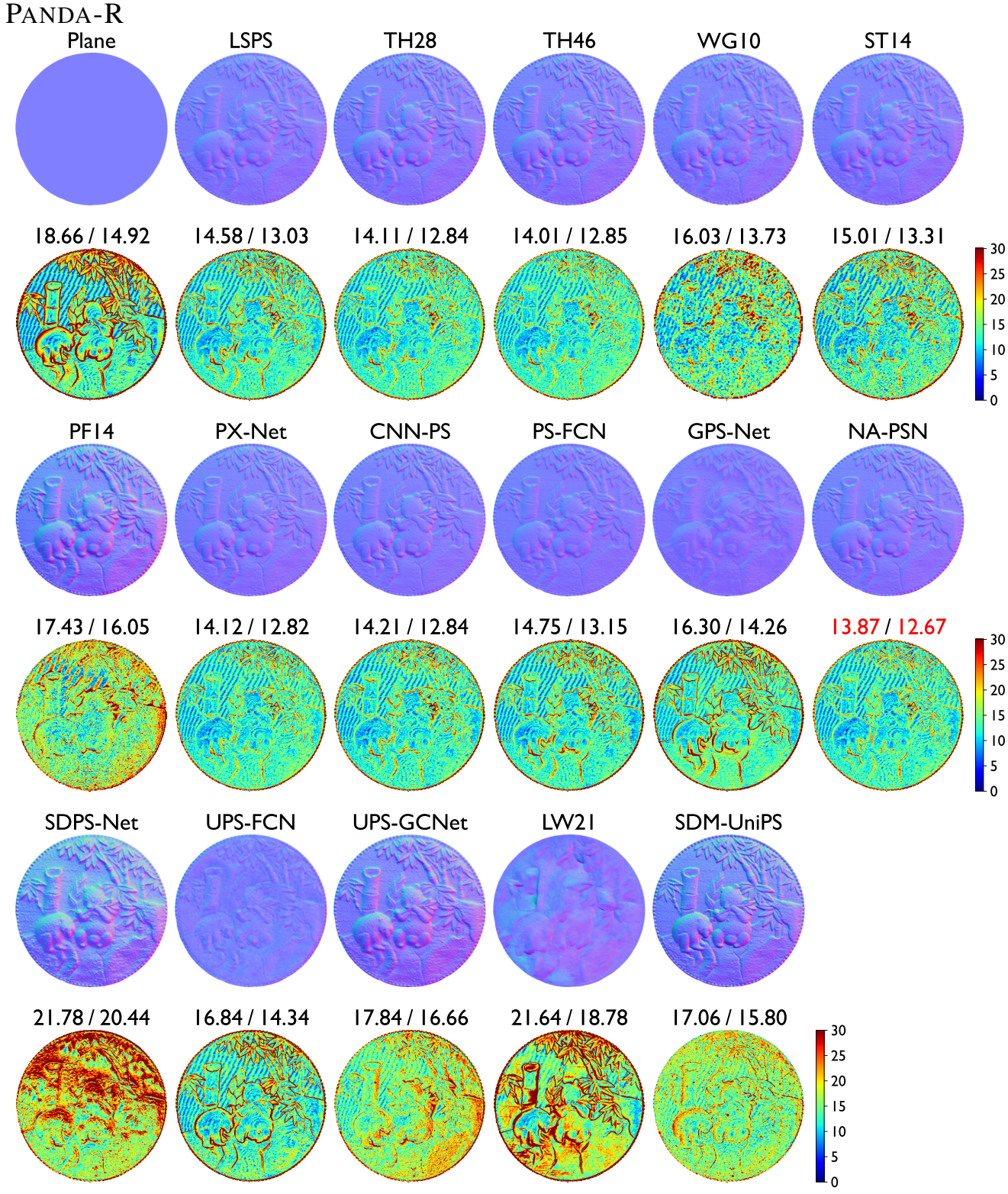


Figure S35: Estimated surface normal maps and angular error maps of PANDA-R.

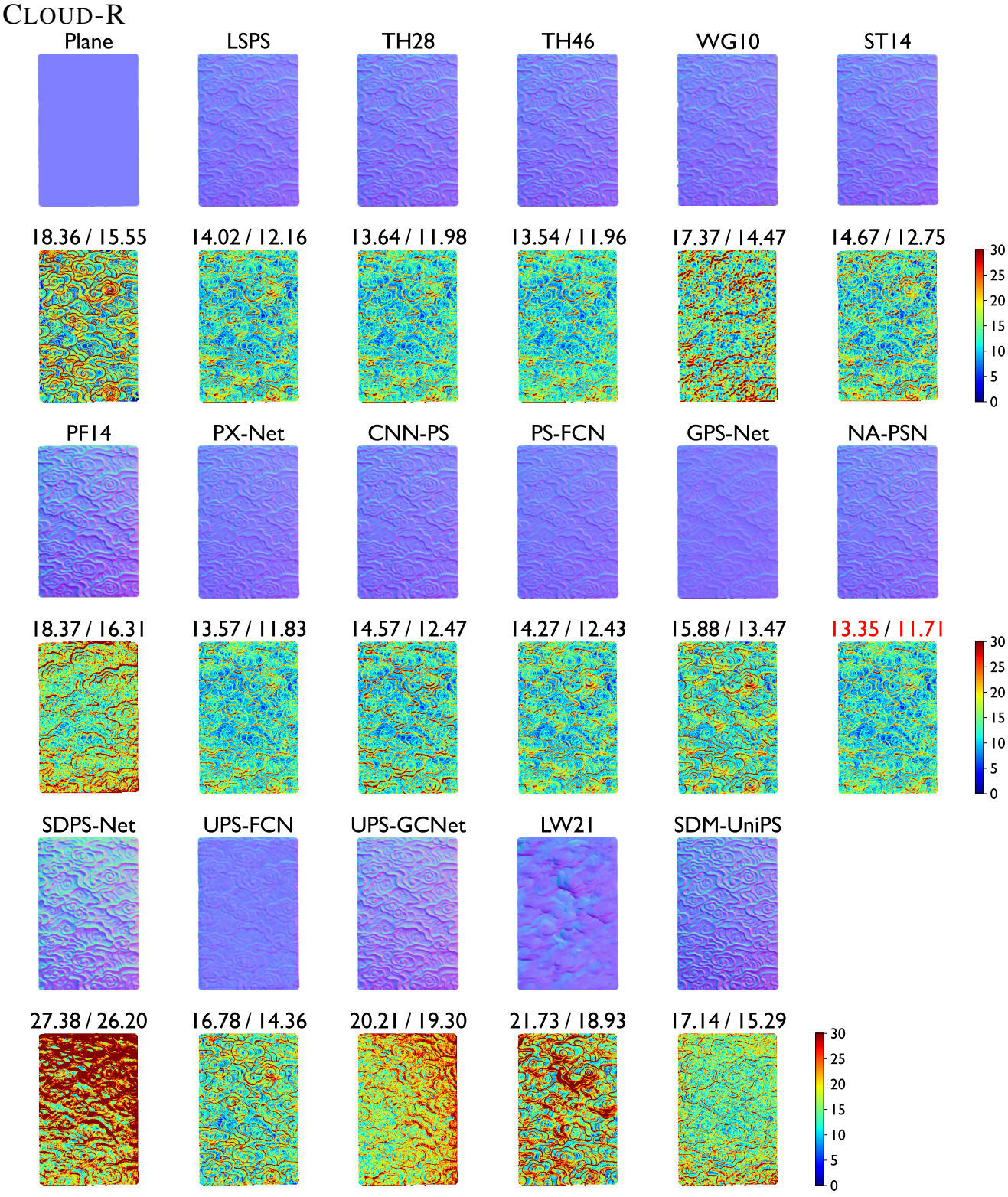


Figure S36: Estimated surface normal maps and angular error maps of CLOUD-R.

References

[1] 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, 5

[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, 5

[3] 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)*, 2020. 5

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

[5] Satoshi Ikehata. Scalable, detailed and mask-free universal photometric stereo. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 5

[6] 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*, 2022. 4, 5

[7] Daniel Lichy, Jiaye Wu, Soumyadip Sengupta, and David W Jacobs. Shape and material capture at home. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 5

[8] Fotios Logothetis, Ignas Budvytis, Roberto Mecca, and Roberto Cipolla. PX-NET: simple and efficient pixel-wise training of photometric stereo networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021. 5

[9] 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, 5

[10] 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. 2

[11] 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, 5

[12] 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, 5

[13] Robert J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 1980. 4, 5

[14] Tai-Pang Wu and Chi-Keung Tang. Photometric stereo via expectation maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010. 4, 5

[15] Zhuokun Yao, Kun Li, Ying Fu, Haofeng Hu, and Boxin Shi. GPS-Net: Graph-based photometric stereo network. *Advances in Neural Information Processing Systems*, 2020. 4, 5