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

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

Abstract

Photometric stereo aims to recover detailed surface shapes from images captured under varying illuminations. However, existing real-world datasets primarily focus on evaluating photometric stereo for general non-Lambertian reflectances and feature bulgy shapes that have a certain height. As shape detail recovery is the key strength of photometric stereo over other 3D reconstruction techniques, and the near-planar surfaces widely exist in cultural relics and manufacturing workpieces, we present a new real-world dataset DiLiGenT-Π containing 30 near-planar scenes with rich surface details. This dataset enables us to evaluate recent photometric stereo methods specifically for their ability to estimate shape details under diverse materials and to identify open problems such as near-planar surface normal estimation from uncalibrated photometric stereo and surface detail recovery for translucent materials. To inspire future research, this dataset will open soruced at https://photometricstereo.github.io/diligentpi.html.

1. Introduction

Photometric stereo [47, 44] aims at single view three-dimensional (3D) reconstruction from image observations captured under varying lights. Compared to structured light-based 3D reconstruction techniques that are widely applied in commercial scanners, the key strength of photometric stereo is the detailed surface shape recovery, which is of great interest for additive manufacturing and rendering.

To evaluate the effectiveness of photometric stereo methods, a batch of real-world benchmark datasets have been built such as DiLiGenT [43, 41] (and its multi-view extension DiLiGenT-MV [29]), DiLiGenT10^2 [39] for distant lights, and LUCES [34] for near lights. Existing datasets focus on evaluating the effectiveness of photometric stereo techniques on non-Lambertian surfaces, revealing the performance of existing methods on real-world scenes. Since the majority of target objects in these datasets are smooth surfaces (some of them with a portion of detailed structures on the surface), it is hard to evaluate the accuracy of surface detail recovery, which is unique to photometric stereo over other 3D reconstruction techniques. On the other hand, near-planar surfaces usually accompanied with rich details are commonly observed in our daily life, such as reliefs, badges, and coins. However, existing photometric stereo datasets mainly choose statue-like objects or other bulgy shapes with a certain height as targets, lacking sophisticated evaluation on near-planar surfaces.

In this paper, we build a new dataset named **DiLiGenT-Π**[1] for photometric stereo focusing on the recovery of near-planar shape and surface details. As shown in Fig. 1, we collect 30 representative real-world planar objects with rapidly varied geometric details. The dataset can be categorized into 4 groups containing **metallic**, **specular**, **rough**, and **translucent** surface reflectance, respectively. The metallic group contains 10 metallic coins; the specular group includes 10 enamel badges; the translucent group has 5 3D-printed objects made by photo-polymer resin, and the rough group contains 5 surfaces sharing the same geometry of the translucent group but sprayed with a matte paint [35]. In addition, our DiLiGenT-Π takes an optical profilometer to capture the ultra-precise surface 3D structure in nanometer

† Equally contributed authors ‡ Corresponding authors.

[1] 'DiLiGenT' [41] as the abbreviation of Directional Lighting, General reflectance, with the 'ground Truth' shapes for photometric stereo benchmarking. As we take the similar assumptions, we refer DiLiGenT as prefix and use Π to indicate planar objects.

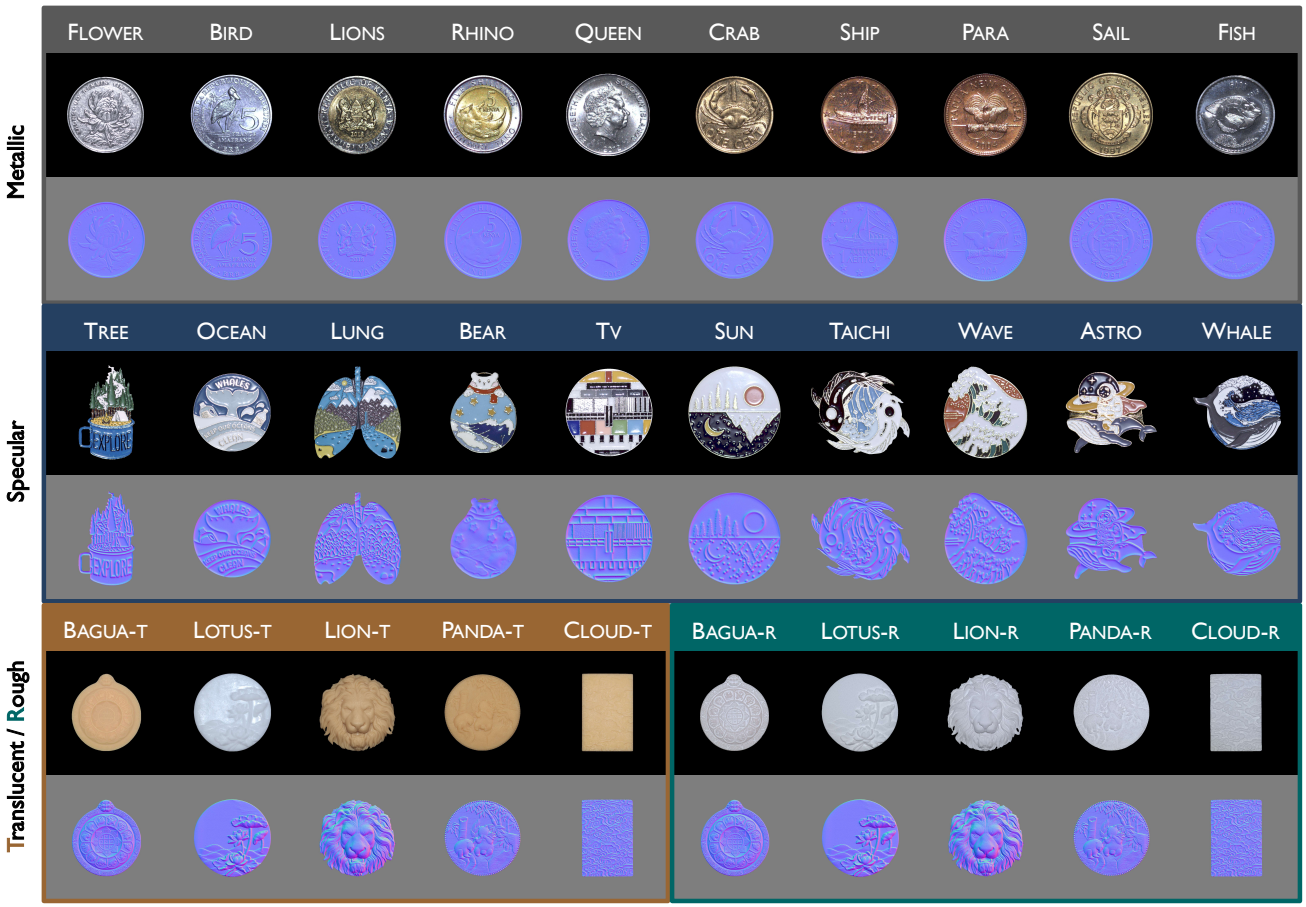


Figure 1: Overview of DiLiGenT-II. We collect 4 groups of near-planar objects with rich surface details and diverse reflectance types (metallic, specular, rough, and translucent). The corresponding 'ground-truth' surface normals shown in the even rows are measured via a precise profilometer in the accuracy of nanometers.

accuracy, providing the 'ground truth' surface normal with well-preserved tiny surface details.

We apply DiLiGenT-II to evaluate up-to-date photometric stereo methods under the settings of calibrated and uncalibrated distant light, and benchmark the reconstruction performance on detailed structures and near-planar surfaces. The evaluation results reveal the difference in surface detail recovery of learning-based photometric stereo methods working with per-pixel and all-pixel manners [54]; the challenging problems of uncalibrated photometric stereo for near-planar surfaces; and the influence of translucent and rough reflectance on surface detail recovery. The analysis on the DiLiGenT-II presents new challenges and open problems for photometric stereo.

To summarize, this paper contributes to photometric stereo benchmark and inspires future research by proposing:

- the first real-world dataset, DiLiGenT-II, that evaluates near-planar surfaces with rich geometric details;
- up-to-date benchmark evaluation of photometric stereo recovering important features for handling surface details; while
- revealing inherent obstacles of planar detailed objects to photometric stereo with open problems.

2. Related Work

This paper focuses on the benchmark dataset for evaluating photometric stereo on near-planar surfaces with rich details. In the following sections, we will discuss the related datasets and representative works in photometric stereo.

2.1. Photometric Stereo Datasets

Synthetic datasets adopt physical-based rendering engines such as Mitsuba [22] and Blender [11] to create image observations and the corresponding surface normal maps of

Table 1: Summary of real-word photometric stereo datasets. Material: controlled (C: fabricated or carefully selected with controlled categories) or uncontrolled (UC: randomly picked up from daily objects); Ground Truth (GT measure): from CAD/Scanned models with registration (+Reg) or from photometric stereo (PS). Number (#) of shapes, lights, and sets (one set means a sequence of photometric stereo images under varying lighting conditions used for computation).

Dataset							
	DiLiGenT-Π	DiLiGenT10^2 [39]	DiLiGenT [41]	LUCES [34]	Harvard [51]	ETHz [27]	Gourd&Apple [4]
GT measure	Scan+Reg	CAD+Reg	Scan+Reg	Scan+Reg	PS	Scan+Reg	PS
Material	C	C	UC	UC	UC	UC	UC
# Shapes	30	10	10	14	7	3	2
# Lights	100	100	96	52	20	260	102/112
# Sets	30	100	10	14	7	3	2

diverse synthetic scenes. DPSN [40] introduces the first synthetic photometric stereo dataset BlobbyPS, comprising 10 smooth Blobby [23] shapes with reflectance assigned by measured MERL BRDFs [33] rendered under 96 light directions. To extend the smooth surface of Blobby to more complex shapes, PS-FCN [8] takes $59,292$ scanned sculptures to render the SculpturePS dataset. CNN-PS [18] proposes the CyclePS dataset to extend material distribution from uniform to spatially-varying, where each sub-region or even each pixel of the surface is assigned with distinct reflectances modeled by Disney Principle BSDF [5]. While synthetic datasets expand photometric stereo data availability, existing synthetic datasets are mostly rendered with statue-like objects such as the shapes in Blobby [23] and Sculpture dataset [6], leading to a shape domain gap to flatten surfaces. In this paper, we render a synthetic dataset containing 127 near-planar surfaces with rich details to enhance learning-based photometric stereo on near-planar surface recovery.

Real-world datasets complement the gap between computer graphics rendering and real-world imaging process. The Gourd&Apple dataset [2] releases image observations of 2 objects with spatially-varying isotropic BRDFs. Harvard dataset [50] contains 7 surfaces with uniform diffuse reflectance. However, the ground truth surface normal of the above two datasets is not provided. DiLiGenT [41] records 10 objects with diverse shapes and general non-Lambertian materials. Starting from DiLiGenT [41], benchmark evaluation of photometric stereo becomes available based on the 'ground truth' surface normal from scanned meshes. The following-up datasets further extends DiLiGenT [41] from the perspective of multi-views (DiLiGenT-MV [30]), controlled materials and shapes (DiLiGenT 10^2 [39]), near-field illumination (LUCES [34]), environment illumination [1, 16, 17], and global illumination effects [27].

As summarized in Tab. 1, objects contained in existing real-world photometric stereo datasets are mostly bulgy and smooth, which do not include flattened or near-planar surfaces though they are commonly seen in our daily life. More importantly, the smooth target shapes are not suitable to evaluate the uniqueness of photometric stereo over other 3D reconstruction techniques, that is, the high-fidelity surface detail recovery, especially for delicate structures. To address these two problems, we newly build a real-world photometric stereo dataset containing near-planar surfaces with rich geometric details.

2.2. Photometric Stereo Methods

We briefly review existing photometric stereo methods based on distant calibrated and uncalibrated light settings. Please refer to the photometric stereo survey of non-learning based methods [41] and learning-based methods [54, 25] for more comprehensive analysis.

Calibrated photometric stereo works with calibrated distant lights. Existing non-learning based methods either treat specular highlights and shadows as sparse outliers [48, 38] or propose parametric or data-based reflectance model to explicitly handle the non-Lambertian reflection [10, 14, 42, 15]. Beginning from DPSN [40], the image observations are directly mapped to the corresponding surface normal via deep neural networks, where the non-Lambertian reflectance is implicitly learned from the synthetic training data. The network structure in existing learning-based methods can be divided into the all-pixel branch (representative work: PS-FCN [8]) and the per-pixel branch (representative work: CNN-PS [18]). Based on these two typical network structures, following-up works further promote photometric stereo by addressing the global illumination effects (*e.g.* PX-Net [32]), reducing the number of inputs (*e.g.* SPLINE-Net [53], LMPS [28]), combining the

merit of per-pixel and all-pixel methods (*e.g.* GPS-Net [52], PS-Transformer [19]). However, few methods focus on the recovery of surface details despite its importance in photometric stereo. Ju *et al.* [24] firstly addressed the high-frequency surface details in photometric stereo based on attention-weighted loss. Their following-up work [26] further enhanced the detail recovery accuracy via a double-gate normalization and a parallel high-resolution structure.

Uncalibrated photometric stereo works under unknown distant light directions, so that photometric stereo becomes more challenging even with Lamebrain reflectance assumption. Non-learning based methods adopt the local diffuse reflectance (LDR) maxima [37], perspective camera [36], specularities [13], albedo entropy [3] to resolve the surface normal estimation ambiguity in uncalibrated Lambertian photometric stereo. Beginning from SDPS-Net [7], learning-based uncalibrated photometric stereo methods capable of handling non-Lambertian reflectance are proposed. SDPS-Net [7] first estimates the light directions and intensities from input image observations, and then feeds them into the normal estimation network. Kaya *et al.* [27] proposed an uncalibrated deep neural network with an inverse rendering module, where the inter-reflections are explicitly modeled in the image formation process. Following the analysis of Chen *et al.* [9], the light calibration in deep uncalibrated photometric stereo is related to the existence of attached shadows and specular highlights in the image observations. For near-planar surfaces where surfaces are flattened and attached shadows are rarely observed, whether the existing methods can be applied is not evaluated. The above uncalibrated photometric stereo methods assume distant light. Existing methods further extend the illumination setting to general natural light. Lichy *et al.* [31] proposes a weakly calibrated method working in the indoor environment, requiring at most 6 images from approximately known lighting directions. Ikehata proposes photometric stereo methods [20, 19] to handle general unknown illumination in real-world scenes.

3. DiLiGenT-Π Dataset

This section introduces our DiLiGenT-Π dataset, which contains 30 near-planar scenes covered by varying materials and geometric details. Each real-world scene contains RGB images under 100 varying light directions and a precisely measured 'ground-truth' normal map. The resolution for each scene is 960×960.

3.1. Objects groups

As shown in Fig. 1, we collect 4 groups of near-planar objects named by their reflectance properties: **metallic**, **specular**, **rough**, and **translucent**. In each group, we select target objects with rich geometric details with size around $15 \sim 25$ mm.

Metallic group. We choose 10 different coins as target objects commonly observed in our daily life. These coins are casted from different metallic materials, such as nickel, brass, aluminum alloy, and bi-metal. The reflectance distributions are either uniform (*e.g.* CRAB) or spatially-varying (*e.g.* RHINO). The surface normal of the coins contains rich and delicate geometric details and even traces of daily use. Their corresponding surface PV (peak to valley) depth value is usually less than 1 mm.

Specular group. We choose 10 sets of badges as the target objects. These badges are made of polished metal, plastic, and enamel, showing strong and sparse specular spikes (*e.g.* BEAR), or broad and soft specular lobes (*e.g.* TREE) on their captured images. All objects in this category contain spatially-varying reflectances and have greater depth variation (PV is around 1.5 mm), making this group different from the metallic group.

Translucent group. We collect 5 sets of 3D-printed relief surfaces to build the translucent object group. The materials for 3D printing for the LOTUS-T and the BAGUA-T are photo-polymer resin[2] with different colors, respectively, which is known to be slightly translucent, bringing unavoidable subsurface scattering in the surface reflectance. The surface shapes in this group have even stronger surface undulating (PV is around 4 mm), leading to stronger shadows and inter-reflections being observed in the captured images.

Rough group. We fabricate another 5 sets of 3D-printed relief surfaces sharing exactly the same shapes with the translucent group but covered by gray matte spray[3], whose reflectance is approximately close to the Oren-Nayar diffuse model [35] and the Torrance-Sparrow specular model [46], as analyzed in [45].

3.2. Capture System and Light Calibration

As shown in Fig. 2, we design and build a photometric stereo image capture setup with the function of automatic illumination and capture at varying light directions. The capture system is placed into a darkroom cage covered with black felt to shield the ambient lights.

On the illumination side, we build a dual-axis rotation platform to control the light direction, which is based on robot arms capable of omni-direction illuminations on the upper hemisphere of the target surface. We attach a single LED light source on our robot arm to ensure the consistency of emitted light intensity among different images. We further adopt a co-concentric rotation design to achieve

[2]Kexcelled Ultradetail: https://www.kexcelled3d.com/products/ultrdetial/. Retrieved August 16, 2023.

[3]FA-5: http://cysygroup.com/en/product.asp?category=NDT&page=5. Retrieved August 16, 2023.

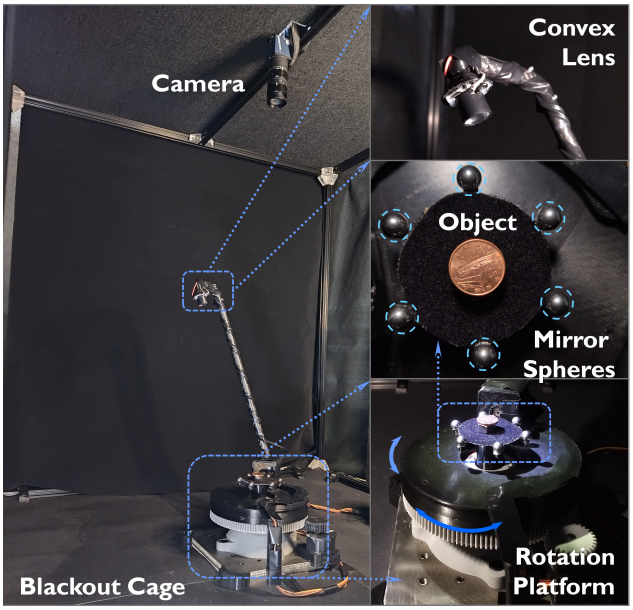


Figure 2: Capture system overview. Our automatic photometric stereo image capture equipment is placed in the blackout cage, containing two rotation axes to provide illumination from an arbitrary direction from the upper hemisphere. A lens is placed in front of the LED to improve the directionality and uniformity. Six mirror balls around the object are used to calibrate light directions 'on the fly'.

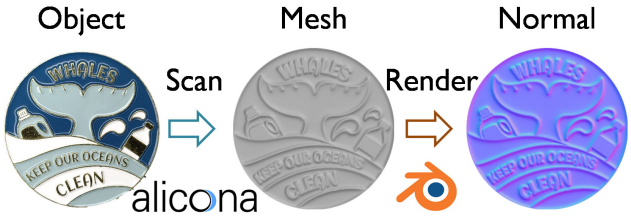


Figure 3: Process of obtaining a 'GT' normal map.

roughly the same distance between light and the target surface so that the received light intensities at different surface positions are roughly the same. To improve the directivity of the LED point light source, A convex lens is placed in front of the LED source.

On the camera side, we use a Daheng MER-503-36U3C camera[4] with a 50 mm lens to record 12-bit raw images (with linear radiometric response). Under each directional light, we capture 10 images with the exposure time settings from 1 ms to 10 ms, from which a high dynamic range (HDR) image that records both dark shadows and bright specular highlights can be reconstructed based on the existing HDR algorithm [12].

During the capture, the object is placed in the center of the rotation platform. The system controls the robot arm to move toward a pre-defined light direction and triggers the camera to shoot at various exposure times to obtain an HDR image measurement. Then, the robot arm is rotated to the next light direction. Totally, 100 HDR images under varying lights are recorded by repeating the above process.

To calibrate the light directions, we place 6 mirror balls with a radius of 8 mm around the object stage, and record the specular highlights at each mirror ball during data capture. The light directions can then be calculated based on the specular positions following the same calibration method described in existing methods [41, 39]. Please check our supplementary material for more details.

3.3. Obtaining the 'GT' Normal Map

Similar to previous real-world datasets: DiLiGenT [41] and LUCES [34], we measure the 'ground truth' surface normal from scanned mesh in DiLiGenT-Π. As shown in Fig. 3, we apply a commercial 3D scanner Bruker Alicona Infinity Focus (up to **10 nm** accuracy)[5] to measure a precise point cloud. The scanner is based on Focus-Variation measurement and can probe vertical surfaces precisely, which is necessary to scan near-planar objects completely.

Given a scanned mesh of the near-planar surface, we manually adjust the camera pose to align the mesh with one captured image by matching key points and geometric features. Then, taking the calibrated intrinsic camera parameters and the extrinsic pose, we render the mesh to the corresponding surface normal map by Blender [11] with the same resolution to the captured images. We try our best to check key points accuracy at the sub-pixel level, but inevitable errors in the manual alignment process might still exist, so we add a quotation on the 'ground truth' like [41, 39].

4. Benchmark Analysis

This section showcases the benchmark results for photometric stereo techniques using the DiLiGenT-Π dataset.

4.1. Baseline Methods & Evaluation Metric

Based on the survey [41] adopt non-Learning photometric stereo, we choose the baseline method (least-square based Lambertian photometric stereo [47], LSPS), baseline method with position thresholding strtegy [41] (TH28 and TH46 reject pixels whose intensities under varying lights are outside the range of $[20\%, 80\%]$ and $[40\%, 60\%]$), respectively), WG10 [49] (a robust photometric stereo method based on outlier rejection), ST14 [42] (showing

[4]Camera website: https://www.get-cameras.com/USB-Camera-Sony-IMX264-MER-503-36U3C. Retrieved August 16, 2023.

[5]Alicona website: https://www.alicona.com/en/products/infinitefocus. Retrieved August 16, 2023.

Table 2: Benchmark results on our real-world dataset DiLiGenT-II. Mean angular error in degree of each object on various methods are presented, and the average angular error for each material group and for all objects is shown in the bottom rows. We denote 'NA-PSN' as the abbreviation of NormAttention-PSN [26].

		Calibrated Methods										Uncalibrated Methods				
		Plane	LSPS [47]	TH28 [41]	TH46 [41]	WG10 [49]	ST14 [42]	PX-Net [32]	CNN-PS [18]	PS-FCN [8]	GPS-Net [52]	NA-PSN [26]	PF14 [37]	UPS-FCN [8]	SDPS-Net [7]	LW21 [31]	SDM-UniPS [20]
Metallic	Flower	5.8	6.8	7.5	6	7.2	7.9	5.5	4.7	4.6	4.6	4.6	35.3	12.7	12.8	23.9	15.2
	Bird	9	8.1	9.1	8	11.6	10.8	8.8	6.8	7.2	6.8	6.9	34.9	15.4	17.6	28.5	26.5
	Rhino	6.8	7.5	7.8	8.9	9.3	8.9	10.6	4.9	5.3	5.6	5.1	30.2	17.7	24.9	26.8	17
	Lions	6.6	6.7	7.7	7.4	9.1	8.1	6.8	4.7	4.5	4.6	4.4	40.9	11.8	19.6	27.1	9.2
	Queen	6.2	7.5	8.1	8.9	8.6	8.3	14.7	5.4	4.7	4.7	4.9	46.4	12.6	16.5	27.2	10.6
	Crab	5.9	7.1	7.6	8.6	8.5	8.4	7.2	4.5	5.3	4.9	4.4	31.7	18.1	20.5	26.8	25.4
	Ship	5.9	7.1	8.9	6.4	9.3	9.1	12.3	4.9	6.1	5.1	4.9	89.7	15.6	19	26.4	22
	Para	6.7	6	6.6	5.1	8.4	8	4.9	3.9	4.7	4	4.2	42.7	17	19.8	24.6	23.2
	Sail	6.9	8.7	10.2	8.8	8.1	8.2	13.4	5.2	5.1	5.5	5.2	40.6	13.5	16.7	27	10.5
	Fish	5.2	6.7	7.9	6.5	8.2	8.1	7.4	4.2	4.6	4.6	4.4	39.9	15.8	23.6	25.6	24.5
Specular	Tree	20.1	9.4	11	9.7	14.9	10.6	11.3	7.8	10.2	9.3	9.7	32	40.6	34.1	35.8	47.2
	Ocean	14.5	6.2	6	5.9	8.4	7.6	5.9	4.6	5.8	5.4	5.8	77.9	26.7	31.4	28.1	34.6
	Lung	20.3	8.2	8.8	8.1	11.8	9.4	7.6	5.7	9.7	7.5	7.8	41.5	36.5	40.2	34.4	46.6
	Bear	9.8	9.1	8.5	8.9	8.6	9.7	7.9	7.4	7.4	6.9	7.4	84.8	27.6	30.7	27	23.8
	TV	19.6	11.7	13.3	13.2	17.1	13.5	12.1	11.3	10.6	10.6	9.6	73	22.8	41.1	35.4	34.4
	Sun	11.9	7.9	9.5	11.7	10.5	8.2	6.7	5.8	6.7	8	5.6	27.9	22.8	31.5	24	26.2
	Taichi	17.3	9.2	11.2	11	17.9	10.7	8.7	8.3	8	8.3	7.6	58.9	29.8	26.9	28.9	36.6
	Wave	15.7	6.9	7.7	7.3	9.2	8.4	7.1	5.3	6.8	6.3	6.1	76	30.8	39.1	25.9	34.9
	Astro	17.6	7.9	8.7	7.9	11.4	9.2	7.1	6	7.2	7.7	7.2	35.2	25.9	37.7	36.5	37.8
	Whale	16.3	8.8	8.9	8.5	13.4	10.1	9.5	11.6	12.2	10.4	8.7	63.2	37.6	29.8	32	33.8
Translucent	Bagua-T	20.6	16.7	16.5	16.1	17.8	17.4	15.8	16.4	16.8	16.1	16.5	20.3	19.9	28.9	27.8	17.1
	Lotus-T	15.9	14	14.1	13.8	14.3	14.5	13.7	13.5	13.6	13.7	13.7	19.8	19.5	26.5	21.3	13.6
	Lion-T	30.1	21	20.5	19.9	21.6	21.3	18.4	20.3	21.2	23.4	21	22.7	25.8	23.6	30.5	16.2
	Panda-T	18.6	16.4	16.4	16.3	16.8	16.8	15.9	16.6	17.2	17	16.6	19.2	18.5	23.7	22.8	17.6
	Cloud-T	18.7	17	17	16.8	18.1	17.7	16.1	17.2	17.8	17.6	17.3	20.6	20.7	27.5	24.2	19.2
Rough	Bagua-R	20.4	13.2	12.2	11.6	16.5	13.6	11.9	12.2	13	16.4	12	19.5	16.8	22.5	25	14.6
	Lotus-R	15.7	12	11.7	11.3	12.3	12.4	11	10.9	11.8	13.4	10.8	18.3	14.7	21.7	23.8	11.8
	Lion-R	30.2	19.4	17.8	17	19.4	19.1	17.9	15.8	18.4	23	16.4	19.5	25.2	20.8	29.7	15.9
	Panda-R	18.7	14.6	14.1	14	16	15	14.1	14.2	14.8	16.3	13.9	17.4	16.8	21.8	21.6	17.1
	Cloud-R	18.4	14	13.6	13.5	17.4	14.7	13.6	14.6	14.3	15.9	13.4	18.4	16.8	27.4	21.7	17.1
	M.Avg.	6.5	7.2	8.1	7.5	8.8	8.6	9.2	4.9	5.2	5.1	4.9	43.2	15	19.1	26.4	18.4
	S.Avg.	16.3	8.5	9.4	9.2	12.3	9.7	8.4	7.4	8.5	8	7.5	57	30.1	34.3	30.8	35.6
	T.Avg.	20.8	17	16.9	16.6	17.7	17.5	16	16.8	17.3	17.6	17	20.5	20.9	26.1	25.3	16.8
	R.Avg.	20.7	14.6	13.9	13.5	16.3	15	13.7	13.5	14.4	17	13.3	18.6	18.1	22.8	24.4	15.3
	Avg.	14.5	10.5	11	10.6	12.7	11.5	10.8	9.2	9.8	10.1	9.2	39.9	21.5	25.9	27.3	23.3

best performance at DiLiGenT reported in [41]). Based on surveys about the learning-based photometric stereo [54, 25], we choose representative networks handling photometric stereo in an all-pixel manner (NormAttention-PSN [26], PS-FCN [8]), per-pixel manner (CNN-PS [18], PX-Net [32]), and the method GPS-Net [52] that considers both the per-pixel and the all-pixel structures. Besides the above calibrated photometric stereo methods, we also evaluate existing uncalibrated photometric stereo approaches, including a non-learning based method PF14 [37] (showing the best performance under uncalibrated setting as reported in [41]), three representative learning-based methods SDPS-Net [7], UPS-FCN [8] and UPS-GCNet [9] published in recent years. We also include photometric stereo methods such as LW21 [31] and SDM-UniPS [20] that are based on more general natural illumination setup. During the evalu-

ation, we adopted the code and pre-trained model released by the authors to process the collected data in DiLiGenT-Π.

Similar to DiLiGenT [41] and DiLiGenT10^2 [39], the mean angular error (MAngE) between the estimated and the ground-truth surface normal is used as the metric for measuring the performance of photometric stereo methods quantitatively.

To clarify any concerns regarding the benchmark results, it's important to note that the input format varies between optimization-based methods (LSPS [47], TH28, TH46, WG10 [49], ST14 [42], and PF14 [37]) and other learning-based methods. Optimization-based methods employ a float format, providing higher precision, while other learning-based methods utilize quantized 16-bit input. This differentiation arises because the optimization-based methods are not sensitive to changes in the pixel value range resulting from quantization. For the learning-based methods, adhering to the standard procedure and employing quantized images as input ensures the reliability of the evaluation results.

4.2. Analysis to Different Baselines

As shown in Table 2, we evaluate the surface normal estimation accuracy for all methods on DiLiGenT-Π. In the following, we first analyze the surface detail recovery from calibrated photometric stereo, followed by the analysis of uncalibrated photometric stereo on near-planar surfaces.

Calibrated photometric stero. From Table 2, we observe that the LSPS [47] shows the best performance over other non-learning-based photometric stereo methods, while the angular error difference between LSPS [47], TH28, TH46, and ST14 [42] are marginal. For surfaces with shadows such as LION_R, or containing dominant Lambertian reflectances and sparse specular highlights such as FLOWER, TH46 is more effective than the LSPS [47]. However, TH46 could be unstable as only 20% of the image observations under varying lights are used for computing surface normals, especially for surfaces with dark reflectances (*e.g.*SUN and TV). These observations are consistent with the previous evaluation on DiLiGenT10^2 [39].

Among learning-based calibrated photometric stereo, CNN-PS [18] and NormAttention-PSN [26] achieve smaller mean angular errors, showing their advantages on detailed surface recovery. As shown in Fig. 4, we visualize the error distributions of PS-FCN [8], CNN-PS [21], and NormAttention-PSN [26] on four representative surfaces belong to the four reflectance groups. PS-FCN [7] outputs blurry normal estimation results as highlighted in Fig. 4, possibly due to the spatial smoothness brought by the fully convolutional networks. As CNN-PS [18] solves photometric stereo in a per-pixel manner, the surface details are not contaminated by the neighboring pixels. NormAttention-PSN [26] is built upon PS-FCN [8] but can handle surface detail recovery by an attention-weighted loss. We summarize the benchmark results of calibrated photometric stereo using DiLiGenT-Π as the following observation:

Observation 1 *Learning-based photometric stereo methods in the per-pixel branch are generally more effective than the all-pixel branch for handling surface detail estimation. A detail-weighted loss can help methods in the all-pixel branch for better recovering tiny structures.*

Uncalibrated photometric stero. From Table 2, uncalibrated photometric stereo methods generally show significant errors on near-planar surfaces in DiLiGenT-Π compared to the case of calibrated photometric stereo. For the non-learning-based method PF14 [37], the near-planar surface could be an ill-posed shape when conducting SVD for obtaining pseudo-surface normals and lights. Also, for learning-based uncalibrated photometric stereo, attached shadows and shading variations are essential in recovering the unknown light directions, as stated in [9]. However, the shadows are much less observable for near-planar surfaces as the tiny detail can barely cast a block of shadows from varying light directions compared with the case of the bulgy objects, making the light estimation more challenging.

Fine-tuning on the synthetic near-planar dataset. To check whether the error of learning-based uncalibrated photometric stereo can be further reduced by learning the data prior, we create a near-planar synthetic dataset PS_RELIEF for finetuning the uncalibrated photometric stereo. As shown in Fig. 5, our synthetic dataset contains 127 near-planar surface normals extracted from CAD meshes. We adopt Disney's principled BSDF [5] as the reflectance model and randomly generate BRDFs by adjusting the parameter to control the diffuse, specular, and metallic reflectance components in a similar manner to the existing synthetic dataset CyclePS [18]. In total, PS_RELIEF contains 3429 scenes. For each scene, we render the object by Blender [11] under 100 distant light directions with the same uniform lighting distribution of DiLiGenT10^2 [39].

Table 3: Ablation study on photometric stereo methods trained with or without fine-tuning (FT.) on the near-planar synthetic dataset PS_RELIEF.

Light	Method	DiLiGenT [41]		DiLiGenT-II	
		w/ FT.	w/o FT.	w/ FT.	w/o FT.
Uncalibrated	SDPS-Net [7]	17.80	9.51	16.03	27.76
	UPS-Net [8]	25.05	15.37	14.99	21.54
Calibrated	PS-FCN [8]	11.75	9.12	8.80	9.84

As shown in Table 3, we fine-tune the UPS-FCN [8] and SDPS-Net [7] with PS_RELIEF and test it on real-world dataset DiLiGenT [41] and DiLiGenT-Π, which contain bulgy shapes and near-planar shapes, respectively. The

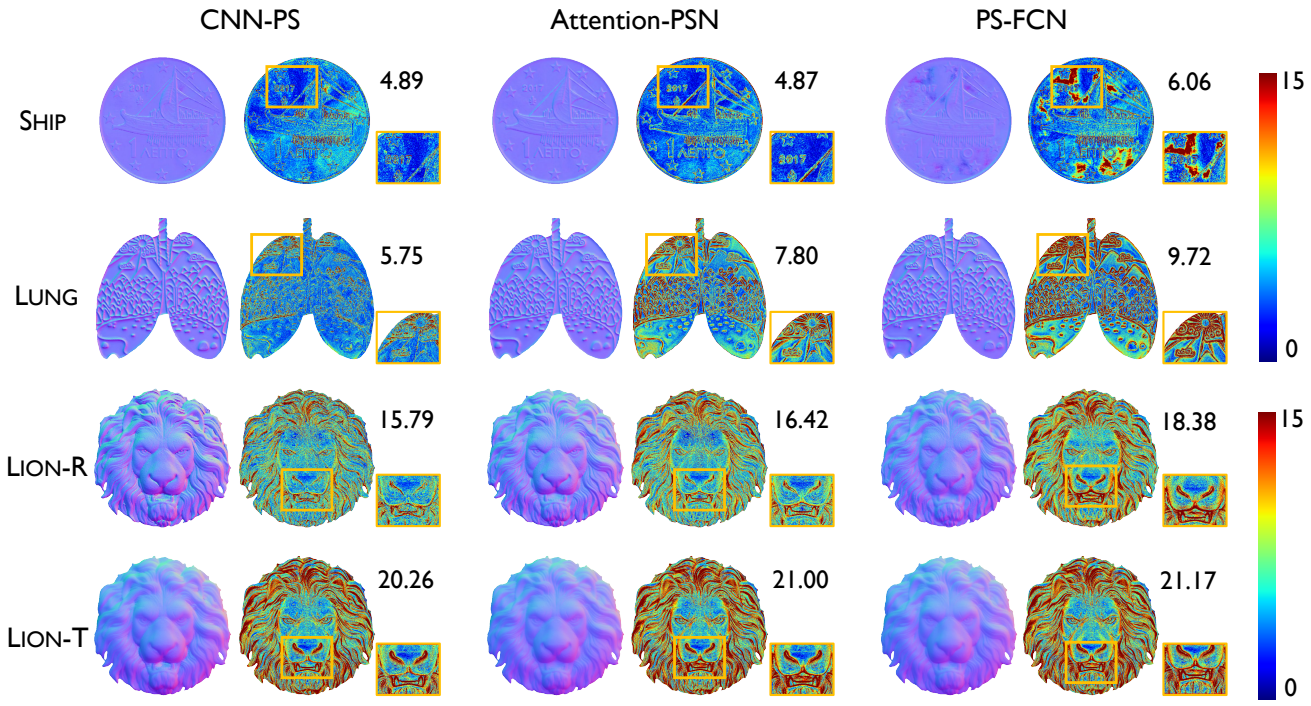


Figure 4: Visualization of surface detail recovery from different photometric stereo methods, where the odd and even columns plot the estimated surface normals and the corresponding angular error maps, respectively. Per-pixel based photometric stereo method CNN-PS [18] is more effective on detail recovery compared to all-pixel based method PS-FCN [8], as highlighted in the yellow boxes (Observation 1).

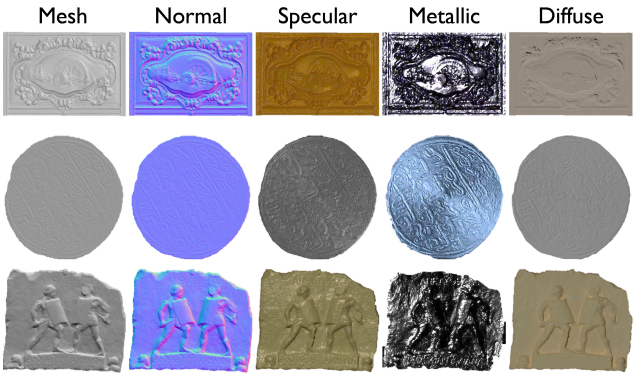


Figure 5: Overview of our synthetic dataset PS_RELIEF for fine-tuning photometric stereo on near-planar surfaces.

mean angular errors on DiLiGenT-II are significantly reduced after the fine-tuning on our near-planar dataset. However, we also observe performance degradation on DiLiGenT [41]. As shown in Fig. 6, the estimated surface normal for the BIRD in DiLiGenT-II is more reasonable after the fine-tuning, but with the consequence that the recovered CAT shape in DiLiGenT becomes more flattened. This behavior is not observed in calibrated photometric stereo methods such as PS-FCN, where the fine-tuning of PS-FCN on PS_RELIEF does not influence the estimation on DiLiGenT [41] dramatically. We summarize the benchmark results of uncalibrated photometric stereo using DiLiGenT-II and fine-tuning existing methods using our synthetic PS_RELIEF dataset as the following observation:

Observation 2 *Near-planar surfaces are challenging for photometric stereo under uncalibrated light settings, and the normal estimation of learning-based uncalibrated photometric stereo is sensitive to the shape distributions present in the training dataset.*

4.3. Analysis to Different Reflectance Groups

As shown in Table 2, we find the four groups in DiLiGenT-II sorted by the MAngE on different photometric stereo methods, arranged from high to low, are translucent, rough, specular, and metallic. The estimation errors come from non-Lambertian reflectance and surface geometry determining the shadows and inter-reflections. On the surface geometry side, we present the angular difference of a pure planar surface normal and the ground-truth surface normal in Table 2 (first column), showing that the shape variations of the translucent and the rough groups are greater than that of the metallic and specular groups. Also, surfaces in rough

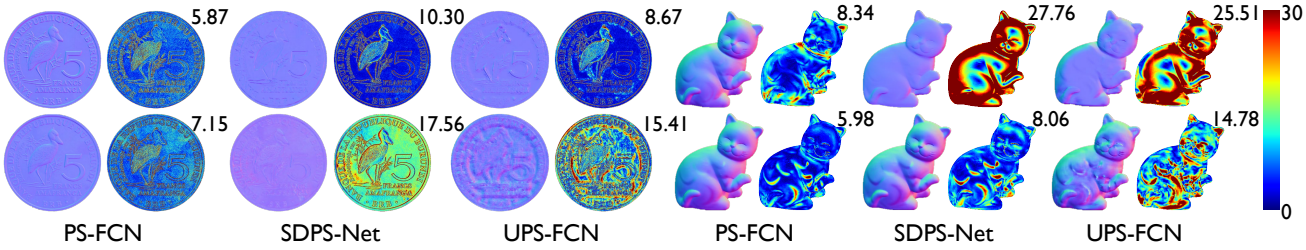


Figure 6: Surface normal estimates of photometric stereo methods with (top row) or without (bottom row) finetuning on PS_RELIEF. Compared to calibrated photometric stereo PS-FCN [8], learning-based uncalibrated photometric stereo (*e.g.* UPS-FCN [8], SDPS-Net [7]) are heavily influenced by the shape prior learned from the training dataset (Observation 2).

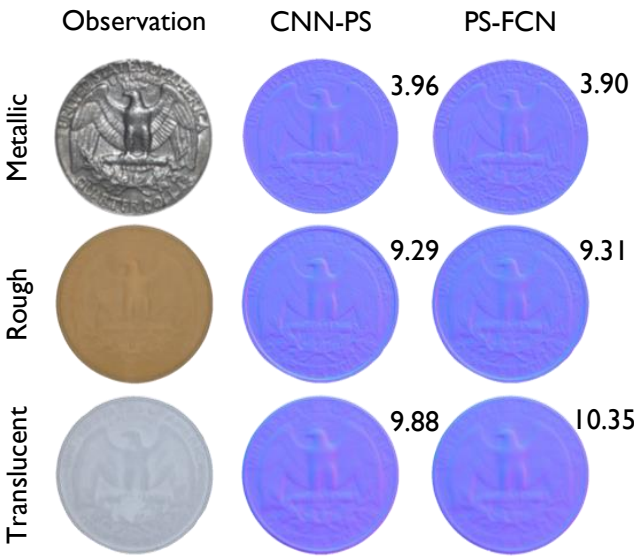


Figure 7: Ablation analysis to the influence of reflectance based on the same shape. The columns, from left to right, display the image observations, the estimated normal maps from CNN-PS [18], and from PS-FCN [8], with MAngE labeled at the top-right corner. The rough and translucent materials are more challenging than the metallic materials for surface detail recovery (Observation 3).

and translucent groups contain more shadows and inter-reflections as their depth variation measured by PV is 4 mm compared to 1.5 mm, and 1 mm in the specular and metallic groups. On the reflectance side, the sub-surface scattering in the translucent group blurs the details of normal estimates as visualized in Fig. 4, resulting in greater estimation errors compared to the metallic and specular reflectance.

To disentangle the influence of reflectance and surface geometry on surface normal estimation, we conduct an ablation study on a metallic coin shown in Fig. 7. We scan this coin's 3D mesh and create another two objects by 3D printing using translucent and rough materials that are the same as those used in our DiLiGenT-II. Based on the same geometry, the surface normal estimation errors from CNN-PS [18] and PS-FCN [8] are higher on rough and translucent surfaces compared to metallic surfaces. Although the recovered details from the rough surface are sharper than the translucent one, the rough surface is still challenging due to its complex reflectance, and no previous photometric stereo method has targeted this kind of reflectance existing in objects covered by matte spray. We summarize these analysis results as our last observation:

Observation 3 *Near-planar surface normal estimation using photometric stereo methods remains a challenging task for translucent and rough surfaces, where surface details are significantly blurred due to the subsurface scattering in translucent surfaces.*

5. Conclusion

This paper builds a real-world photometric stereo dataset DiLiGenT-II focusing on near-planar surfaces with rich details, which are important to show the core strength of photometric stereo. We conduct benchmark evaluations on the dataset and draw three key observations. However, the evaluation metric utilized in this study is MAngE, which assigns equal weights to surface normals regardless of the spatial distribution of surface details. Therefore, it is desired to devise a new evaluation metric that can measure the performance of surface detail recovery. Overall, we hope that DiLiGenT-II and the key observations will offer useful insights to further photometric stereo methods for detailed recovery of near-planar surfaces.

Acknowledgement

This work is supported by the National Key Research and Development Program of China under Grant No. 2019YFA0706701, National Natural Science Foundation of China under Grant No. 62136001, 62088102, 52175477. Heng Guo was supported by JSPS KAKENHI (Grant No. JP23H05491). The authors also want to thank *openbayes.com* for providing part of the computing resource.

References

[1] Jens Ackermann, Fabian Langguth, Simon Fuhrmann, and Michael Goesele. Photometric stereo for outdoor webcams. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 3

[2] Neil Alldrin, Todd Zickler, and David 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] 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. 4

[4] 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

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

[6] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. ShapeNet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 3

[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. 4, 6, 7, 9

[8] 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, 6, 7, 8, 9

[9] 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. 4, 6, 7

[10] Lixiong Chen, Yinqiang Zheng, Boxin Shi, Art Subpa-Asa, and Imari Sato. A microfacet-based reflectance model for photometric stereo with highly specular surfaces. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017. 3

[11] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2021. 2, 5, 7

[12] Paul E Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 classes*. 2008. 5

[13] Ondřej Drbohlav and Radim Šára. Specularities reduce ambiguity of uncalibrated photometric stereo. In *Proc. of European Conference on Computer Vision (ECCV)*, 2002. 4

[14] Kenji Enomoto, Michael Waechter, Kiriakos N Kutulakos, and Yasuyuki Matsushita. Photometric stereo via discrete hypothesis-and-test search. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 3

[15] 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

[16] Bjoern Haefner, Songyou Peng, Alok Verma, Yvain Quéau, and Daniel Cremers. Photometric depth super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 3

[17] Yannick Hold-Geoffroy, Paulo Gotardo, and Jean-François Lalonde. Single day outdoor photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 3

[18] Satoshi Ikehata. CNN-PS: CNN-based photometric stereo for general non-convex surfaces. In *Proc. of European Conference on Computer Vision (ECCV)*, 2018. 3, 6, 7, 8, 9

[19] Satoshi Ikehata. PS-Transformer: Learning sparse photometric stereo network using self-attention mechanism. *Proc. of the British Machine Vision Conference (BMVC)*, 2022. 4

[20] Satoshi Ikehata. Scalable, detailed and mask-free universal photometric stereo. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 4, 6

[21] 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. 7

[22] Wenzel Jakob. Mitsuba renderer, 2010. 2

[23] 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

[24] Yakun Ju, Kin-Man Lam, Yang Chen, Lin Qi, and Junyu Dong. Pay attention to devils: A photometric stereo network for better details. In *Proc. of International Joint Conference on Artificial Intelligence*, 2021. 4

[25] Yakun Ju, Kin-Man Lam, Wuyuan Xie, Huiyu Zhou, Junyu Dong, and Boxin Shi. Deep learning methods for calibrated photometric stereo and beyond: A survey. *arXiv preprint arXiv:2212.08414*, 2022. 3, 6

[26] 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, 6, 7

[27] 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

[28] 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

[29] 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*, 2020. 1

[30] 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*, 2020. 3

[31] 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. 4, 6

[32] 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. 3, 6

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

[34] 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, 5

[35] Michael Oren and Shree K Nayar. Generalization of Lambert's reflectance model. In *Proc. of Annual conference on Computer Graphics and Interactive Techniques*, 1994. 1, 4

[36] Thoma Papadhimitri and Paolo Favaro. A new perspective on uncalibrated photometric stereo. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013. 4

[37] 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, 7

[38] Yvain Quéau, Tao Wu, François Lauze, Jean-Denis Durou, and Daniel Cremers. A non-convex variational approach to photometric stereo under inaccurate lighting. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 3

[39] 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, 5, 7

[40] 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

[41] 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, 5, 6, 7, 8

[42] 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

[43] Boxin Shi, Zhe Wu, Zhipeng Mo, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation for non-Lambertian and uncalibrated photometric stereo. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 1

[44] William M. Silver. *Determining shape and reflectance using multiple images*. PhD thesis, Massachusetts Institute of Technology, 1980. 1

[45] Bo Sun, Kalyan Sunkavalli, Ravi Ramamoorthi, Peter N Belhumeur, and Shree K Nayar. Time-varying brdfs. *IEEE Transactions on Visualization and Computer Graphics*, 2007. 4

[46] Kenneth E Torrance and Ephraim M Sparrow. Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America*, 1967. 4

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

[48] 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

[49] Tai-Pang Wu and Chi-Keung Tang. Photometric stereo via expectation maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010. 5, 6, 7

[50] 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

[51] 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

[52] 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, 6

[53] 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

[54] Qian Zheng, Boxin Shi, and Gang Pan. Summary study of data-driven photometric stereo methods. *Virtual Reality & Intelligent Hardware*, 2020. 2, 3, 6

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.

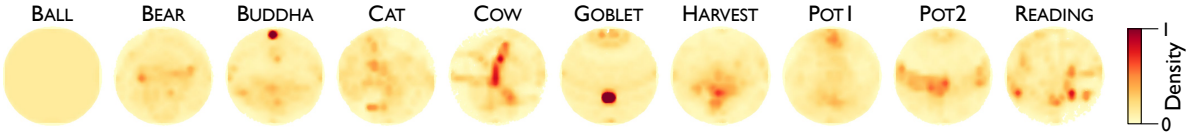


Figure S1: Surface normal distributions of objects in DiLiGenT [11].

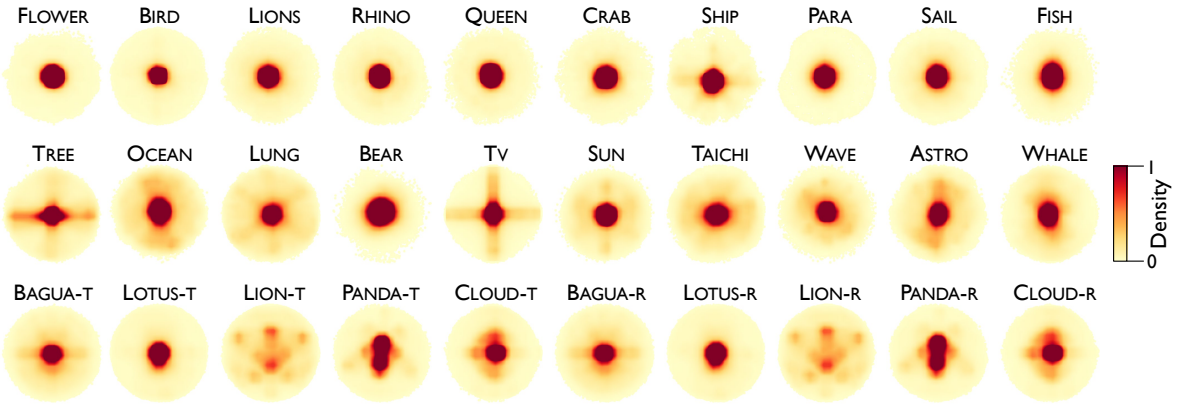


Figure S2: Surface normal distributions of objects in DiLiGenT-Π.

† 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.

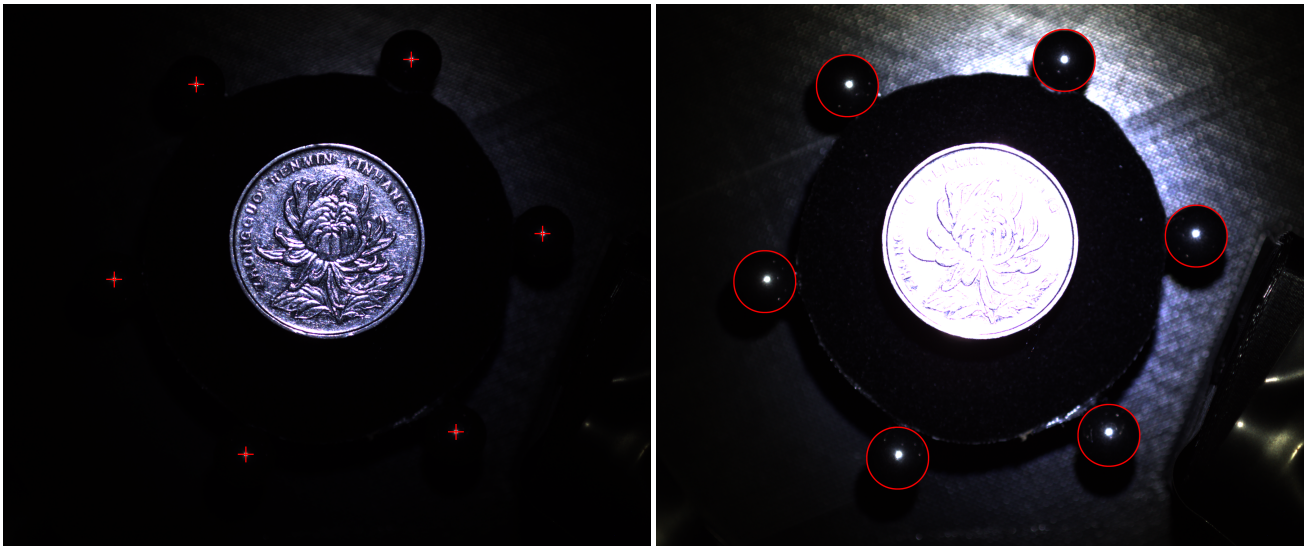


Figure S3: Images of FLOWER captured in low (left) and high (right) exposure time for detecting the specular spots and contour of the 6 mirror spheres for light calibration.

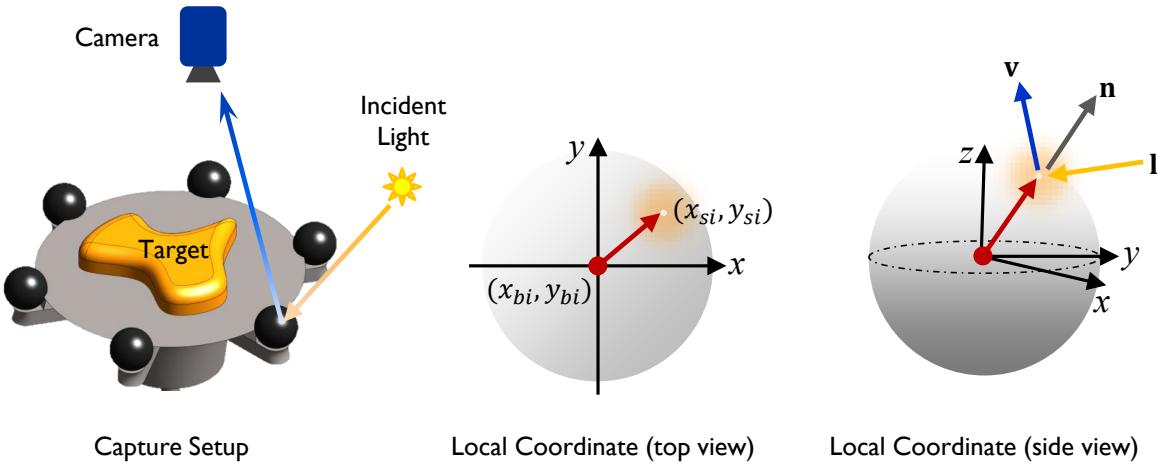


Figure S4: The geometrical relationship between the captured object and a mirror sphere in our light calibration process, showing the setup (left), the sphere in camera coordinate (middle), and the geometrical relationship on the sphere in 3D coordinate (right).

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:

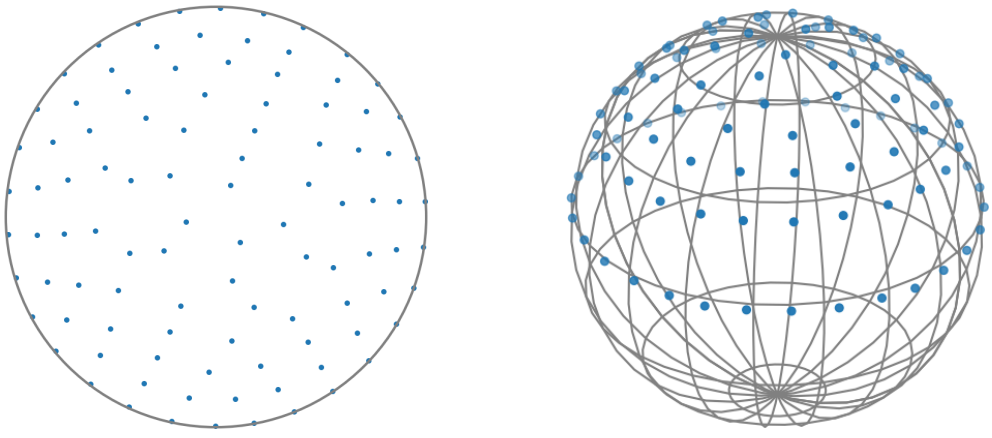
$$\begin{aligned} h_x &= \frac{\sum_{i=1}^{6}(x_{si} - x_{bi}) r_i}{\sum_{i=1}^{6} r_i^2}, \\ h_y &= \frac{\sum_{i=1}^{6}(y_{si} - y_{bi}) r_i}{\sum_{i=1}^{6} r_i^2}, \end{aligned} \tag{4}$$

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.

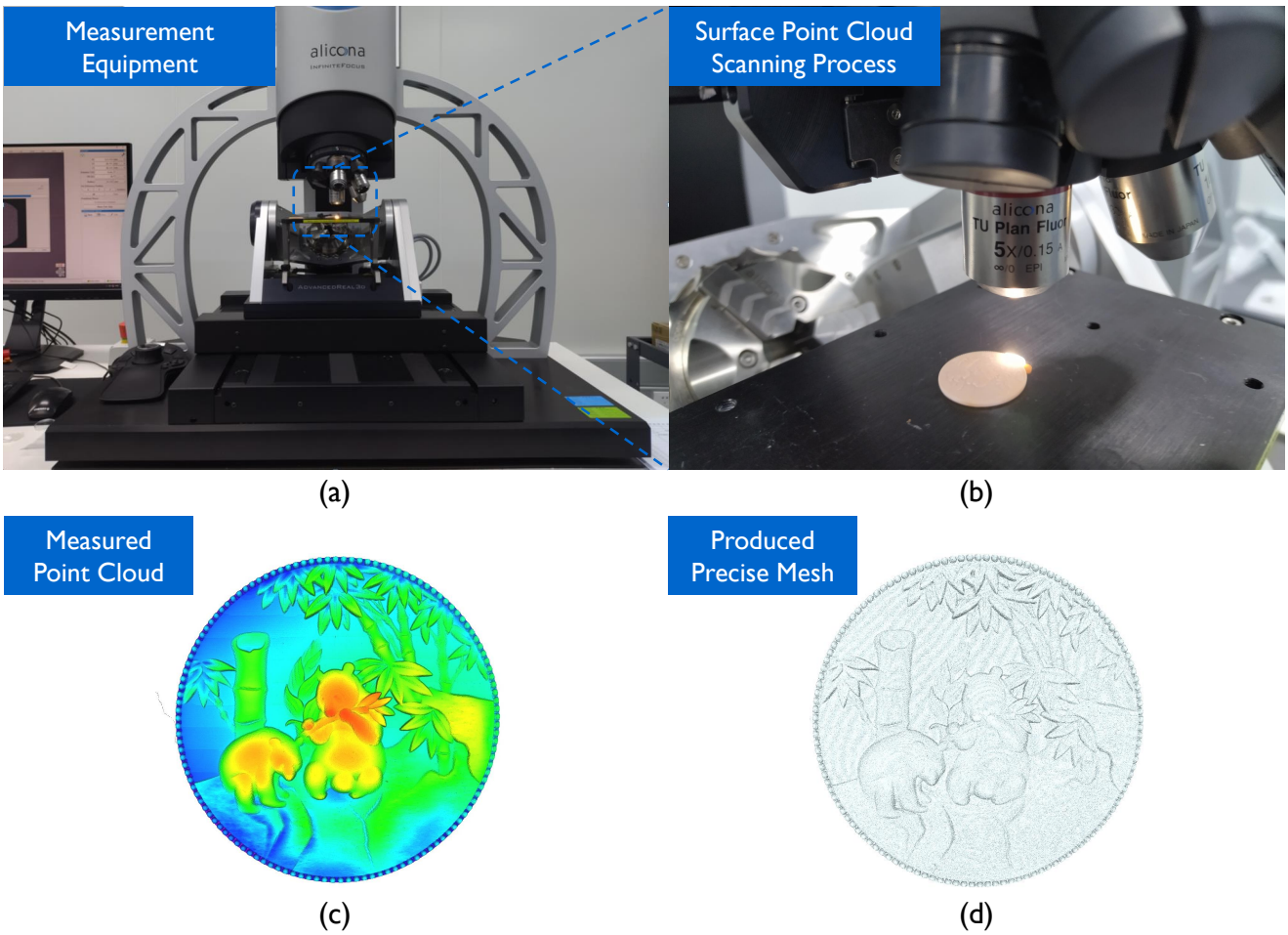


Figure S6: Surface measurement process of the proposed dataset. A precise mesh is obtained via Alicona Infinity Focuse to generate the 'GT' surface normal map.

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.

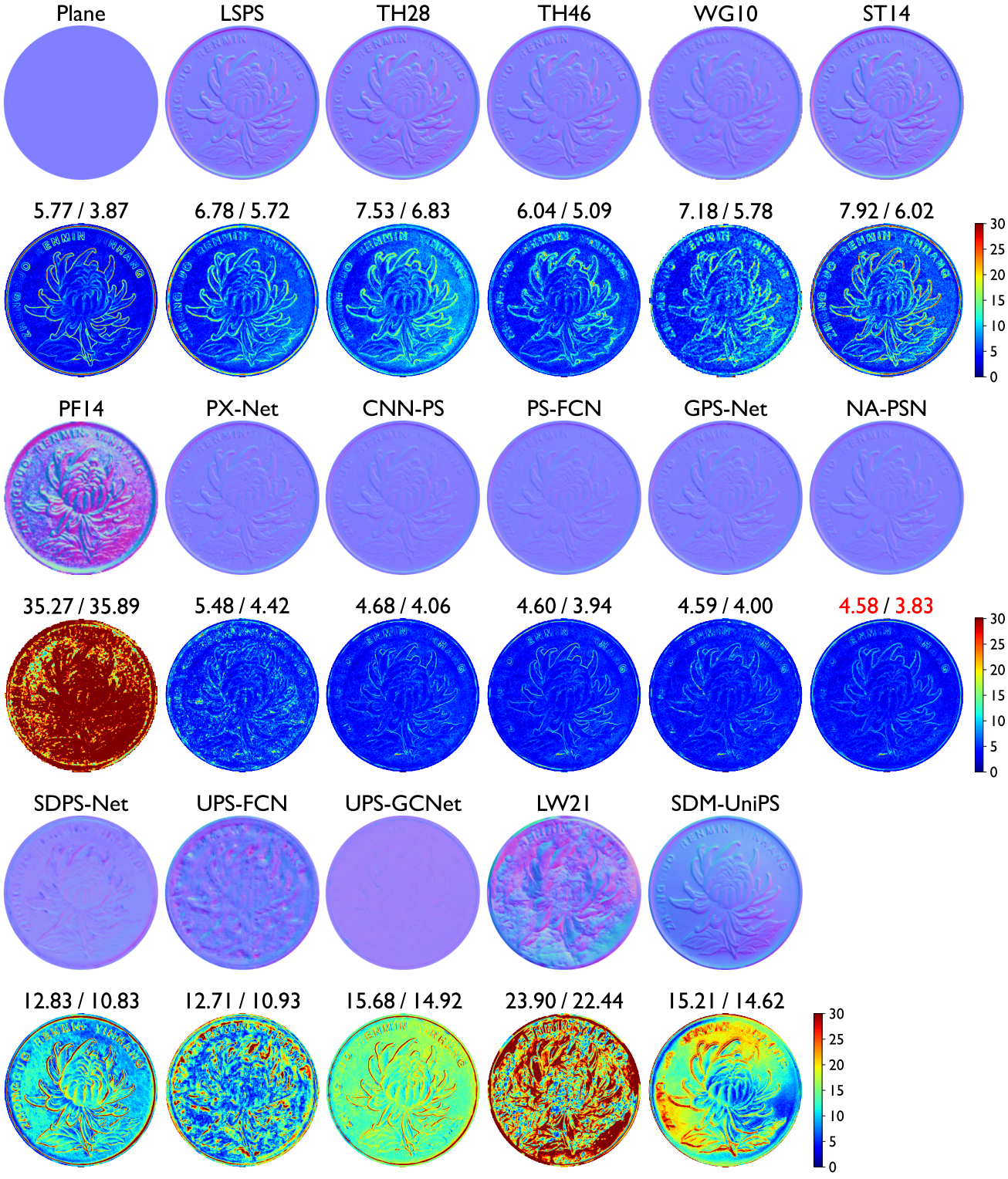


Figure S7: Estimated surface normal maps and angular error maps of FLOWER.

BIRD

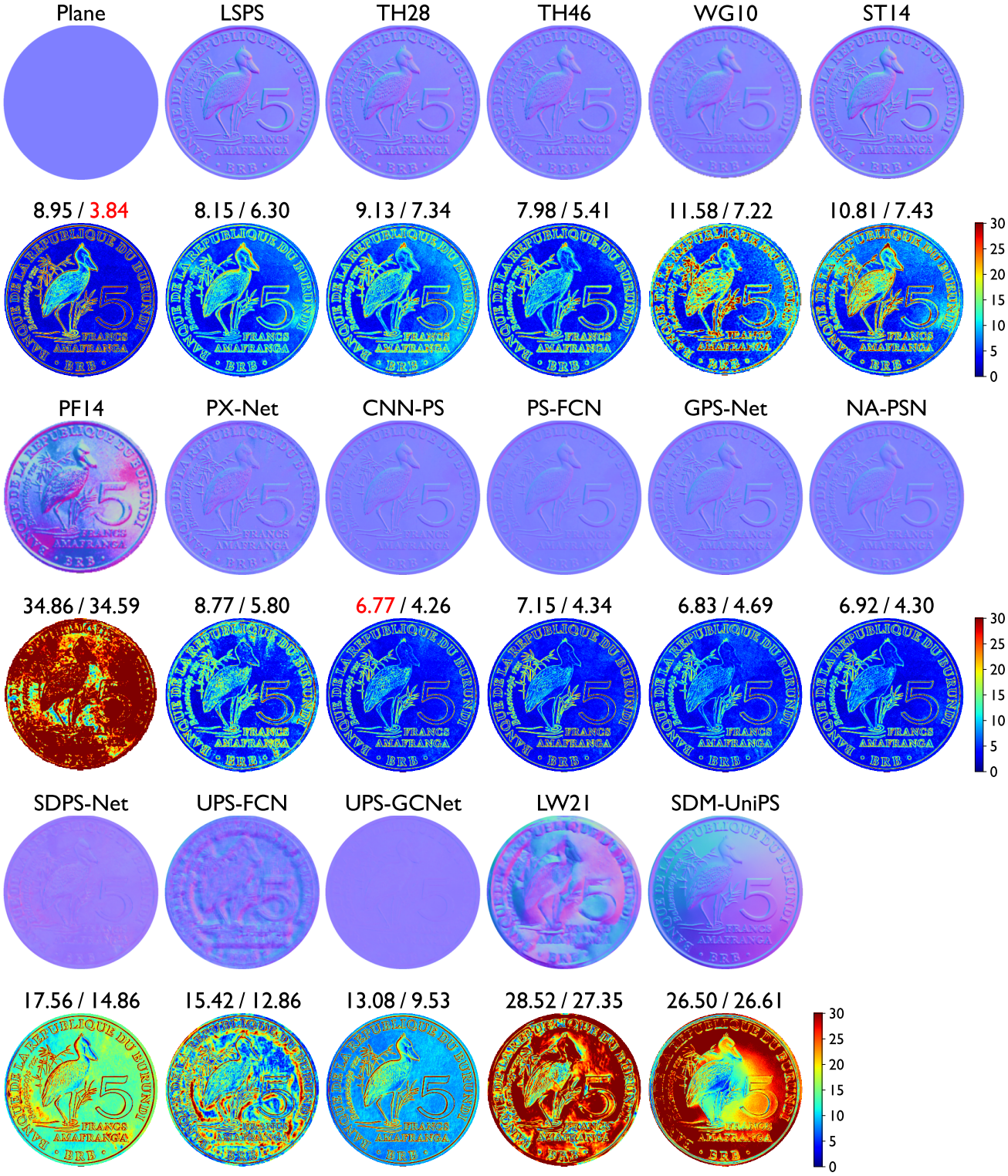


Figure S8: Estimated surface normal maps and angular error maps of BIRD.

LIONS

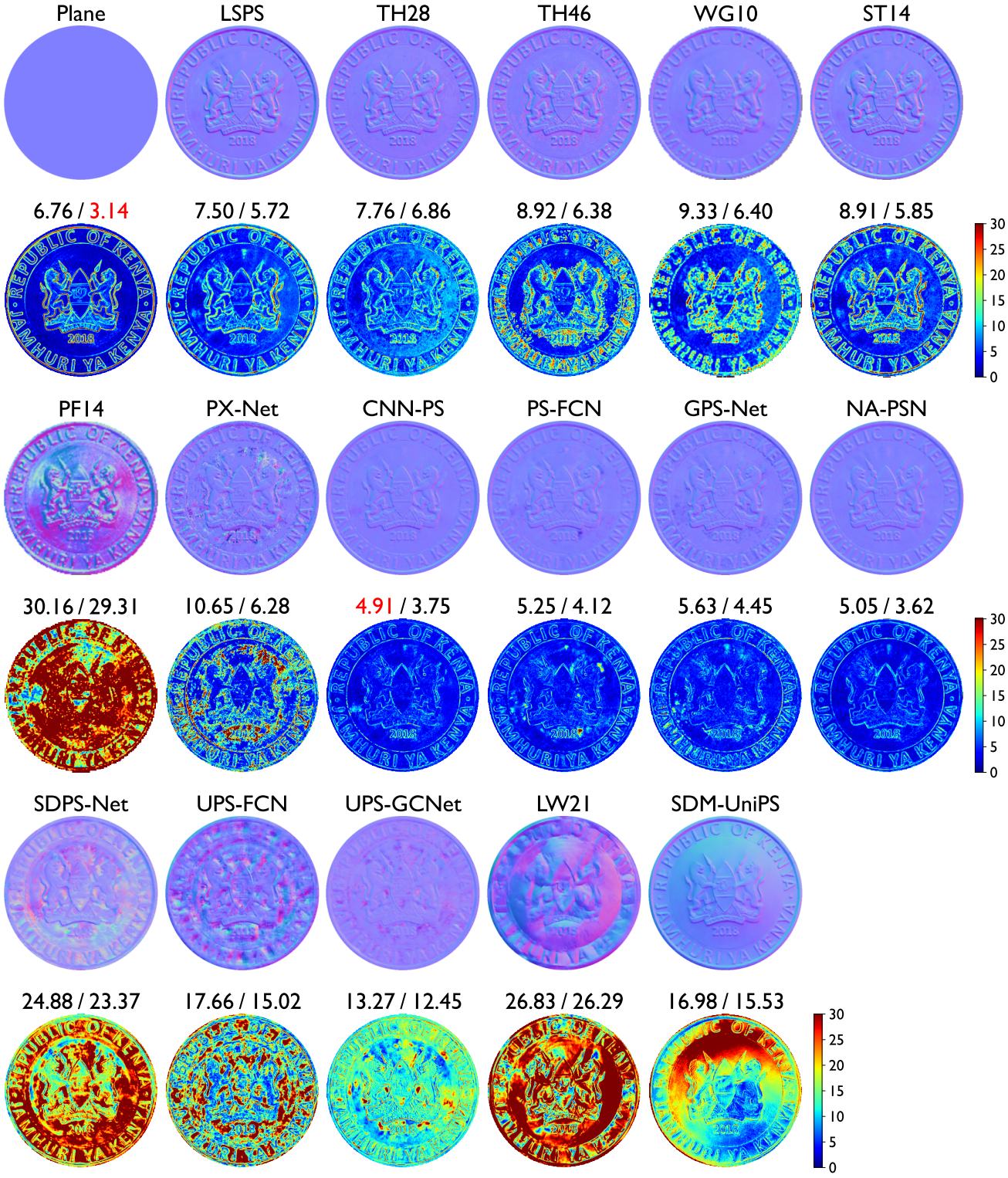


Figure S9: Estimated surface normal maps and angular error maps of LIONS.

RHINO

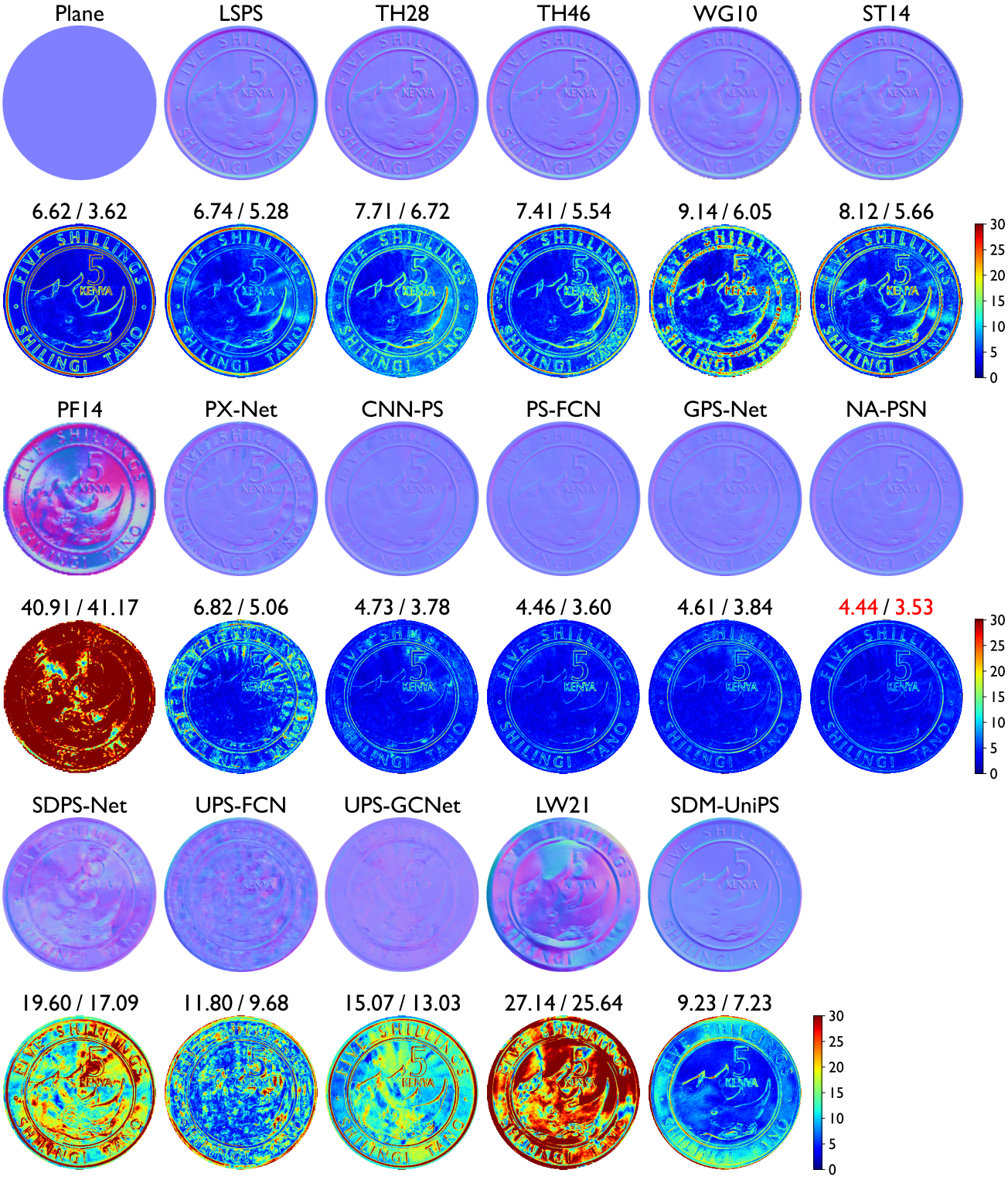


Figure S10: Estimated surface normal maps and angular error maps of RHINO.

QUEEN

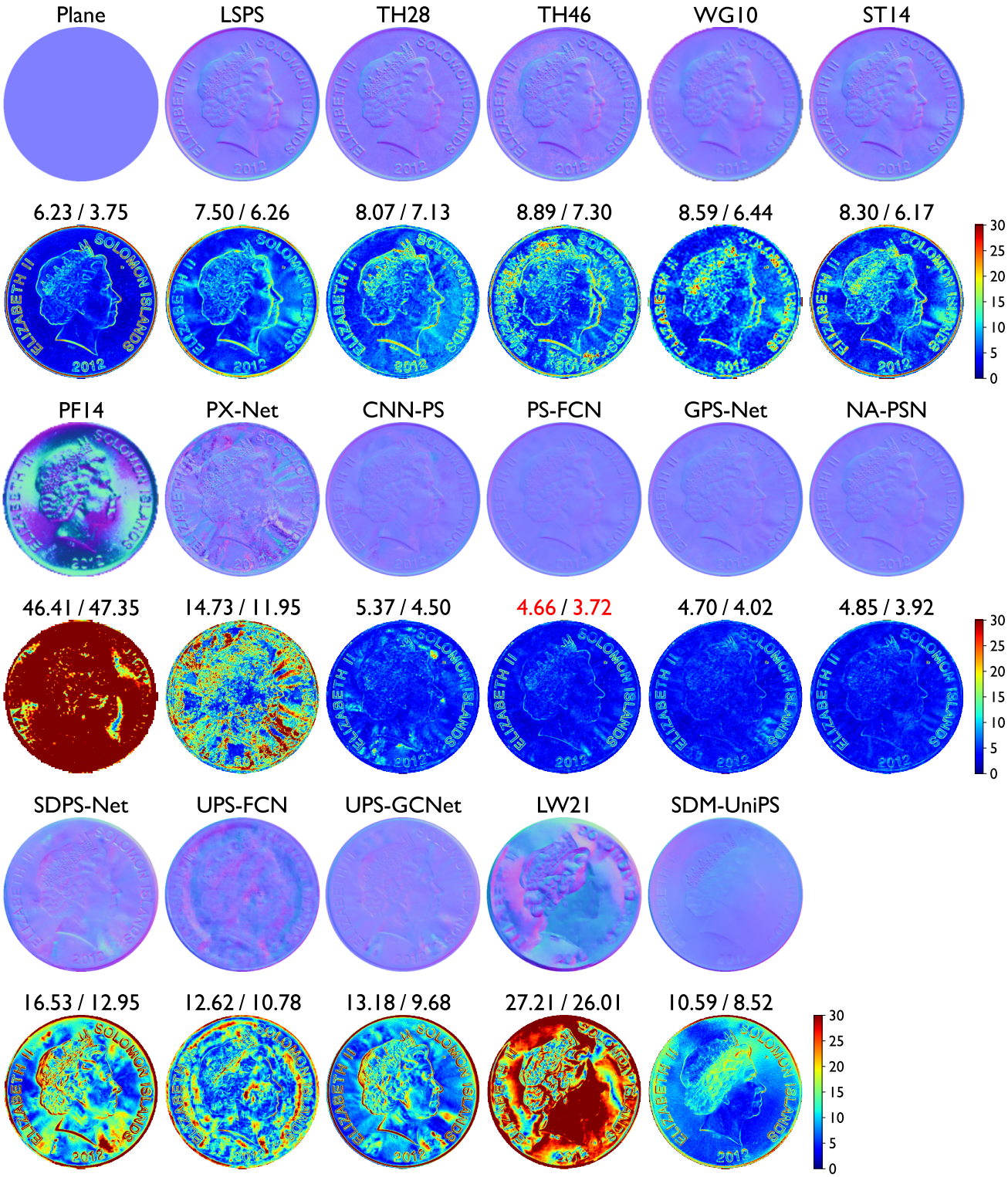


Figure S11: Estimated surface normal maps and angular error maps of QUEEN.

CRAB

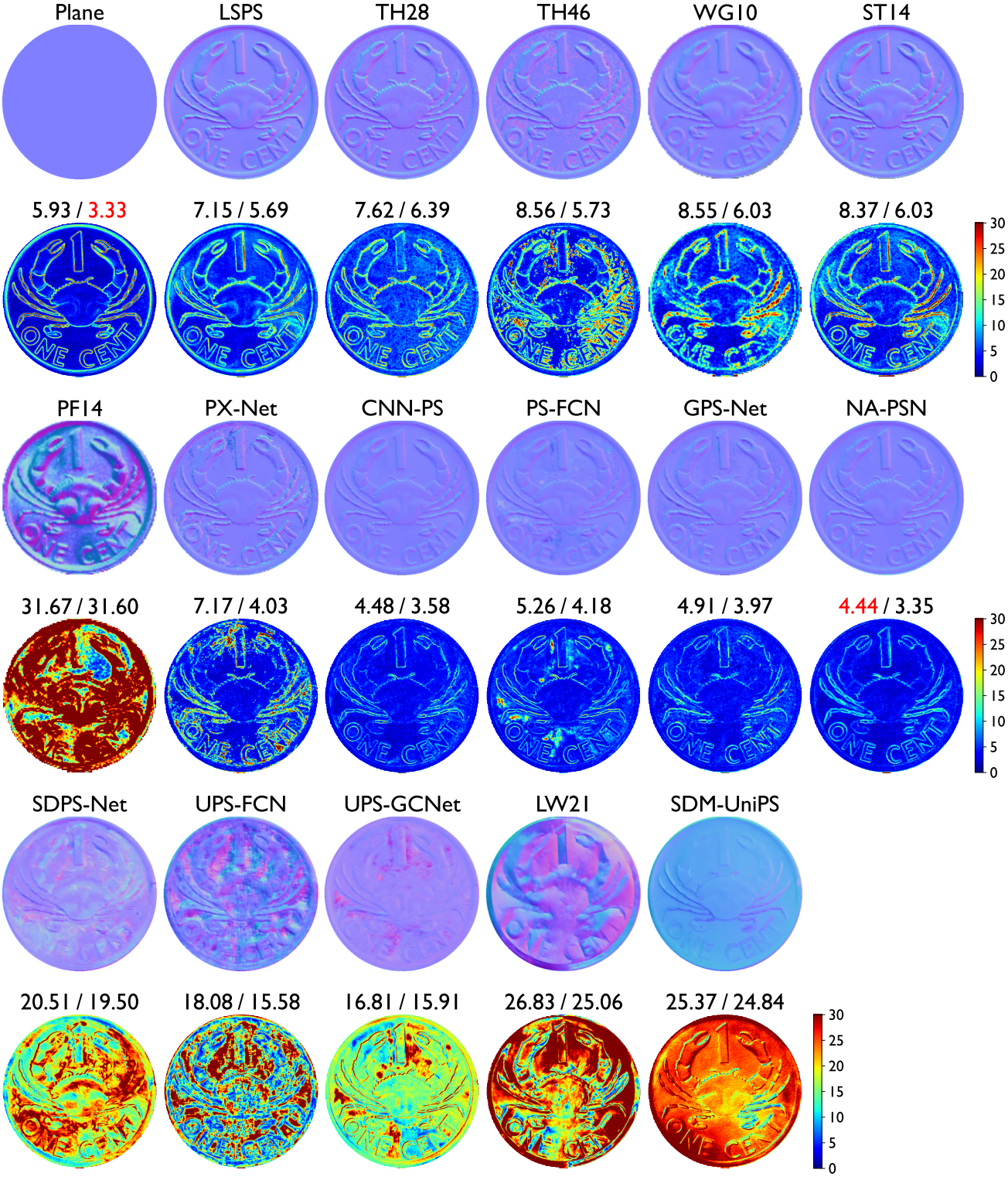


Figure S12: Estimated surface normal maps and angular error maps of CRAB.

SHIP

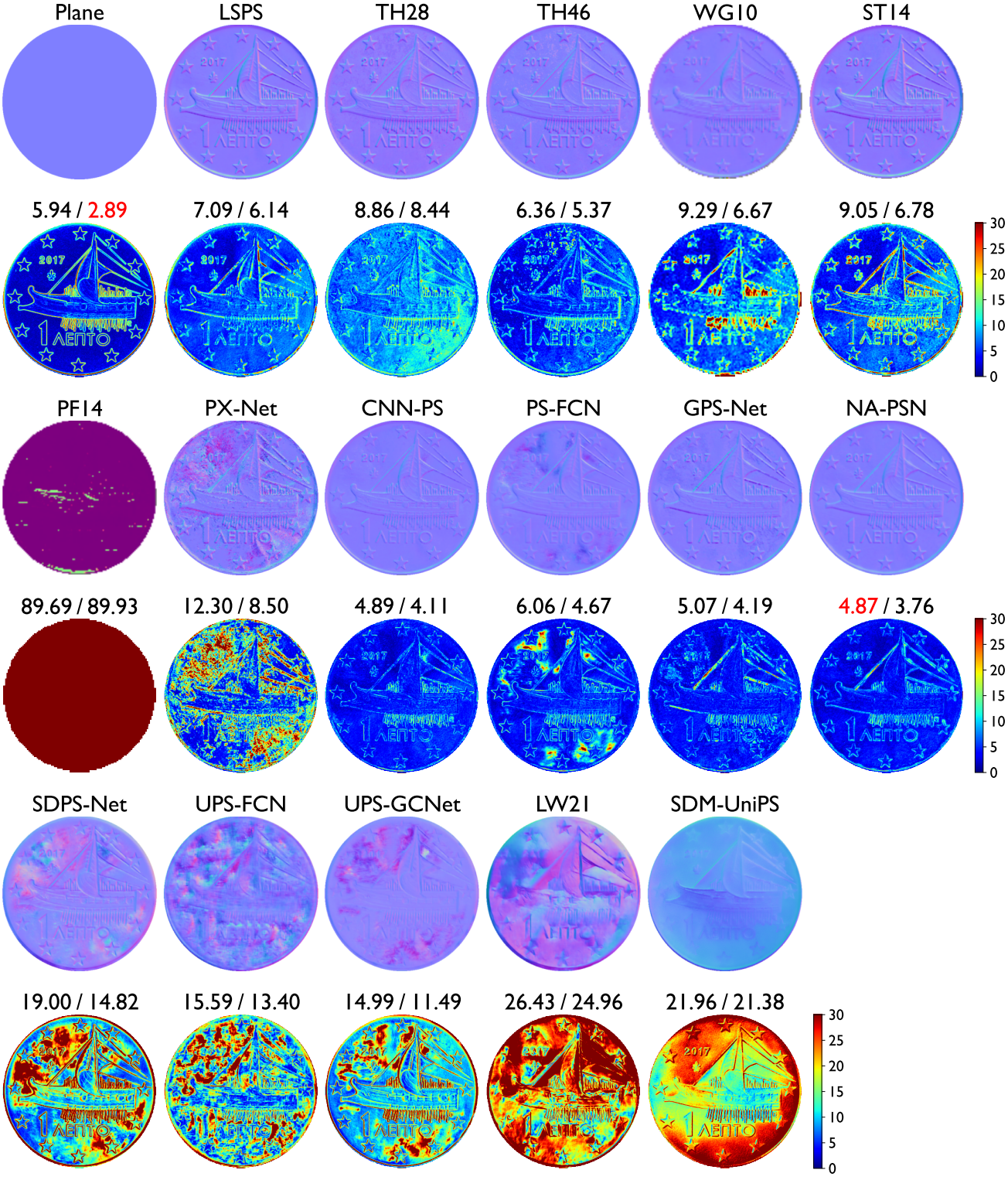


Figure S13: Estimated surface normal maps and angular error maps of SHIP.

PARA

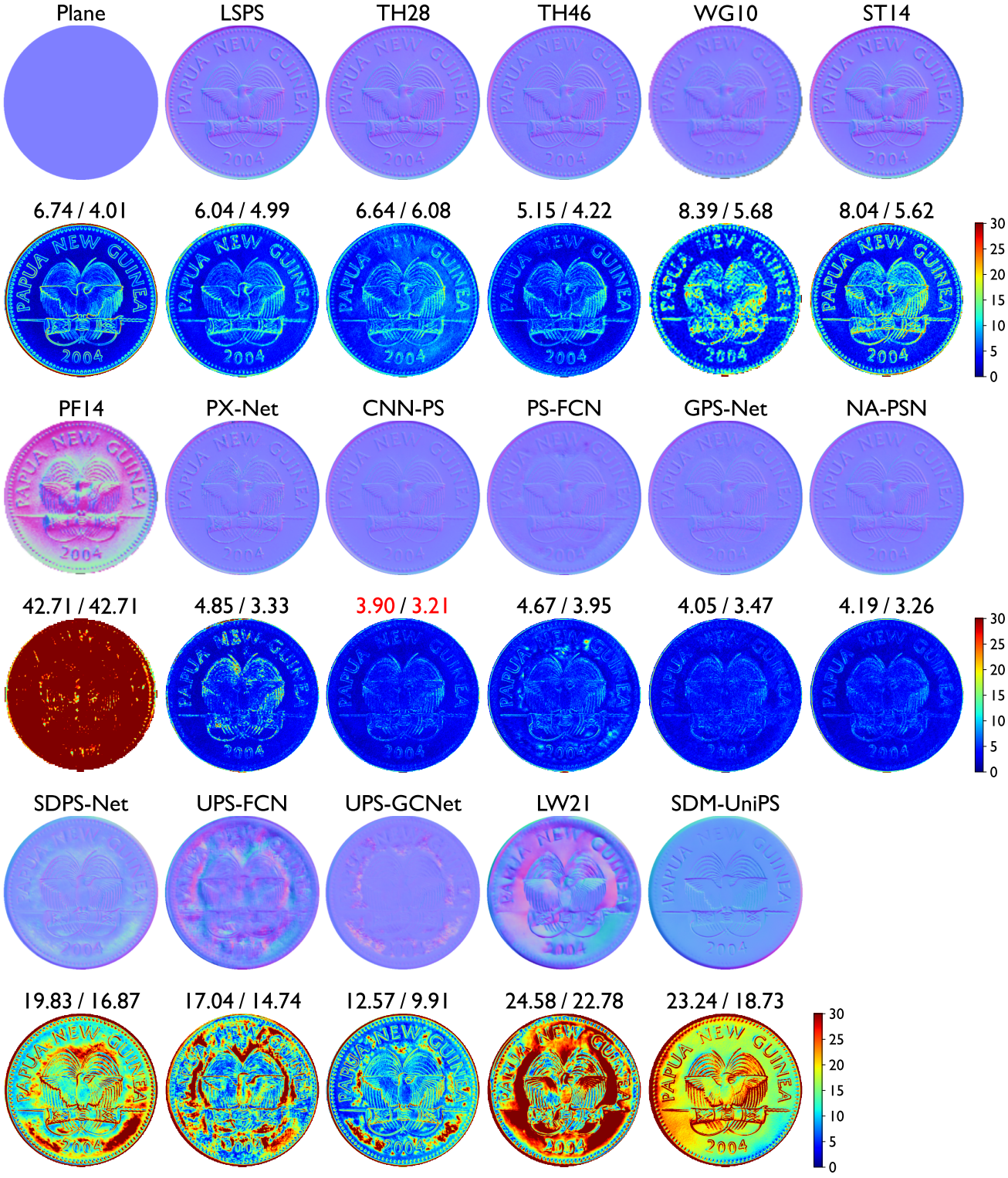


Figure S14: Estimated surface normal maps and angular error maps of PARA.

SAIL

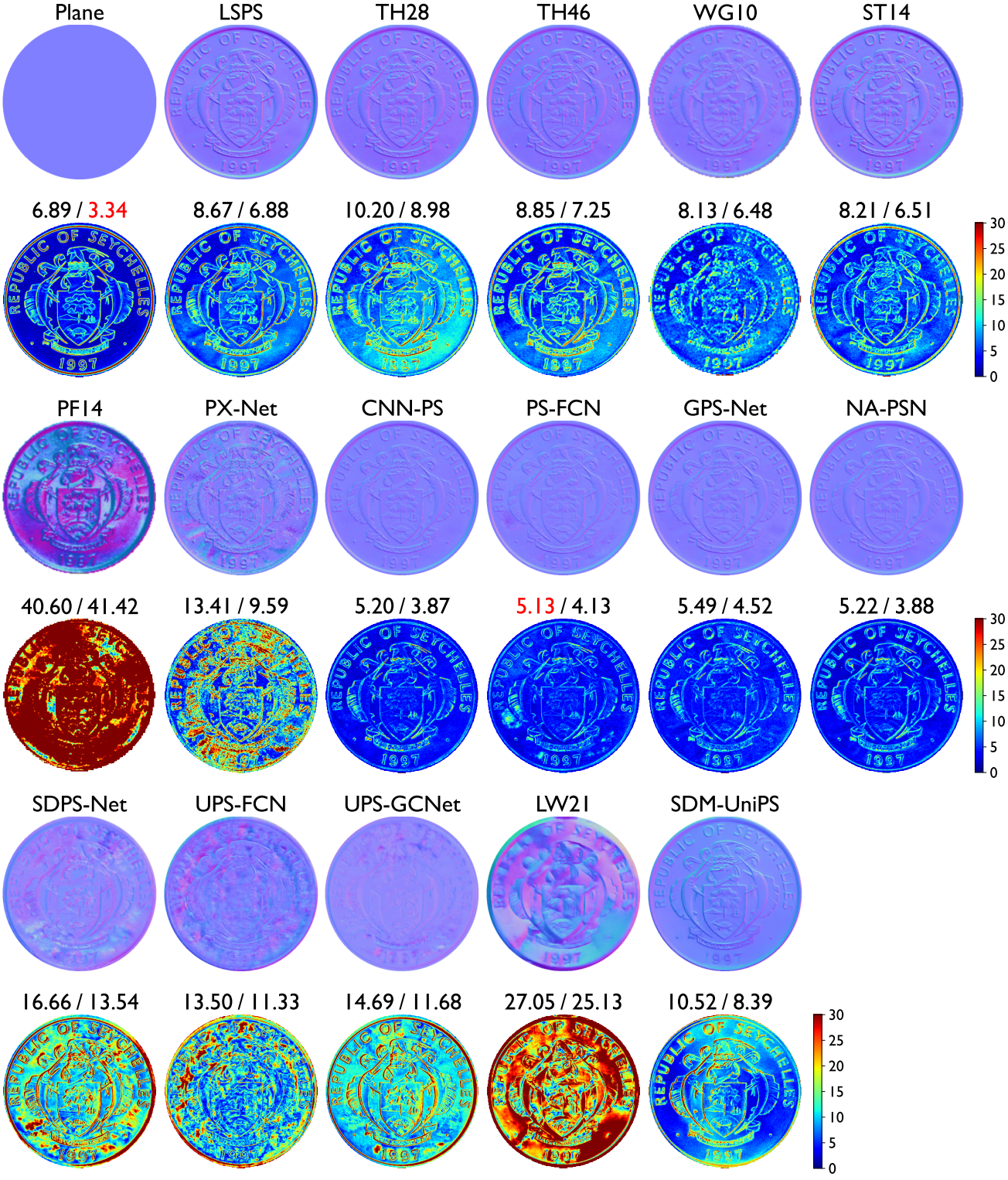


Figure S15: Estimated surface normal maps and angular error maps of SAIL.

FISH

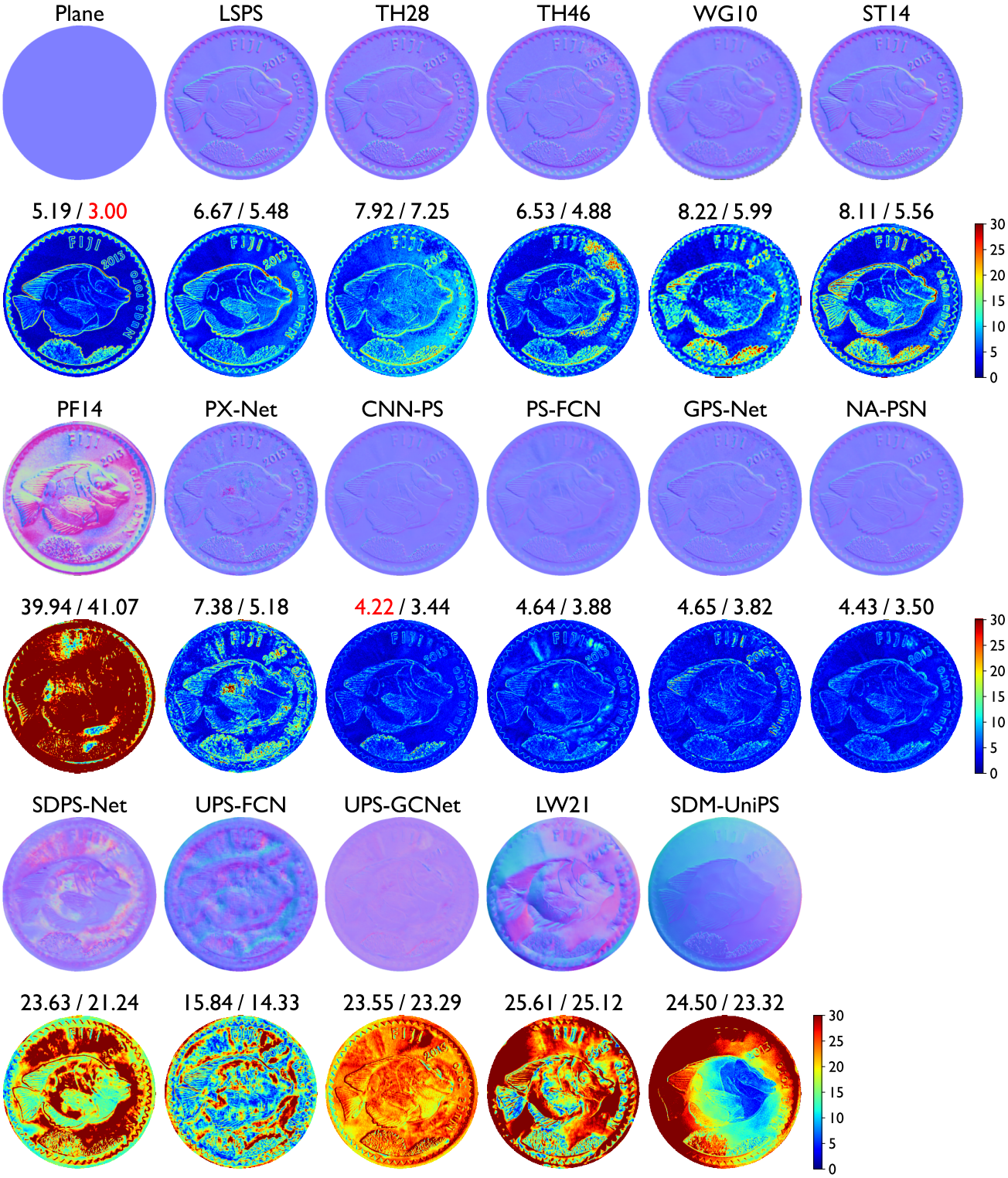


Figure S16: Estimated surface normal maps and angular error maps of FISH.

TREE



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

OCEAN

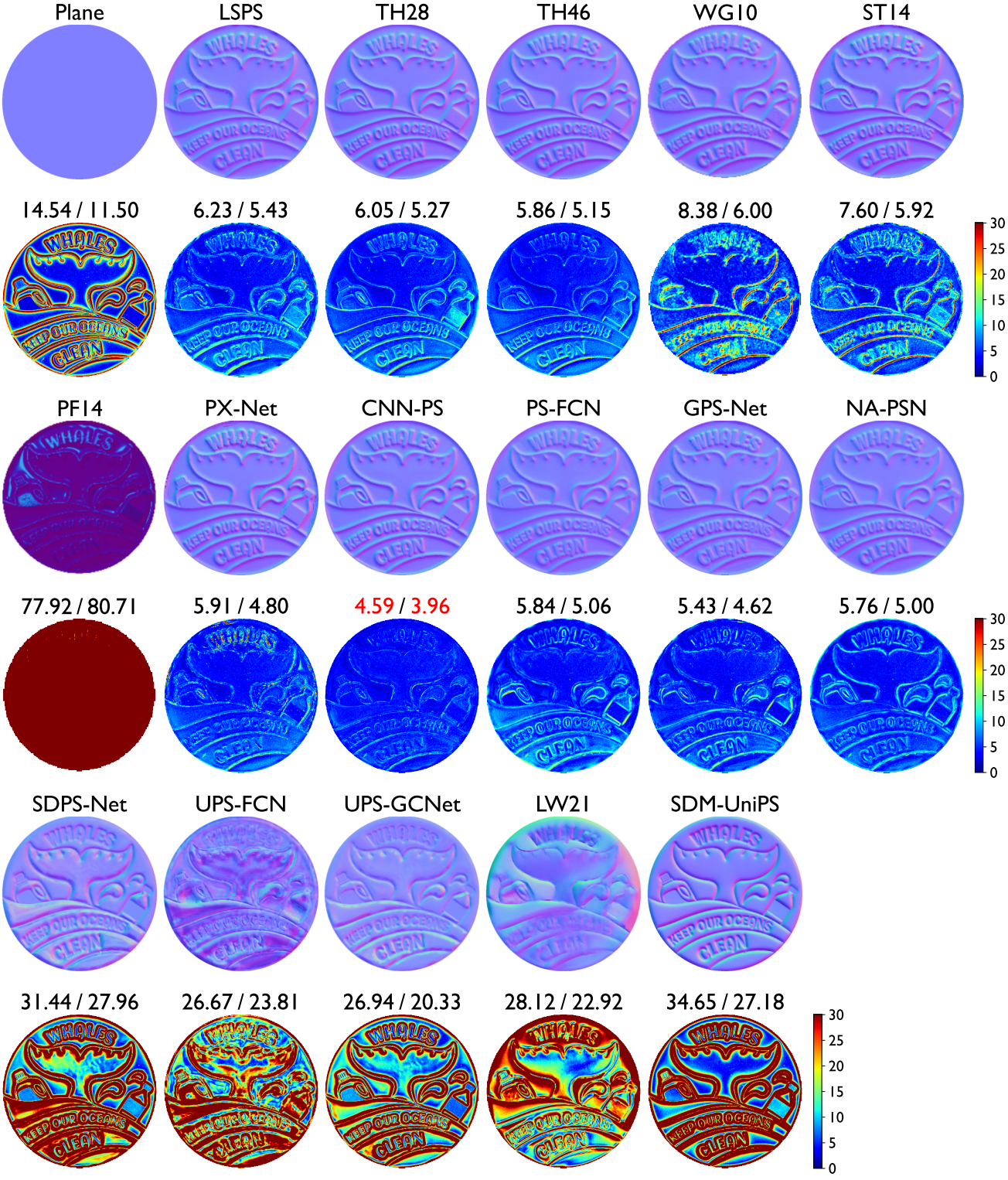


Figure S18: Estimated surface normal maps and angular error maps of OCEAN.

LUNG

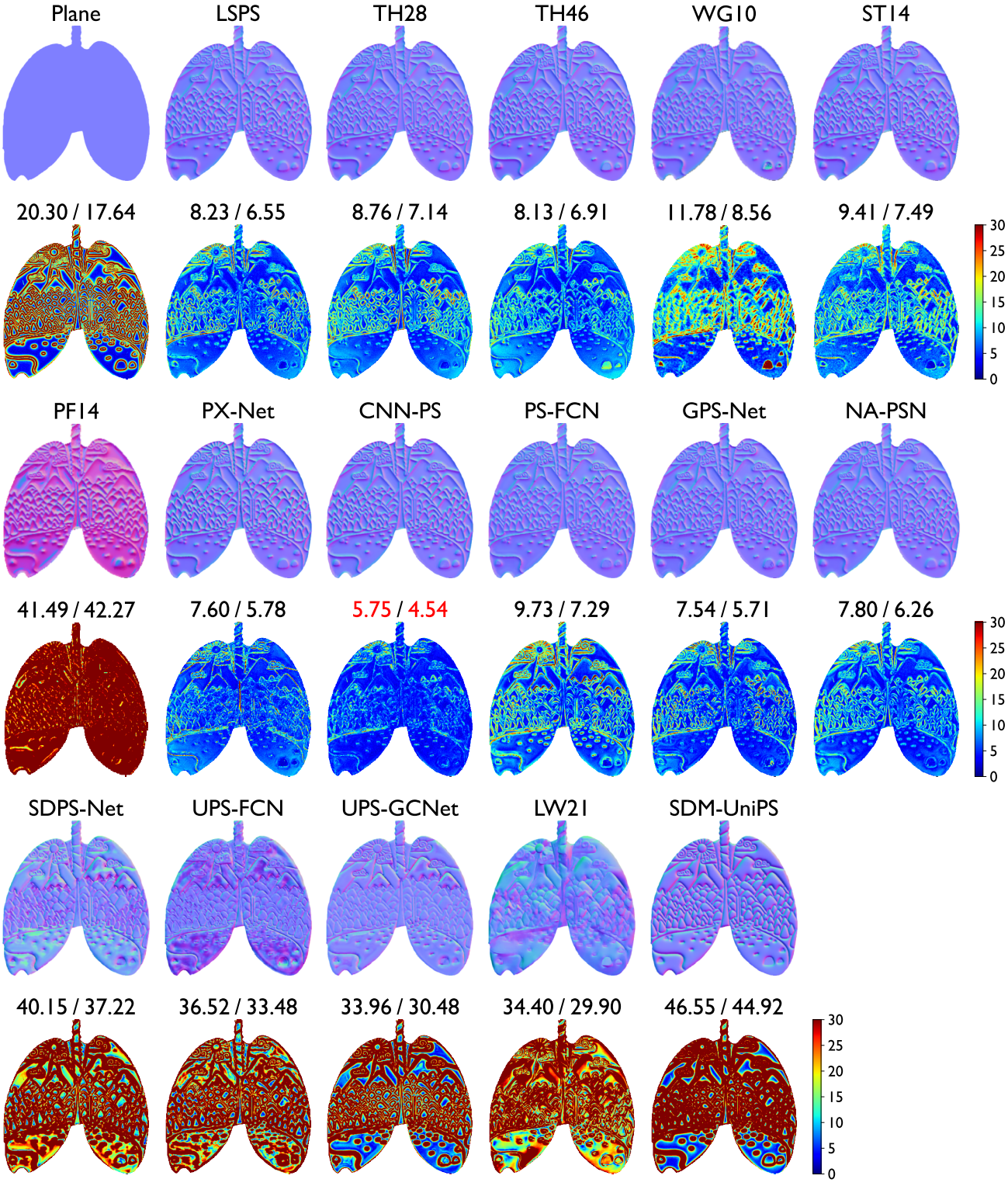


Figure S19: Estimated surface normal maps and angular error maps of LUNG.

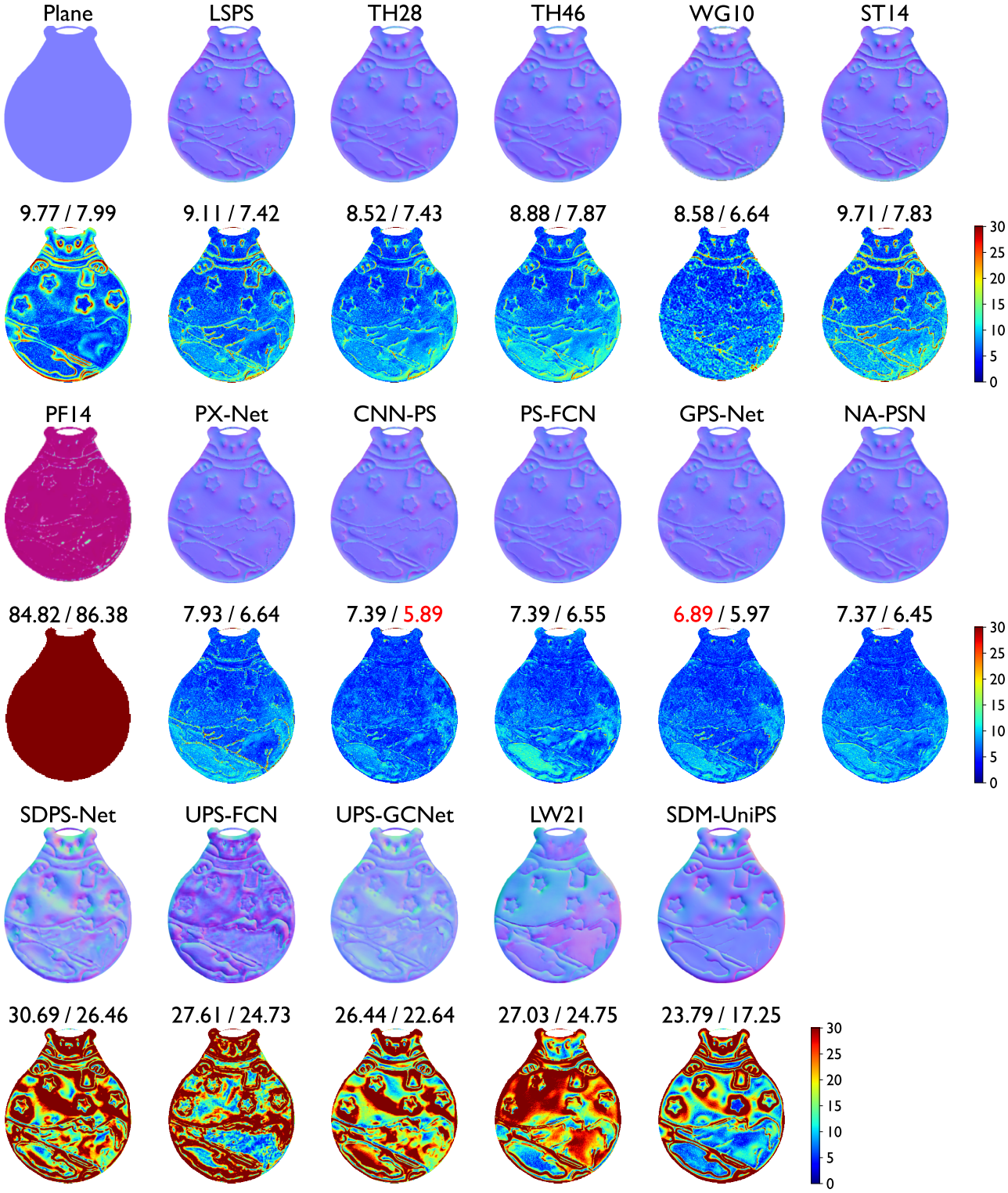


Figure S20: Estimated surface normal maps and angular error maps of BEAR.

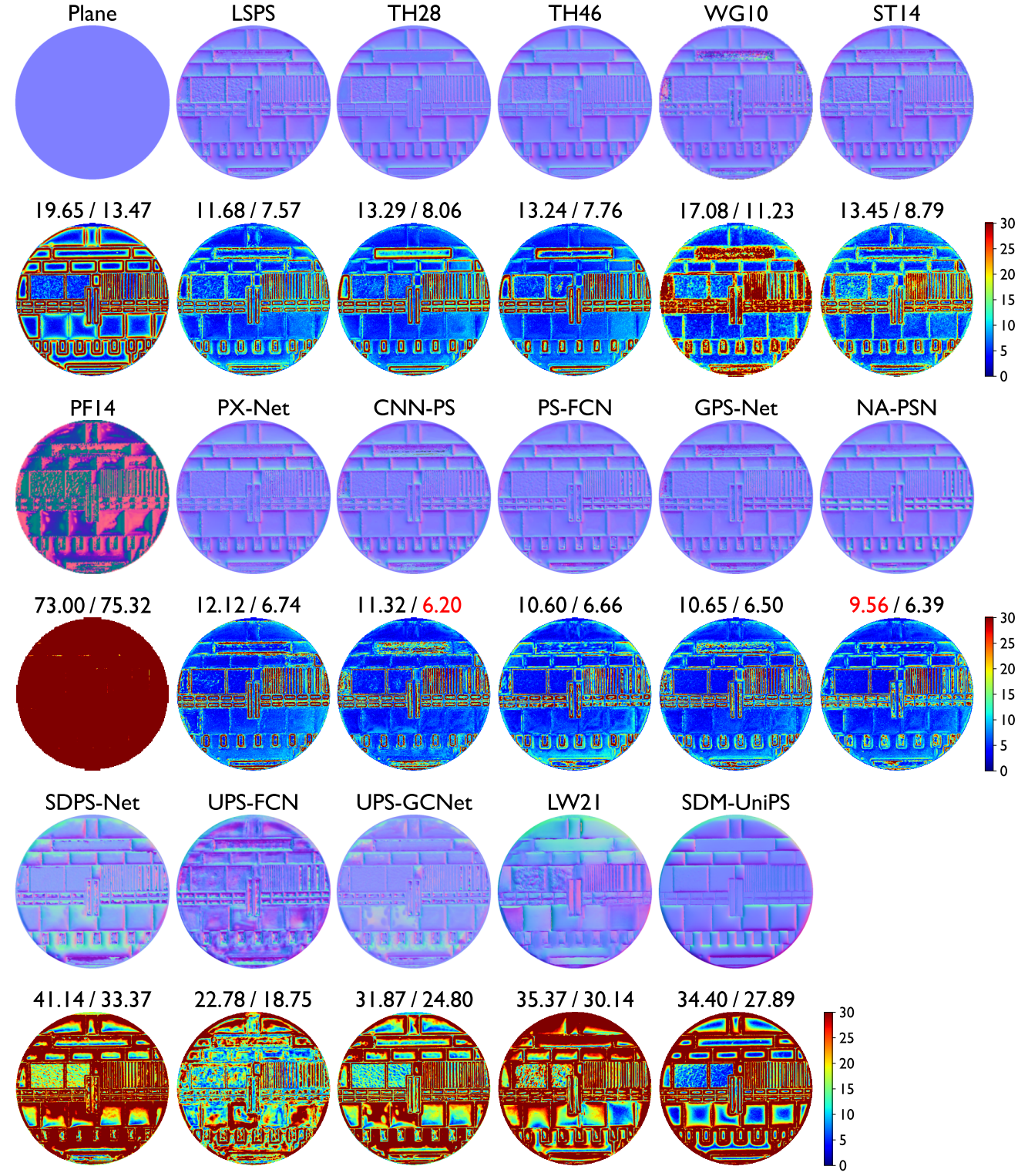


Figure S21: Estimated surface normal maps and angular error maps of Tv.

SUN

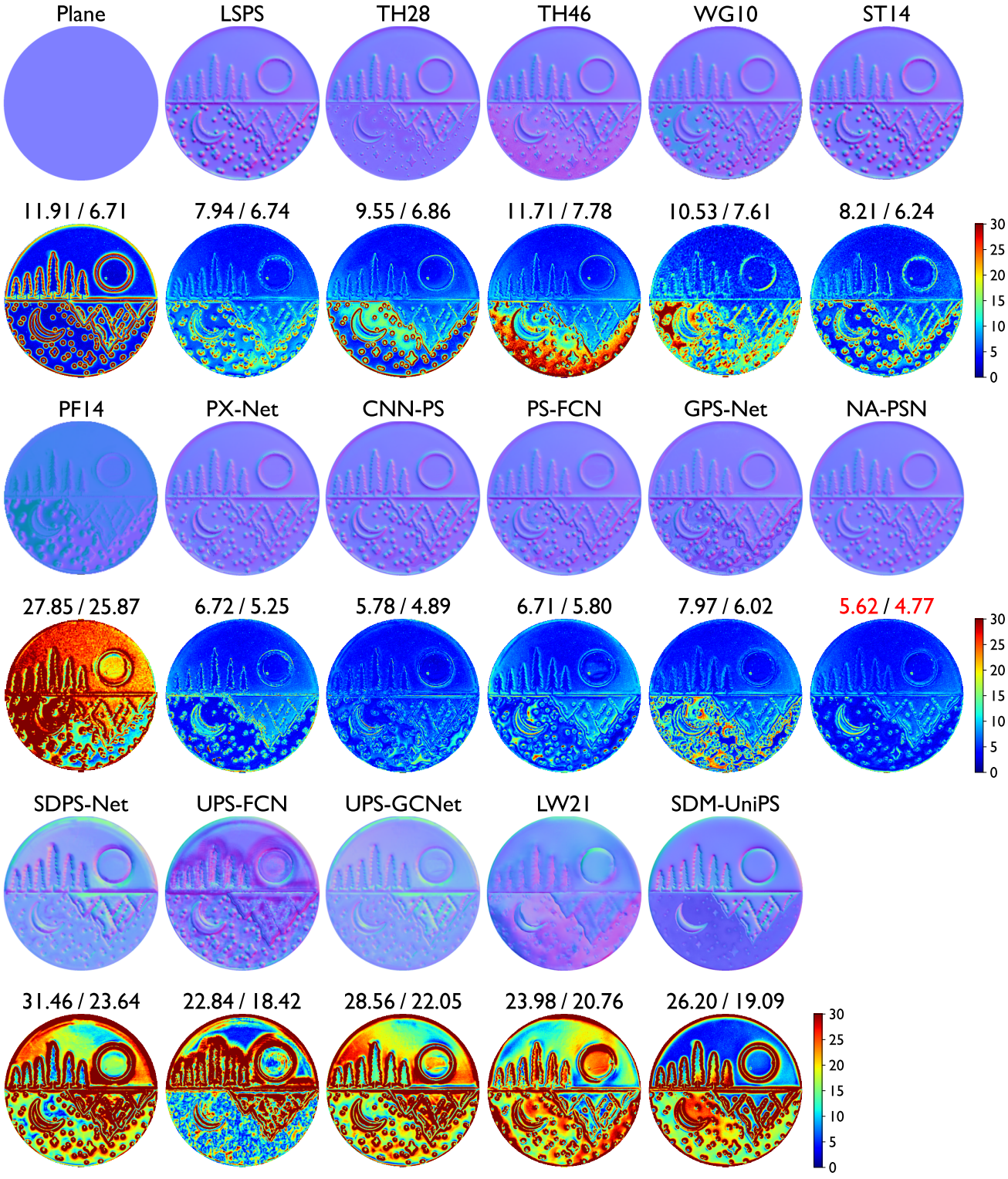


Figure S22: Estimated surface normal maps and angular error maps of SUN.

TAICHI

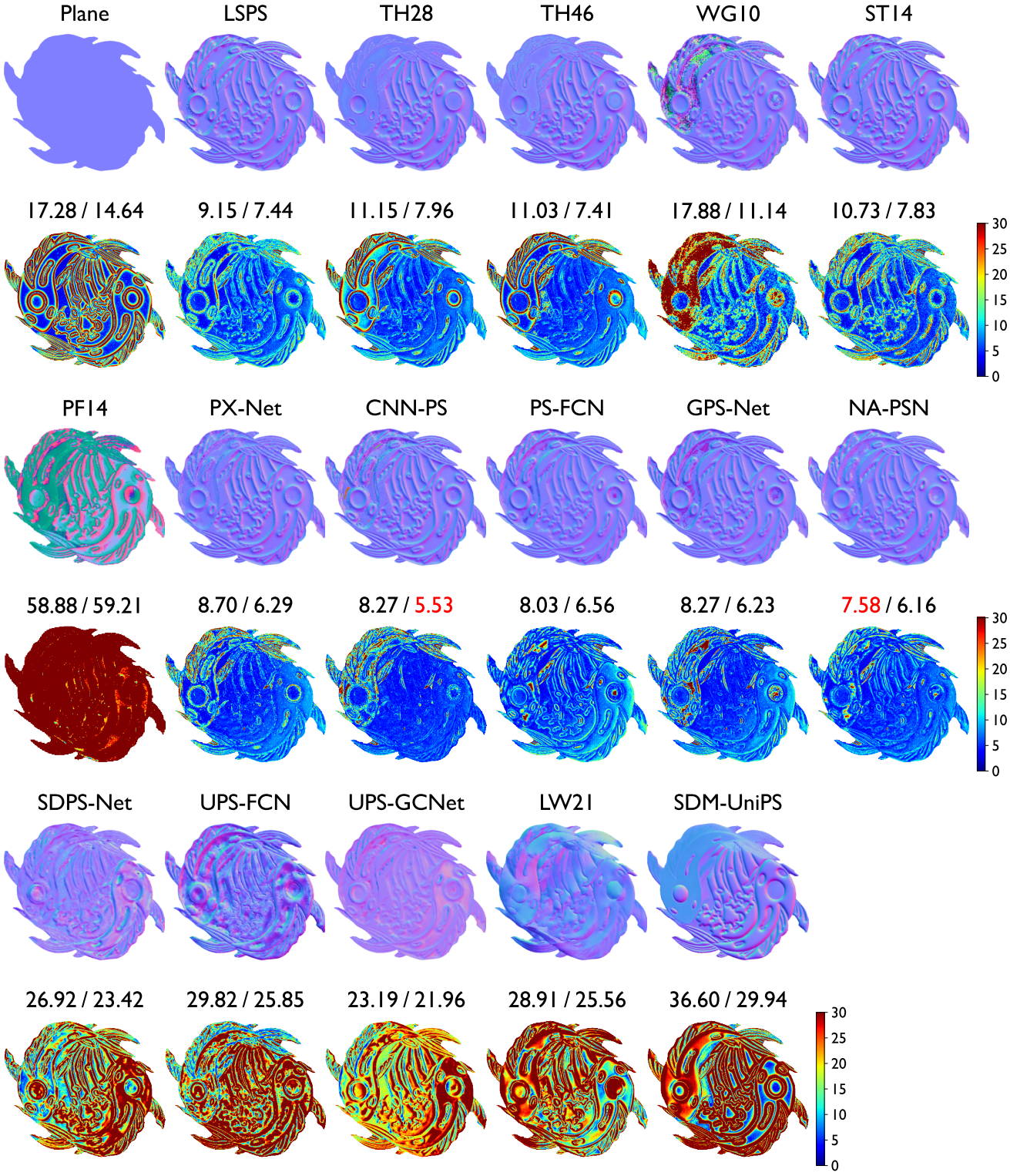


Figure S23: Estimated surface normal maps and angular error maps of TAICHI.

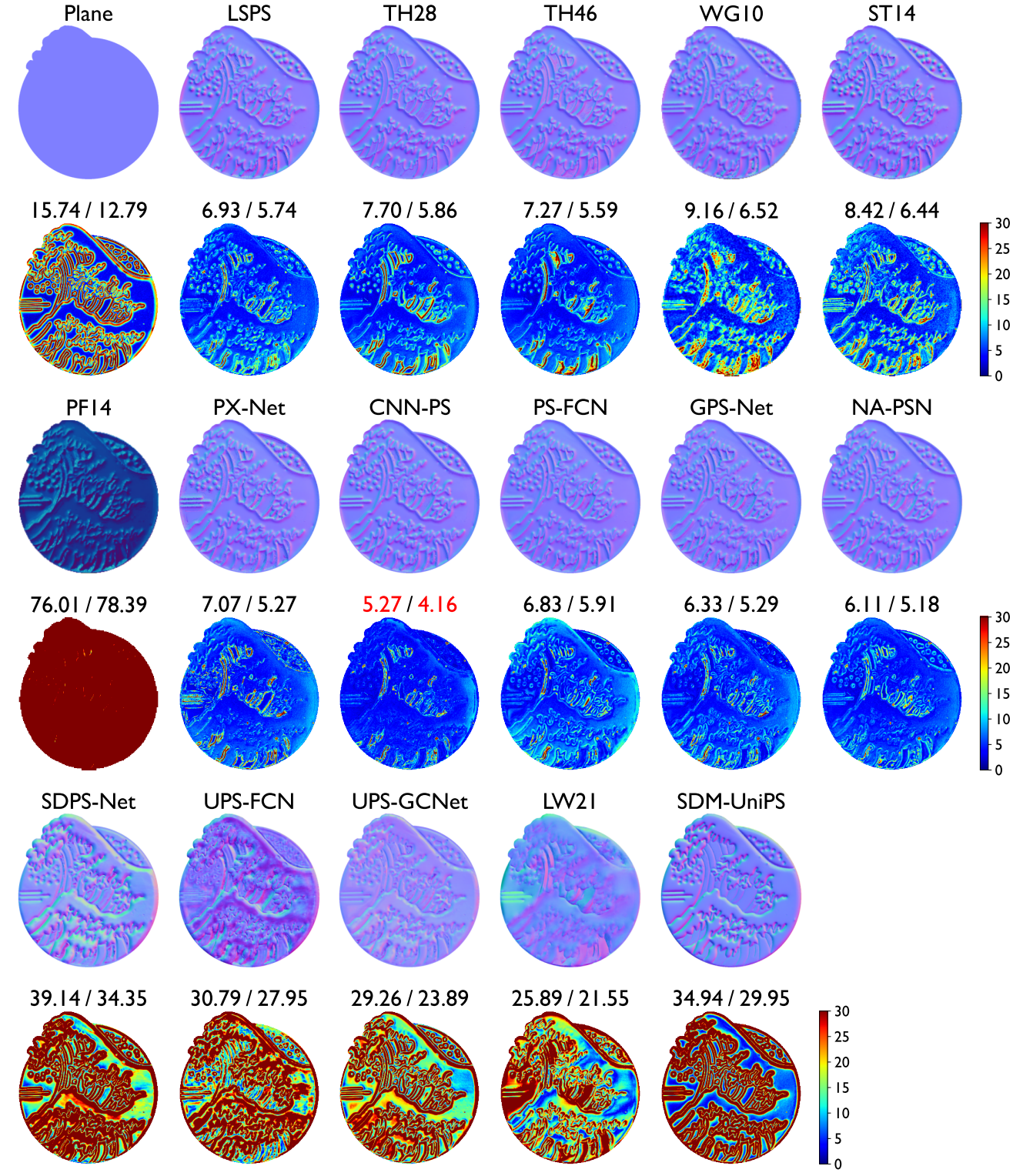


Figure S24: Estimated surface normal maps and angular error maps of WAVE.

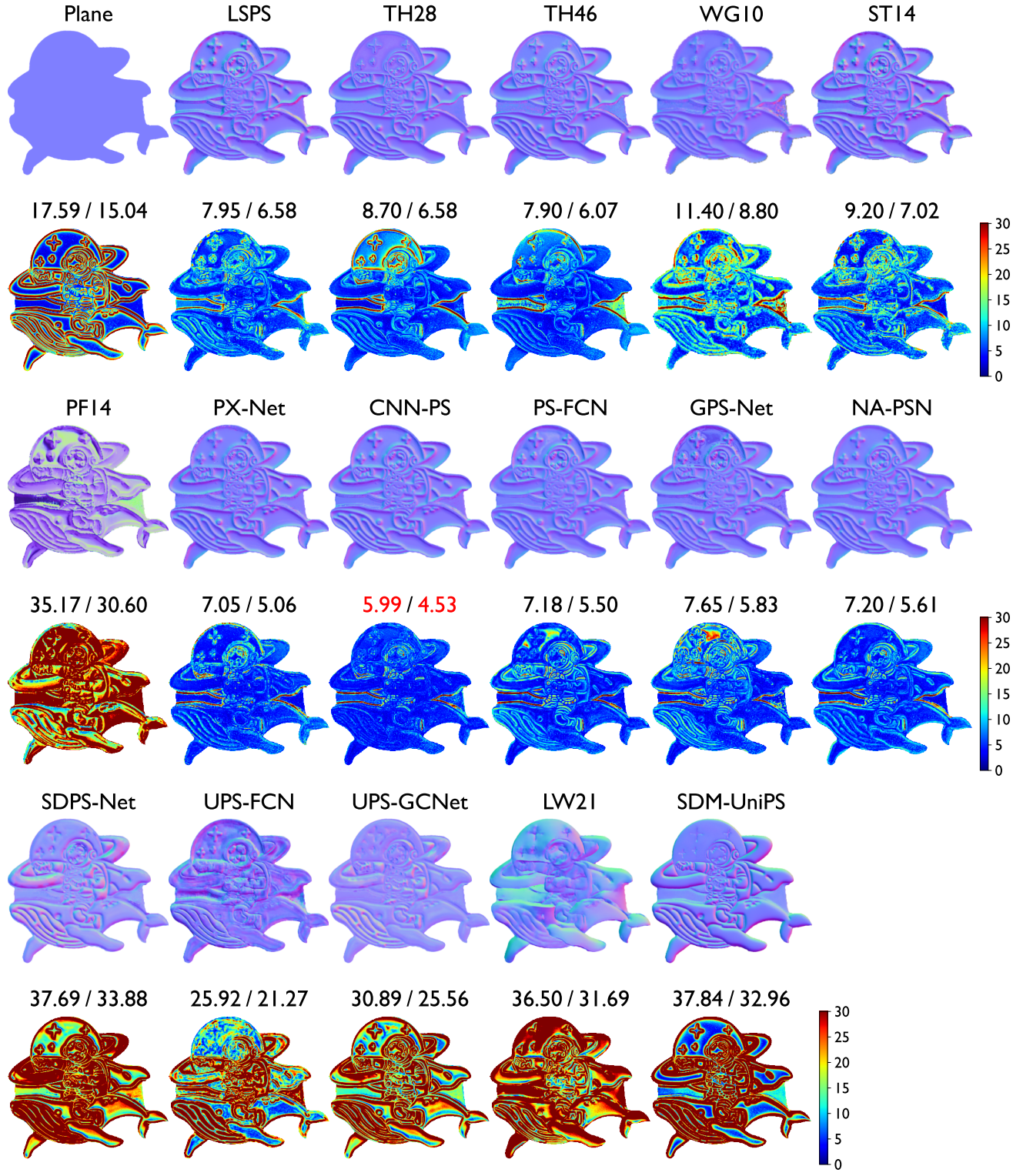


Figure S25: Estimated surface normal maps and angular error maps of ASTRO.

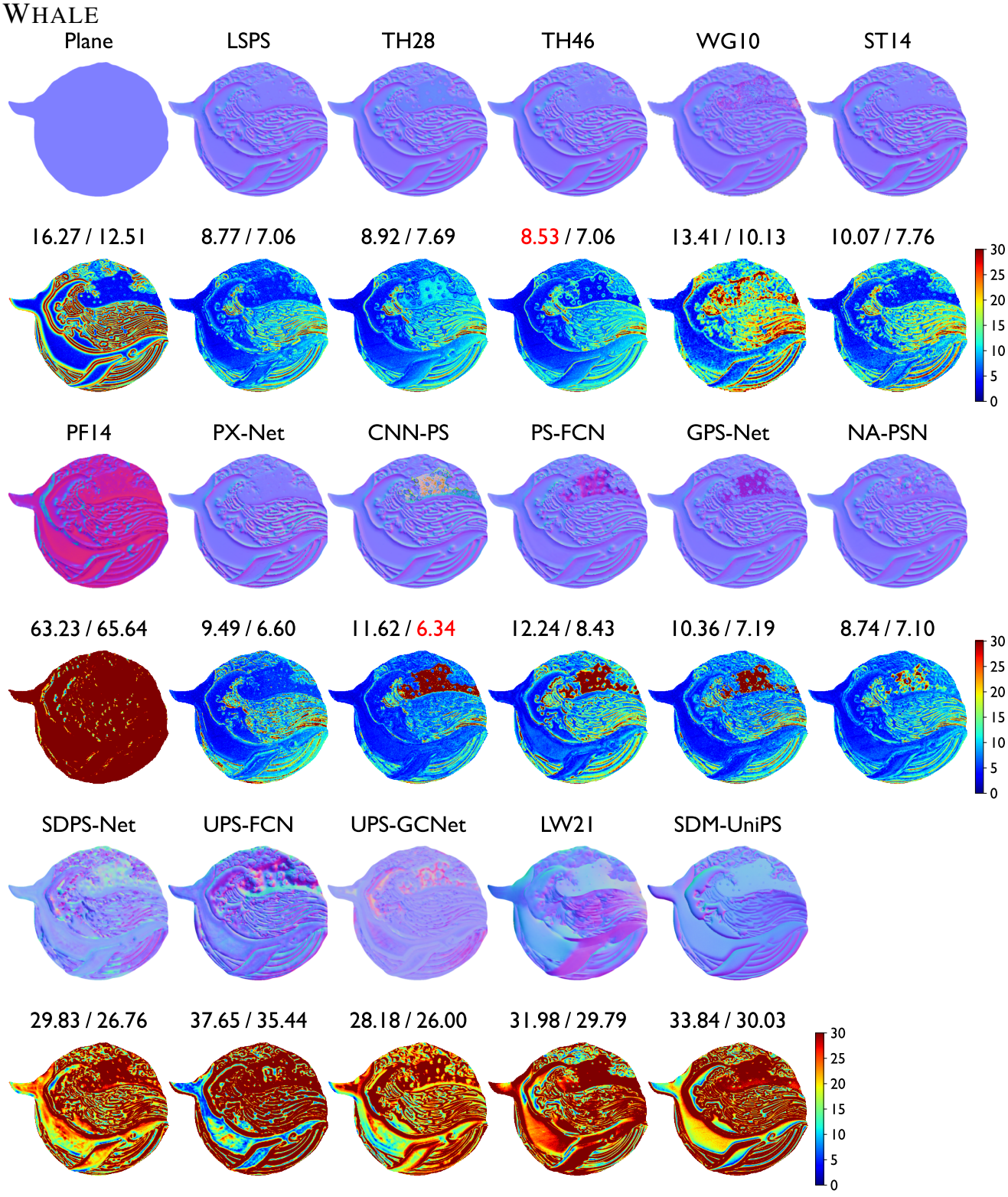


Figure S26: Estimated surface normal maps and angular error maps of WHALE.

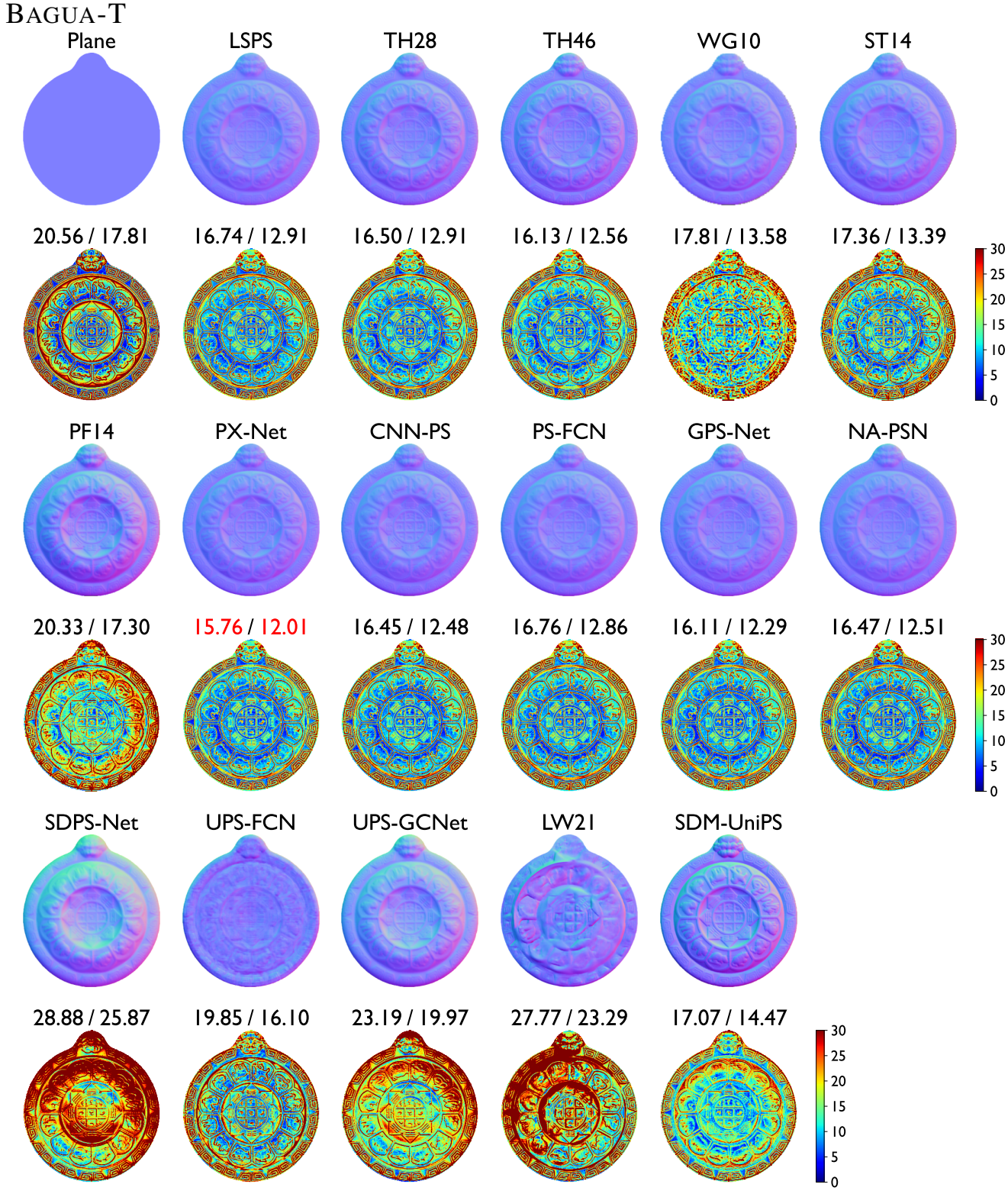


Figure S27: Estimated surface normal maps and angular error maps of BAGUA-T.

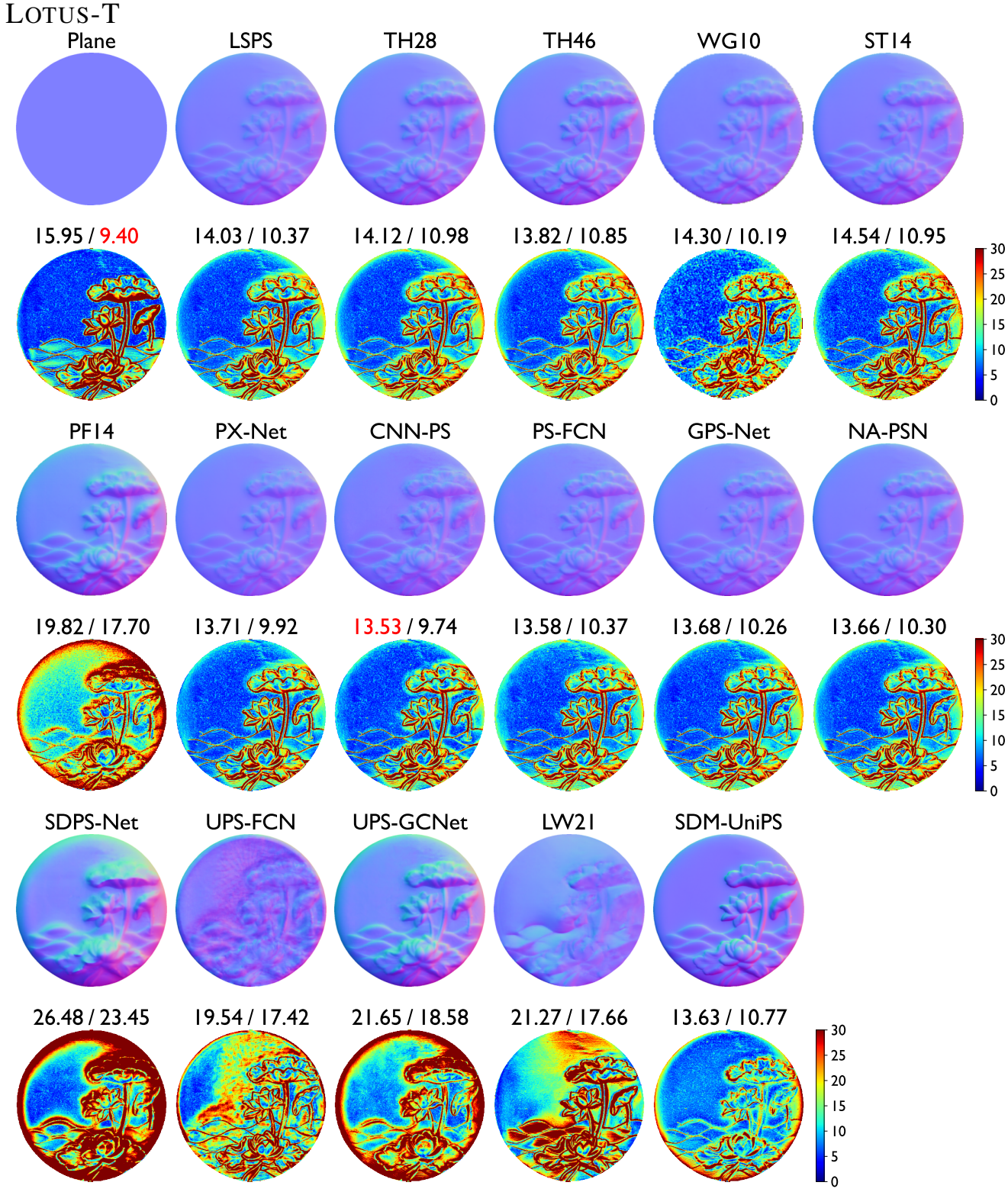


Figure S28: Estimated surface normal maps and angular error maps of LOTUS-T.

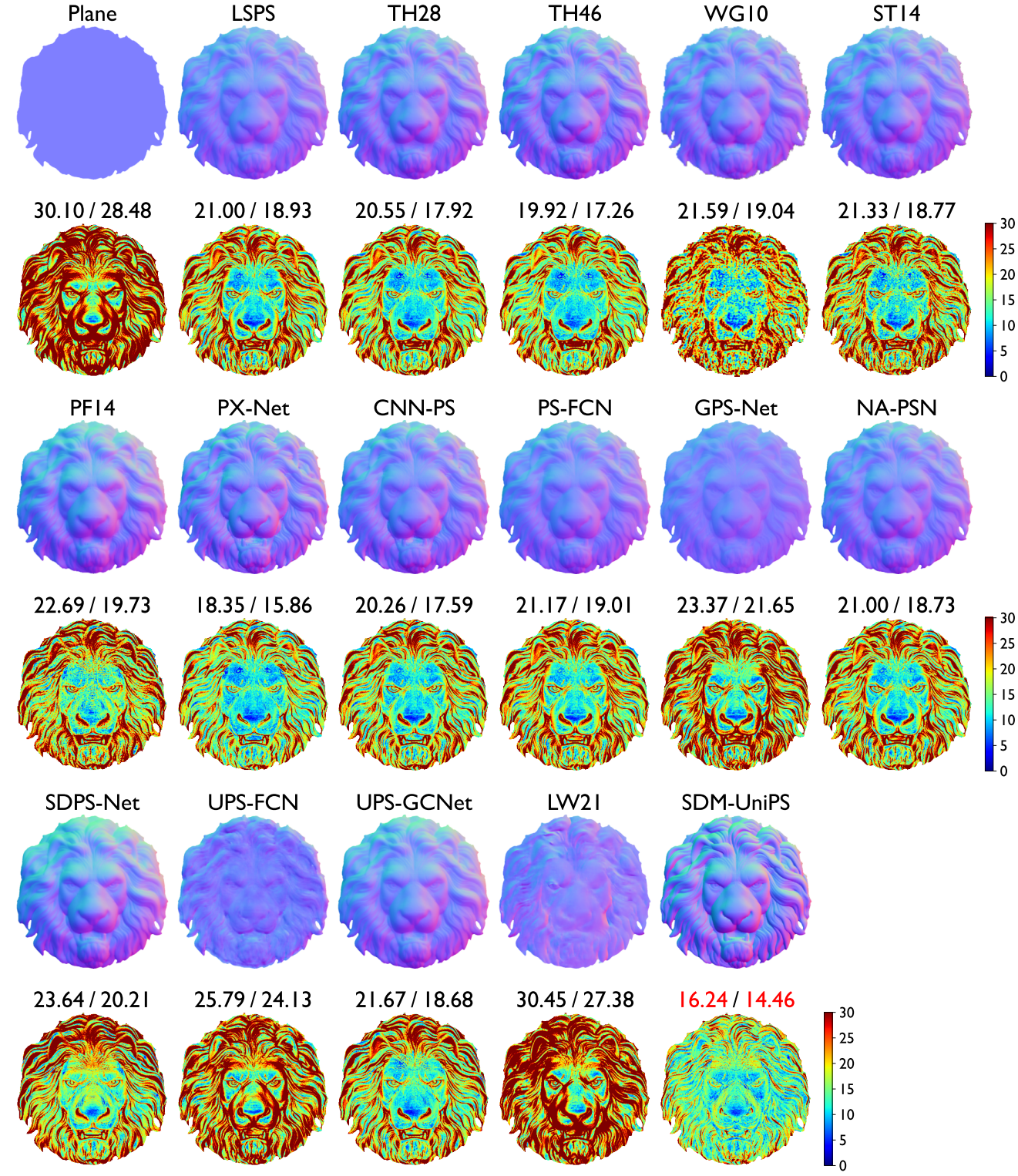


Figure S29: Estimated surface normal maps and angular error maps of LION-T.

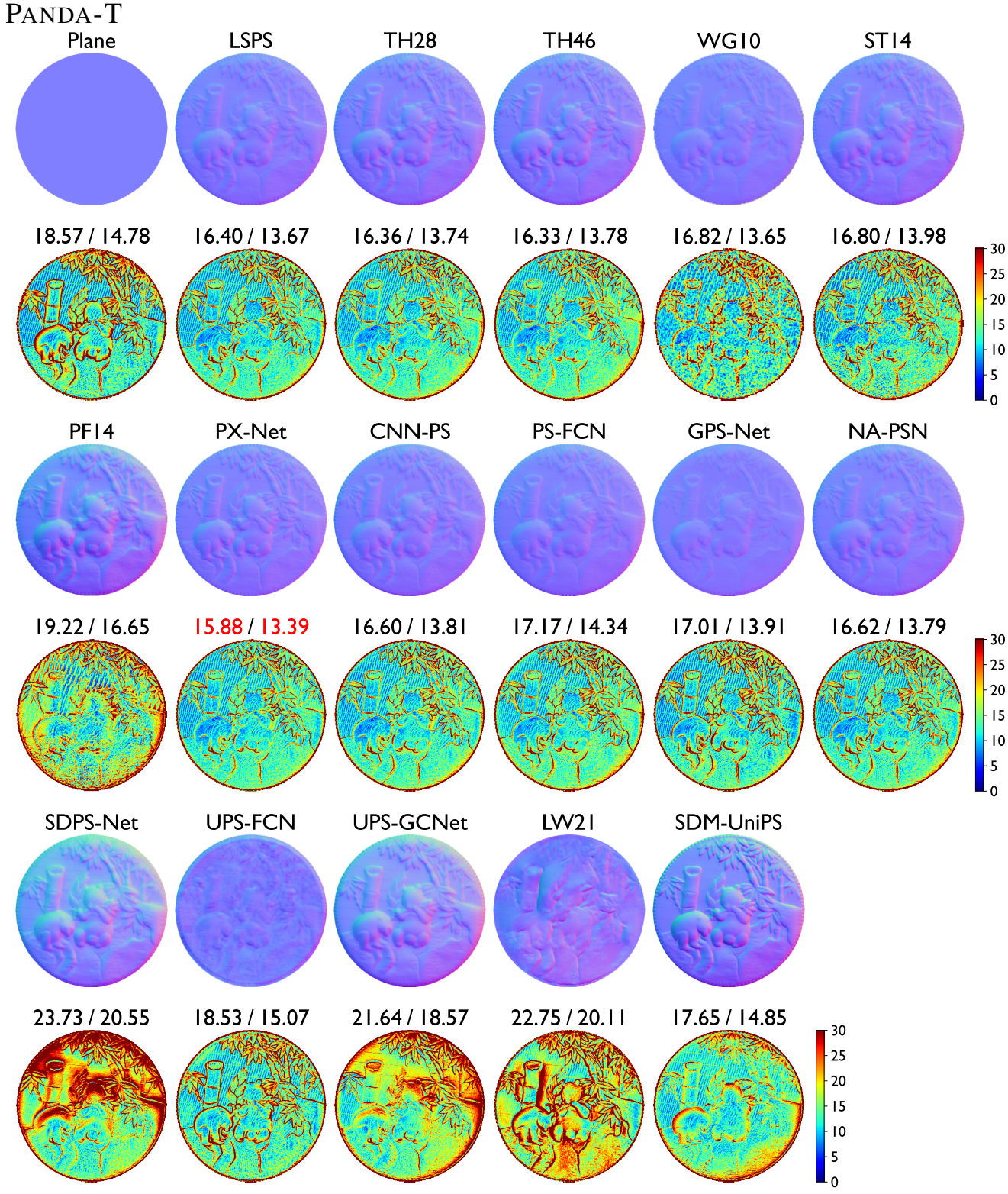


Figure S30: Estimated surface normal maps and angular error maps of PANDA-T.

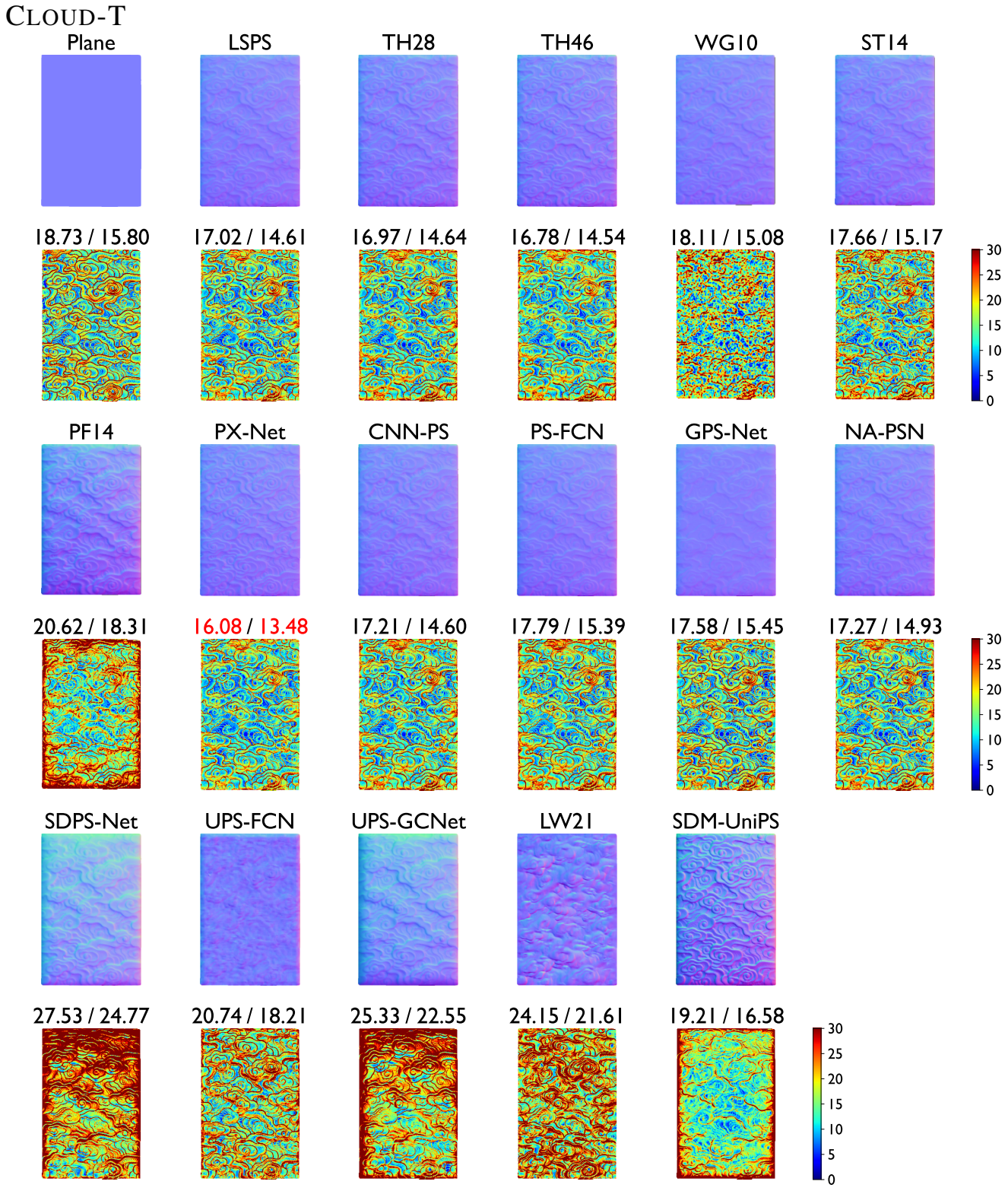


Figure S31: Estimated surface normal maps and angular error maps of CLOUD-T.

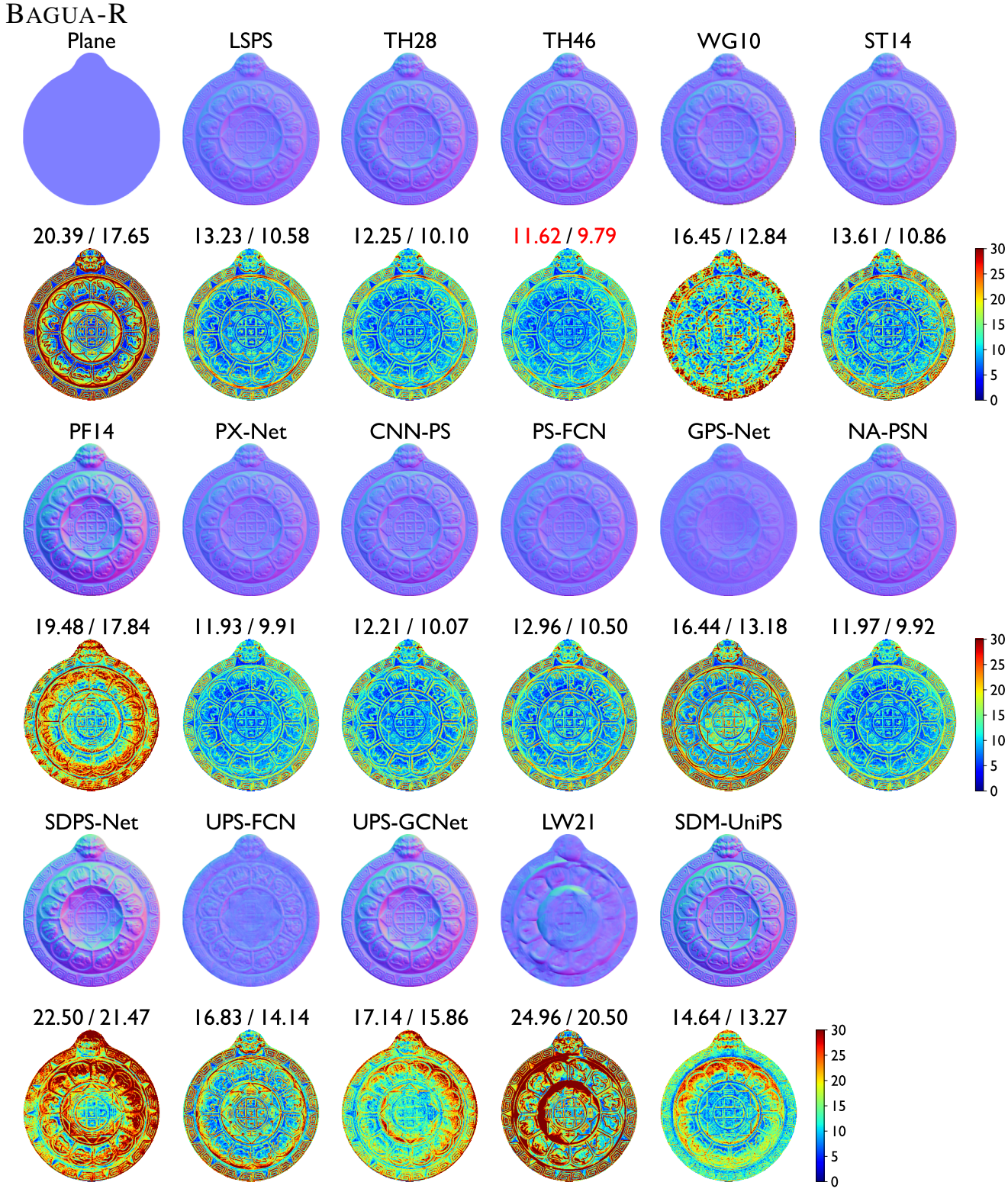


Figure S32: Estimated surface normal maps and angular error maps of BAGUA-R.

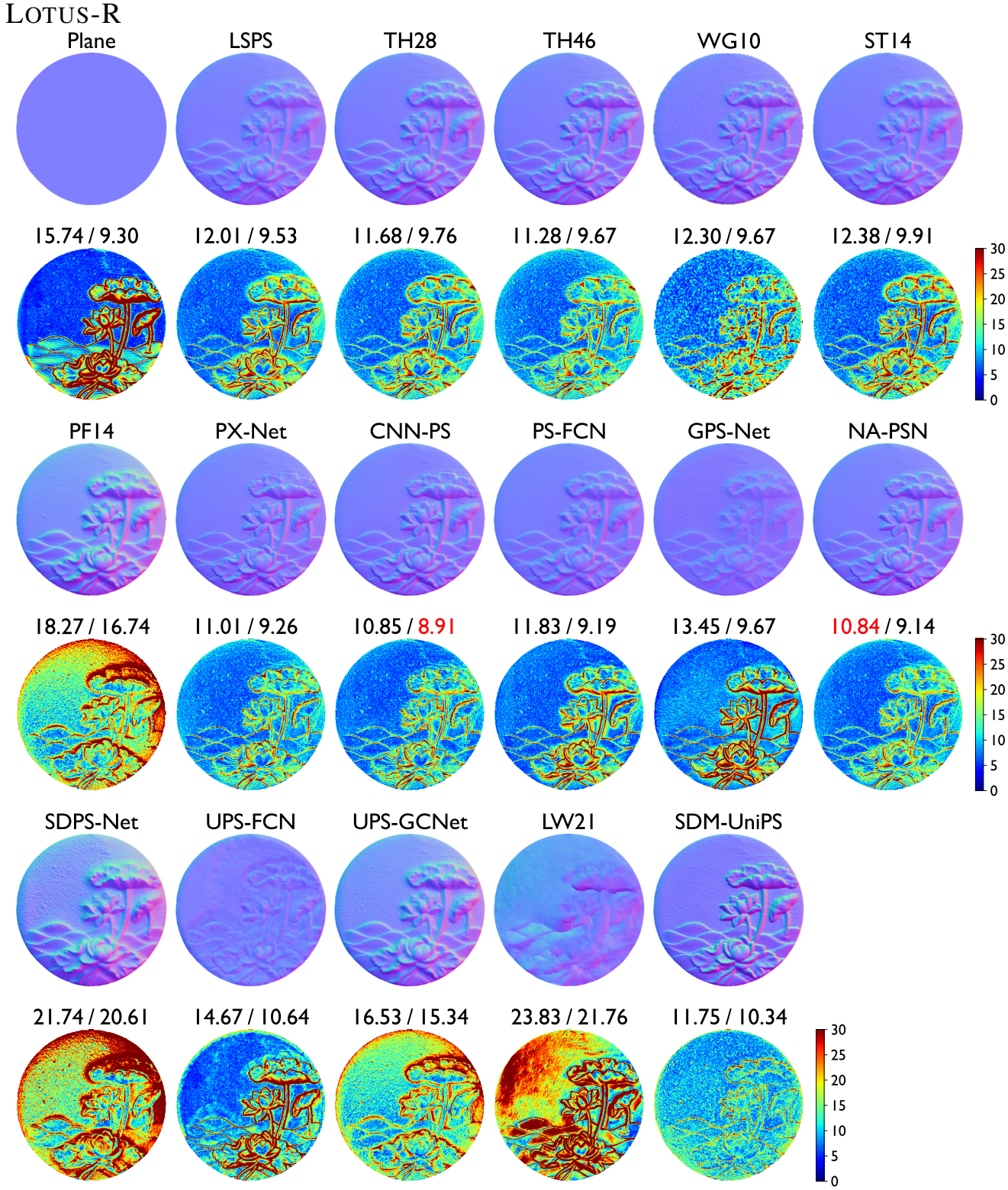


Figure S33: Estimated surface normal maps and angular error maps of LOTUS-R.

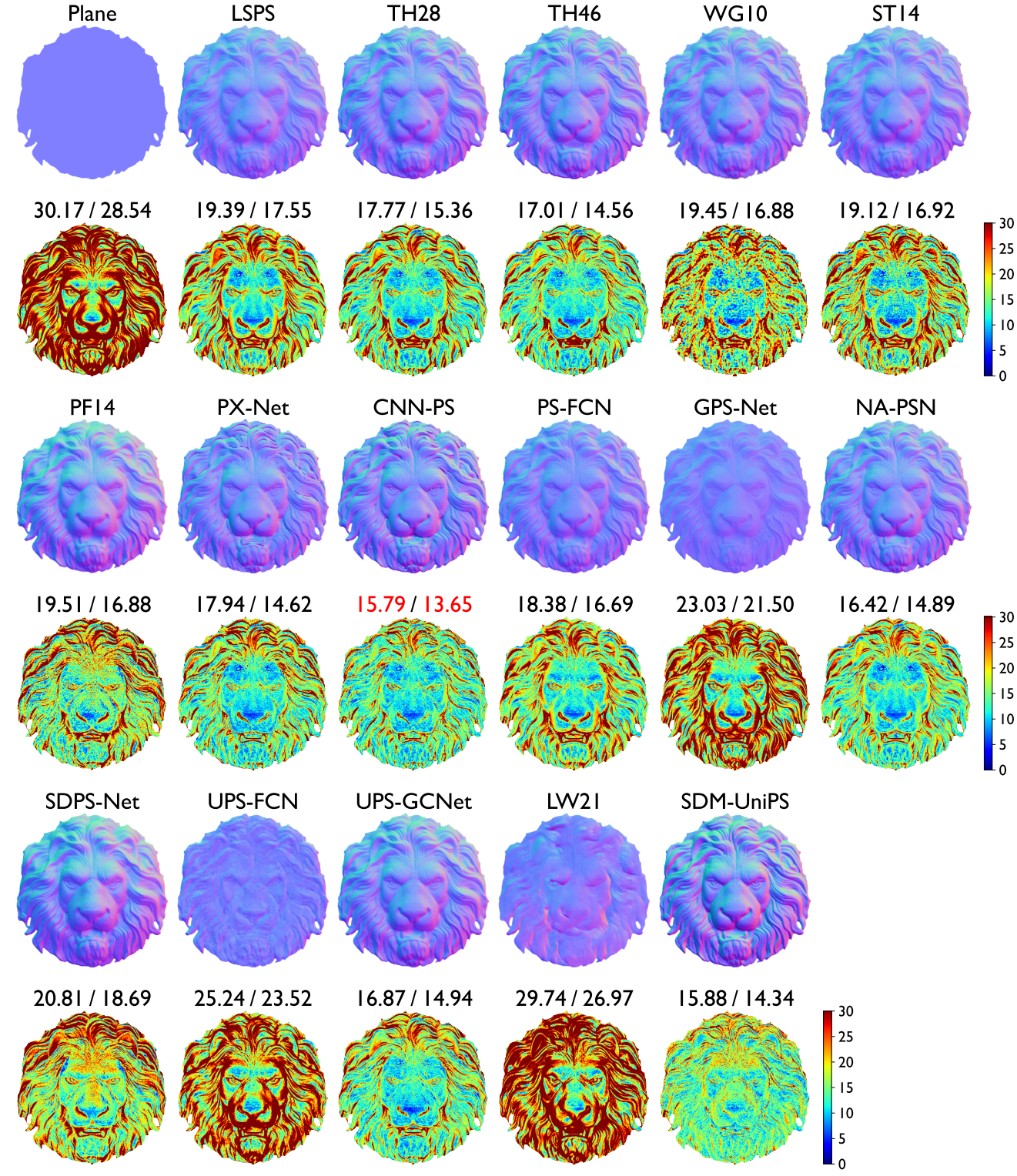


Figure S34: Estimated surface normal maps and angular error maps of LION-R.

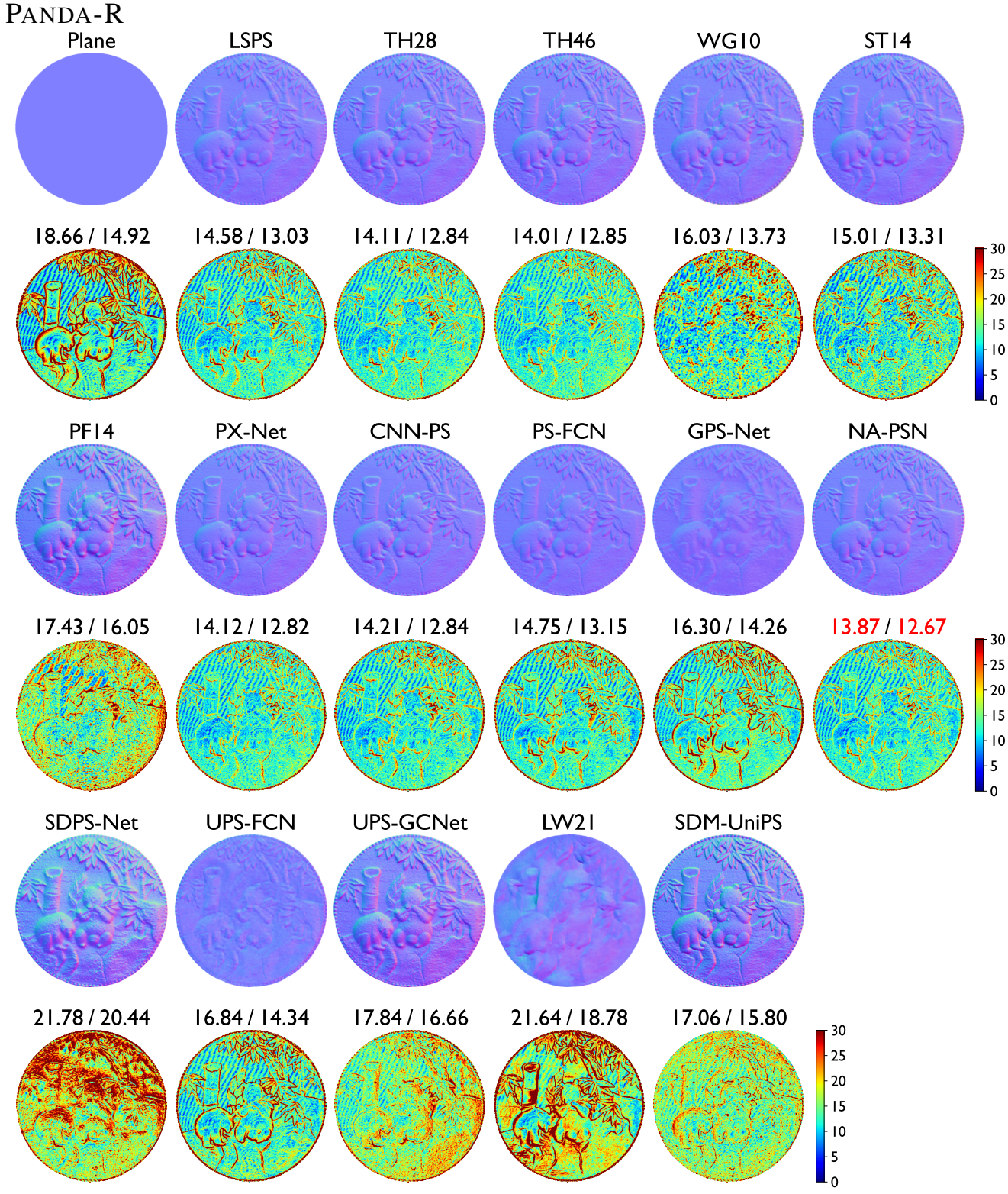


Figure S35: Estimated surface normal maps and angular error maps of PANDA-R.

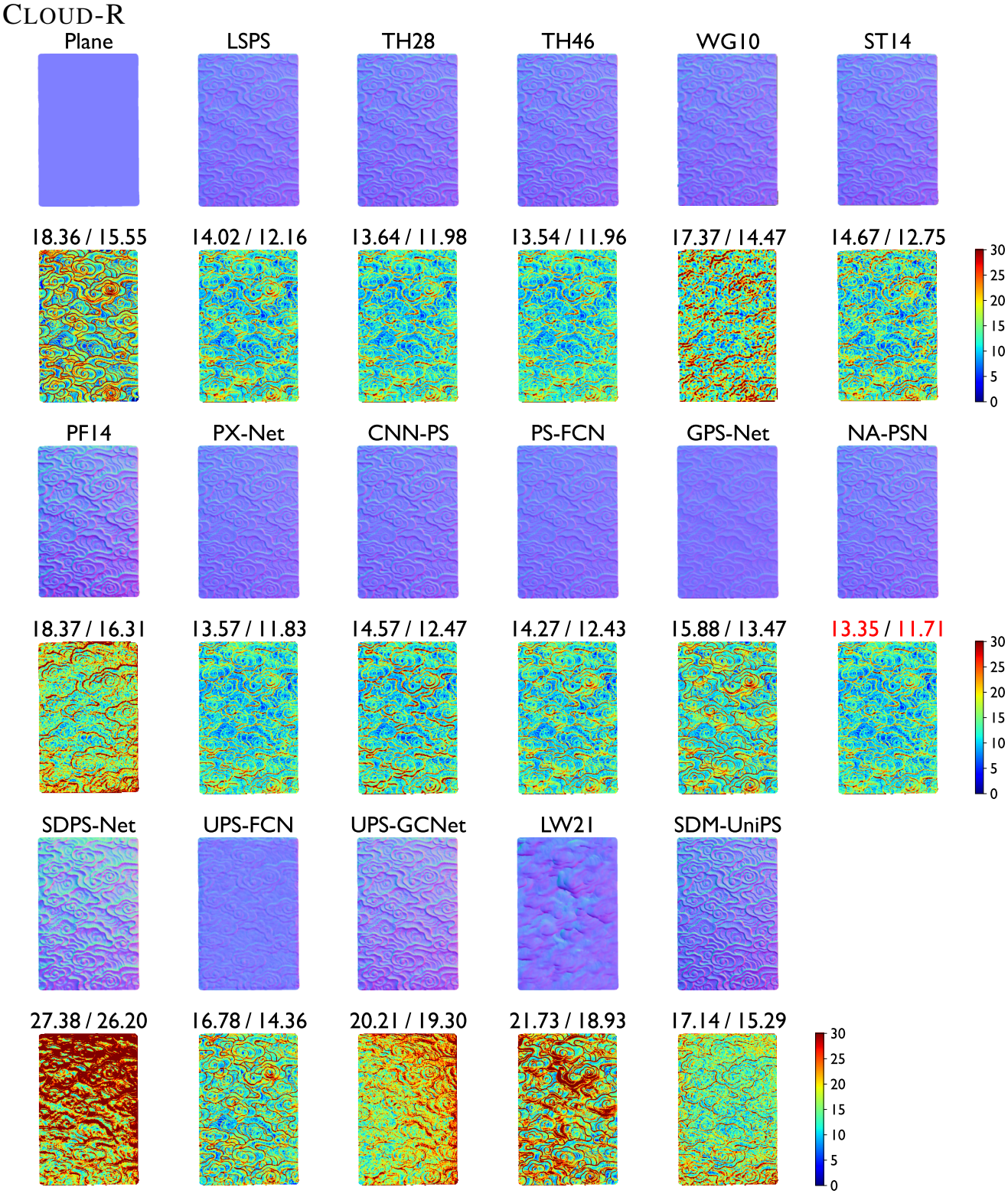


Figure S36: Estimated surface normal maps and angular error maps of CLOUD-R.

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