Supplementary Material for “DANI-Net: Uncalibrated Photometric Stereo by Differentiable Shadow Handling, Anisotropic Reflectance Modeling, and Neural Inverse Rendering”

Zongrui Li[1] Qian Zheng[2,3,*] Boxin Shi[4,5] Gang Pan[2,3] Xudong Jiang[1]
[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[2]The State Key Lab of Brain-Machine Intelligence, Zhejiang University, Hangzhou, China
[3]College of Computer Science and Technology, Zhejiang University, Hangzhou, China
[4]National Key Laboratory for Multimedia Information Processing, School of Computer Science, Peking University, Beijing, China
[5]National Engineering Research Center of Visual Technology, School of Computer Science, Peking University, Beijing, China
{zongrui001, EXDJiang}@ntu.edu.sg, {qianzheng,gpan}@zju.edu.cn, shiboxin@pku.edu.cn

In this supplementary material,

1. we provide more details about the network architectures in Sec. 6 (footnote 5 in the main paper);
2. we show the complete qualitative comparison of 10 objects from DiLiGenT [12] dataset between DANI-Net and other state-of-the-art methods [4, 6, 7] in Sec. 7 (footnote 9 in the main paper); we also conduct additional experiments to validate the effectiveness of our normal fitting method and visualize the generated svBRDF in this section (footnote 3 in the main paper);
3. we conduct more ablation studies about our three-stage training schema and silhouette loss in Sec. 8 (footnote 7 in the main paper);
4. we show the complete quantitative comparison of 100 objects (*i.e.*, 10 shapes multiplying 10 different materials) from DiLiGenT10^2 dataset [11] and visualization of the fitted material on 2 selected shapes (*i.e.*, Ball, Bunny) with 3 anisotropic materials (*i.e.*, Al, Cu, and Steel) in Sec. 9 (footnote 11 in the main paper); we also analyze the effects of different initialization methods in this section (footnote 12 in the main paper);
5. we show the qualitative and quantitative results on 3 objects from Apple & Gourd dataset [1] and 6 objects from Light Stage Data Gallery dataset [2] in Sec. 10 (footnote 8 in the main paper).

The code is available at `https://github.com/LMozart/CVPR2023-DANI-Net`.

---

[*] Corresponding author

6. Implementation Details

Positional encoding. We apply the same positional encoding module in NeRF [8][1] given the following equation,

$$E(p) = \left(\sin\left(2^0 \pi p\right), \cos\left(2^0 \pi p\right), \cdots, \sin\left(2^{L-1} \pi p\right), \cos\left(2^{L-1} \pi p\right)\right), \tag{10}$$

where L is the dimension of the positional code, p is 2D coordinate (u_i, v_i) in the image plane for DepthMLP and MaterialMLP. We compute the positional code $E(p)$ for each dimension and concatenate these codes to get our positional code.

Network structure. We show the details of the DepthMLP and MaterialMLP in Fig. 7. We use the same pre-trained light model as [7] for light initialization. More details of the pre-trained light model can be found in the supplementary materials of SCPS-NIR [7].

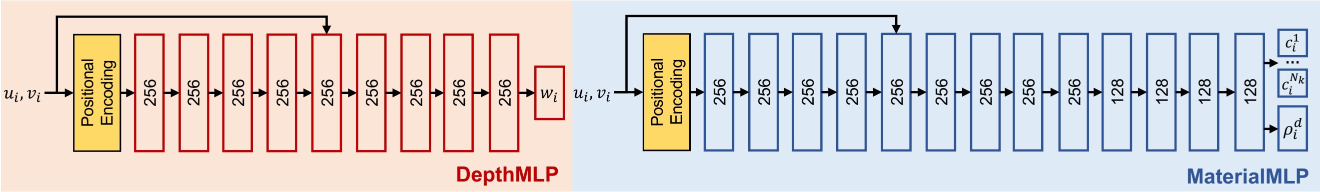


Figure 7. Network structures of DepthMLP and MaterialMLP.

---

[1] https://github.com/facebookresearch/pytorch3d/tree/main/projects/nerf

7. Additional Results on DiLiGenT Dataset [12]

Qualitative results on DiLiGenT dataset [12]. From Fig. 8 to Fig. 11 we show the normal map, error map, and light map of 10 objects from DiLiGenT Dataset [12] predicted by DANI-Net and its alternatives (*i.e.*, DANI-Net *w* [25] and DANI-Net *w/o s*). We further compare DANI-Net with recent UPS methods [4, 7] and state-of-the-art unsupervised PS methods [6]. The results illustrate the effectiveness of the differentiable shadow handling and anisotropic material modeling method in DANI-Net.

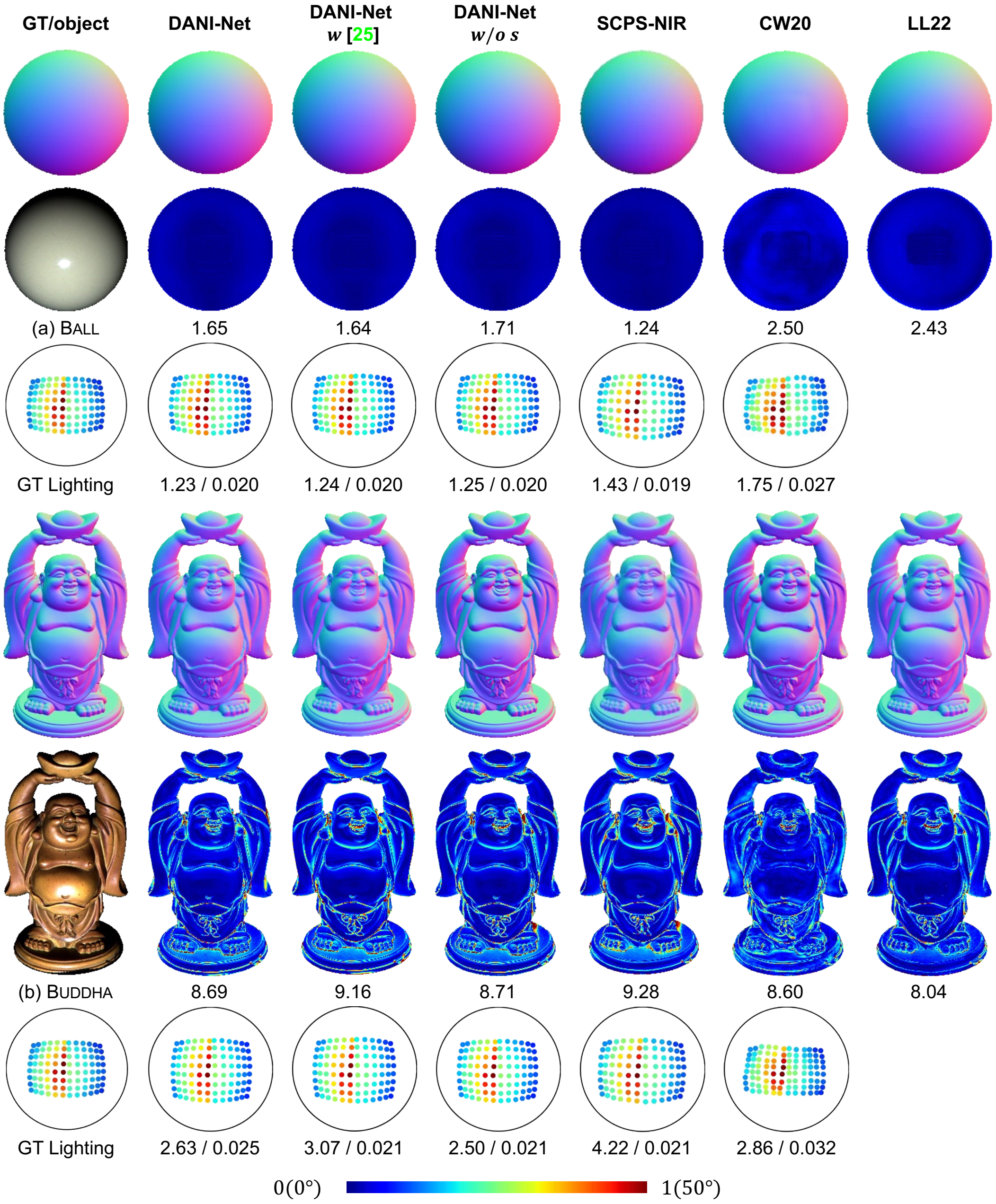


Figure 8. The visual quality comparison among DANI-Net, DANI-Net *w* [25], DANI-Net *w/o s*, SCPS-NIR [7], CW20 [4], and LL22 [6] on Ball and Buddha from DiLiGenT [12] in terms of normal map (row 1,4), error map (row 2, 5), and light map (row 3, 6). Numbers indicate the MAE (for surface normal or light directions) or scale-invariant relative error (for light intensity).

Figure 9. The visual quality comparison among DANI-Net, DANI-Net *w* [25], DANI-Net *w/o s*, SCPS-NIR [7], CW20 [4], and LL22 [6] on BEAR, CAT, and HARVEST from DILIGENT [12] in terms of normal map (row 1, 4, 7), error map (row 2, 5, 8), and light map (row 3, 6, 9). Numbers indicate the MAE (for surface normal or light directions) or scale-invariant relative error (for light intensity).

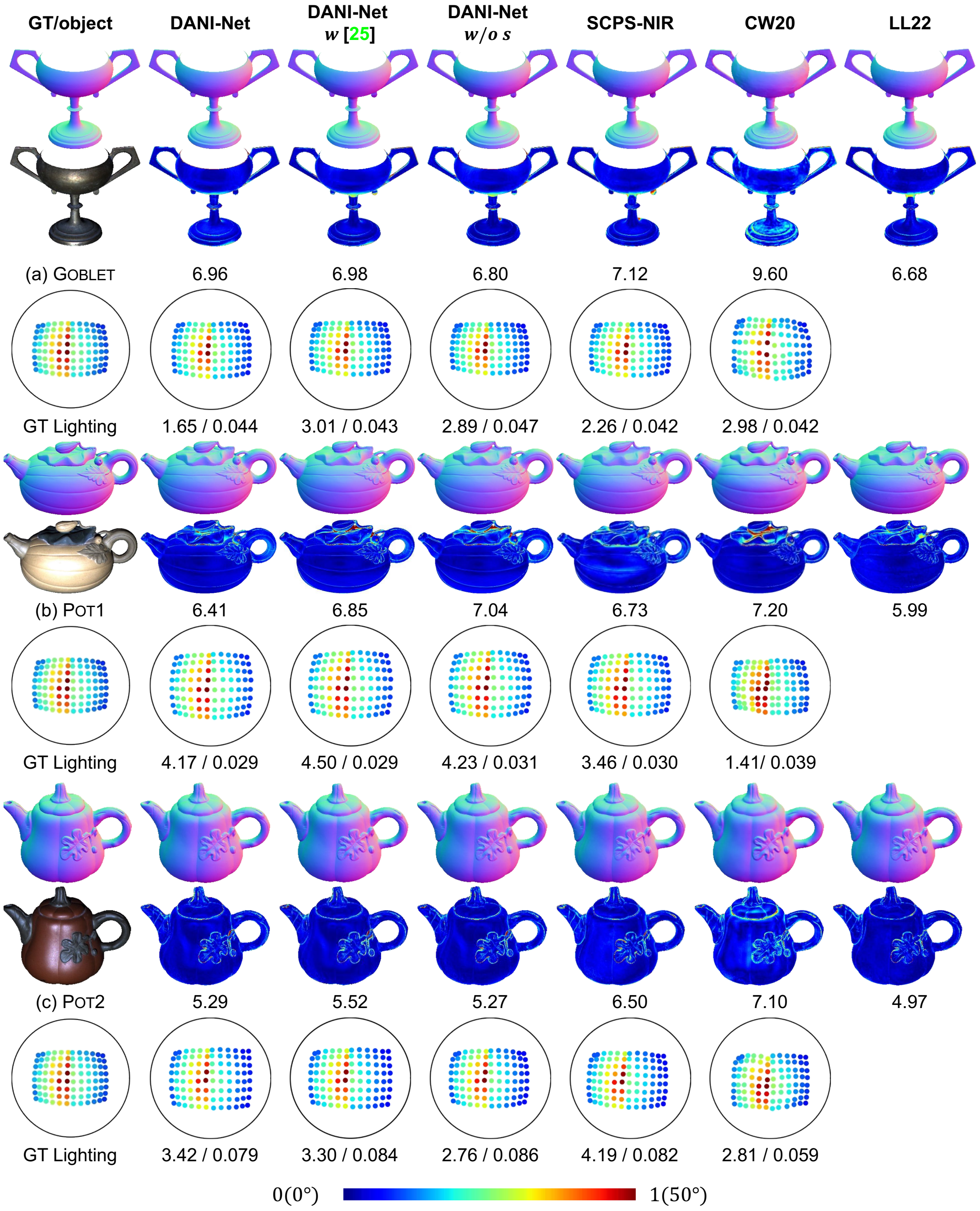


Figure 10. The visual quality comparison among DANI-Net, DANI-Net *w* [25], DANI-Net *w/o s*, SCPS-NIR [7], CW20 [4], and LL22 [6] on GOBLET, POT1, and POT2 from DILIGENT [12] in terms of normal map (row 1, 4, 7), error map (row 2, 5, 8), and light map (row 3, 6, 9). Numbers indicate the MAE (for surface normal or light directions) or scale-invariant relative error (for light intensity).

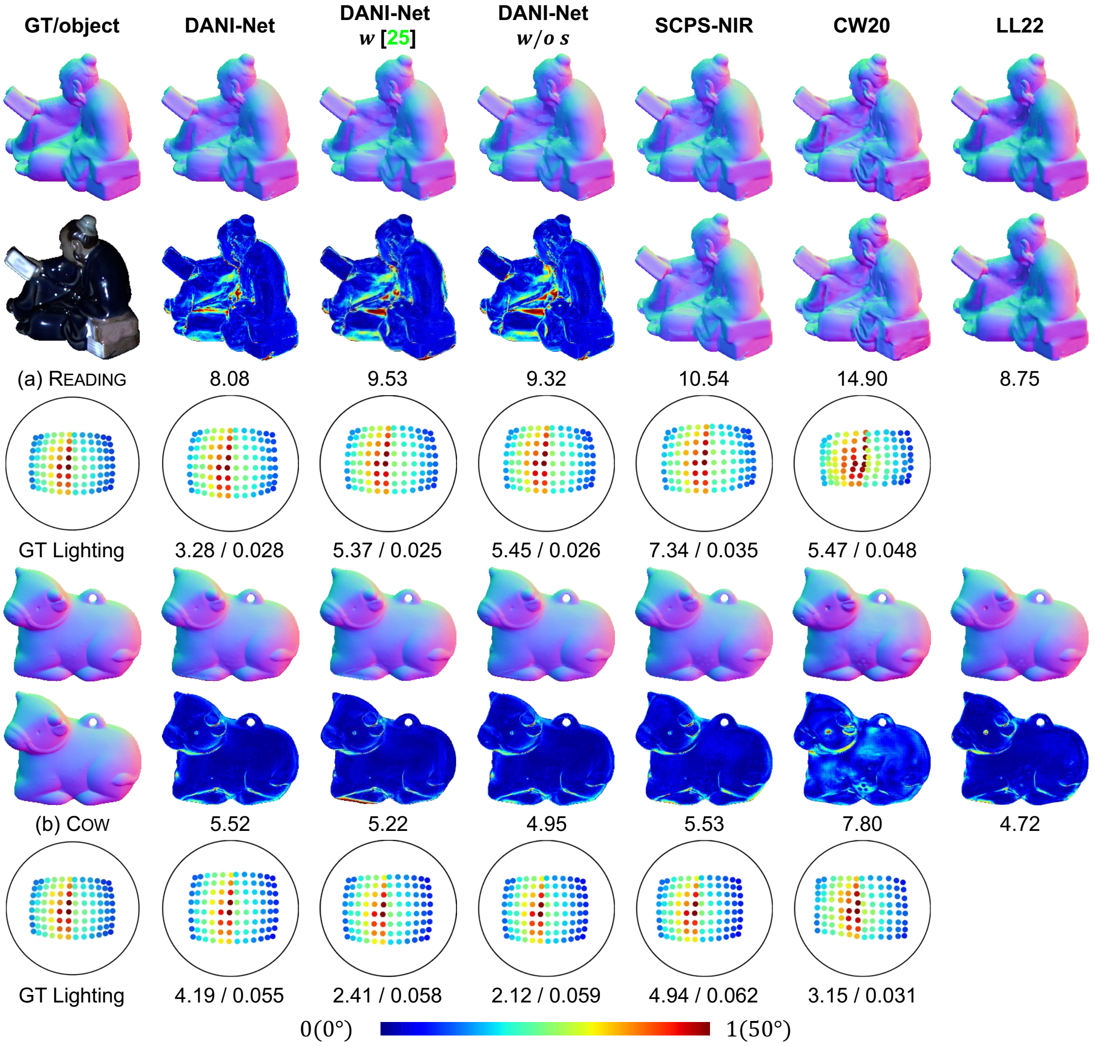


Figure 11. The visual quality comparison among DANI-Net, DANI-Net w [25], DANI-Net w/o s, SCPS-NIR [7], CW20 [4], and LL22 [6] on Reading and Cow from DiLiGenT [12] in terms of normal map (row 1,4), error map (row 2, 5), and light map (row 3, 6). Numbers indicate the MAE (for surface normal or light directions) or scale-invariant relative error (for light intensity).

Validation of weighted interpolation. To validate the effectiveness of our weighted interpolation for normal fitting, we compare DANI-Net with 'DANI-Net *w* sobel' (*i.e.*, Sobel operator), 'DANI-Net *w* cross' (*i.e.*, method in [6]), and 'DANI-Net *w* triangle' (*i.e.*, method in [9]). Those alternatives are only different in the normal fitting method. As shown in Table 6, DANI-Net outperforms all alternatives on average, especially on objects like Harvest with a complicated shape. We observe a similar performance between 'DANI-Net *w* sobel' and 'DANI-Net *w* cross', whose average MAE on normal estimation increase $0.15°$ and $0.09°$, respectively since they either ignore query point's information or using excessive neighbor points. The average MAE of 'DANI-Net *w* triangle' increase about $0.39°$ due to its ineffectiveness in backpropagation (*i.e.*, limit points affected by the gradients).

Table 6. Quantitative comparison in terms of MAE of surface normal on DiLiGenT benchmark dataset [12] for different normal fitting methods. **Bold numbers** indicates the best results.

Method	Ball	Bear	Buddha	Cat	Cow	Goblet	Harvest	Pot1	Pot2	Reading	AVG
DANI-Net *w* triangle	2.33	4.27	9.19	4.87	5.65	7.44	15.22	6.61	5.69	8.01	6.93
DANI-Net *w* sobel	1.89	**3.97**	9.05	4.71	**5.31**	7.27	14.85	6.72	5.42	7.68	6.69
DANI-Net *w* cross	1.88	3.98	9.08	**4.70**	5.36	**6.67**	14.93	6.63	5.43	**7.68**	6.63
DANI-Net	**1.65**	4.11	**8.69**	4.73	5.52	6.96	**13.99**	**6.41**	**5.29**	8.08	**6.54**

Validation of svBRDF. To validate the spatially varying BRDF (svBRDF) generated by DANI-Net, we visualize the BRDF spheres at different points of Harvest, Reading, and Pot1 in Fig. 12. The spheres show polychrome appearances with different roughness, demonstrating the visually pleasant svBRDF generated by DANI-Net.

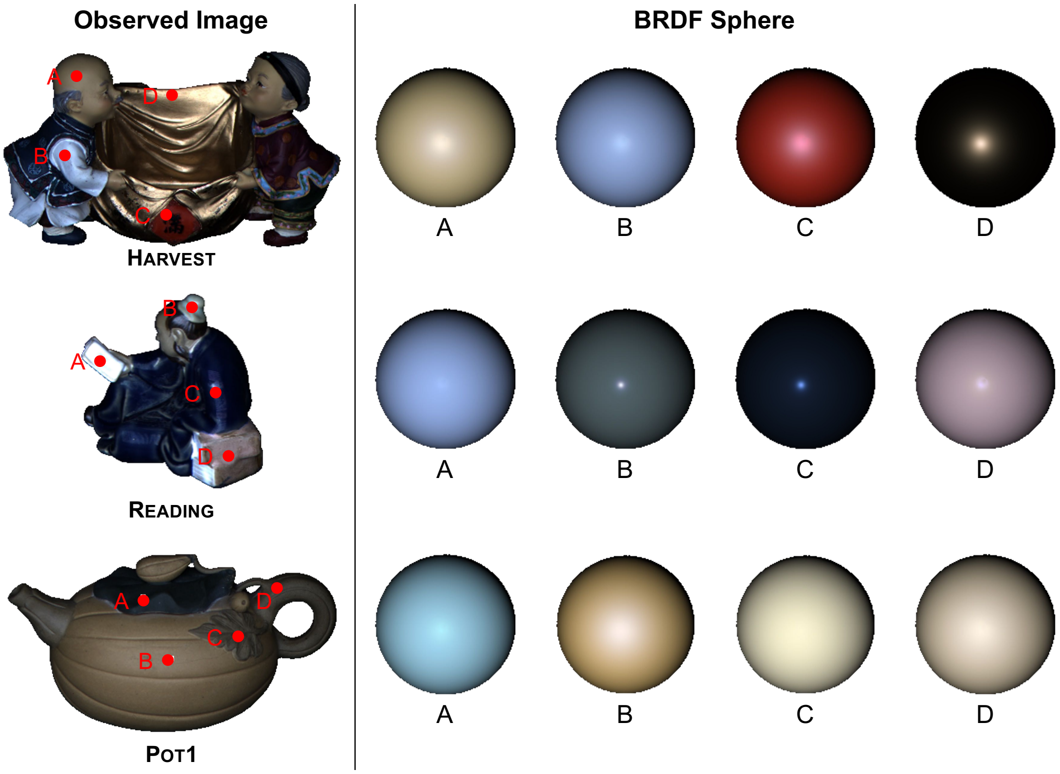


Figure 12. Visualization of the estimated svBRDF on Harvest, Reading, and Pot1 from DiLiGenT dataset [12]. On each object, we select four points (red dots on the first column's images, labeled ABCD) at different positions with different materials to showcase the predicted svBRDF spheres. The overall intensity of the observed images and BRDF spheres is scaled for better visualization.

8. Additional Ablation Studies

8.1. Three-stage Training Schema

The smoothness terms incorporated in the three-stage training schema (*i.e.*, $\mathcal{L}_{R^d}, \mathcal{L}_W$, and $\mathcal{L}_N$) facilitate DANI-Net's convergence at an early stage (stage 1); while they are gradually dropped in the subsequent stage (stage 2, 3) to avoid over-smooth results. To validate the effectiveness of the three-stage training schema, we compare DANI-Net with: 1) 'DANI-Net *w* $\mathcal{L}_N$' that keeps $\mathcal{L}_N$ till the training complete; 2) 'DANI-Net *w/o* $\mathcal{L}_{R^d}$' that drops material smoothness term (*i.e.*, $\mathcal{L}_{R^d}$) at the beginning on DiLiGenT dataset [12]; 3) 'DANI-Net *w/o* $\mathcal{L}_N + \mathcal{L}_W$' that drops geometry smoothness terms (*i.e.*, $\mathcal{L}_N$ and $\mathcal{L}_W$) at the beginning. As shown in Table 7, the average MAE of normal estimation on 10 objects increases $0.22°$, $0.28°$, and $0.10°$ for 'DANI-Net *w* $\mathcal{L}_N$', 'DANI-Net *w/o* $\mathcal{L}_{R^d}$', and 'DANI-Net *w/o* $\mathcal{L}_N + \mathcal{L}_W$', respectively, indicating the necessity of the three-stage training schema.

Table 7. Quantitative comparison in terms of MAE of surface normal on DiLiGenT benchmark dataset [12] for different alternatives on three-stage training schema. **Bold numbers** indicates the best results.

Method	Ball	Bear	Buddha	Cat	Cow	Goblet	Harvest	Pot1	Pot2	Reading	AVG
DANI-Net *w* $\mathcal{L}_N$	1.57	4.10	9.76	4.99	**5.16**	6.96	14.41	7.28	5.88	7.47	6.76
DANI-Net *w/o* $\mathcal{L}_{R^d}$	1.47	**4.05**	8.79	5.15	5.69	6.97	14.28	7.13	**5.24**	9.43	6.82
DANI-Net *w/o* $\mathcal{L}_N + \mathcal{L}_W$	**1.44**	4.28	**8.19**	4.98	5.92	**6.88**	15.29	**6.08**	5.28	**8.06**	6.64
DANI-Net	1.65	4.11	8.69	**4.73**	5.52	6.96	**13.99**	6.41	5.29	8.08	**6.54**

8.2. Silhouette Normal

Although the fitted silhouette normal is only reliable at the occluding boundaries, the silhouette loss $\mathcal{L}_{\text{Si}}$ is crucial in alleviating GBR ambiguity that facilitates DANI-Net's convergence and prevents it from falling into the local optimum. We compare DANI-Net with 'DANI-Net *w/o* $\mathcal{L}_{\text{Si}}$' (or $\mathcal{L}_{\text{Si}}$) that drops $\mathcal{L}_{\text{Si}}$ at the beginning (or at stage 3) on DiLiGenT dataset [12]. As shown in Table 8, the average MAE of normal estimation on 10 objects increases $0.52°$ (or $0.14°$), indicating the necessity of silhouette loss.

For objects in Light Stage Data Gallery [2] and DiLiGenT 10^2 [11] with non-occluding silhouette, DANI-Net uses silhouette loss in a more flexible way. That is, we either drop the silhouette loss at stage 3 (Knight Kneeling, Knight Fighting, and Knight Standing from Light Stage Data Gallery [2]), or roughly estimate the silhouette's normal instead of pre-computing it through silhouette's perpendicular vectors (Bunny, Hexagon, Nut, Pentagon, Propeller, Square, and Turbine from DiLiGenT 10^2 [11])[2]. To clarify the necessity of this strategy, we visualize the estimated normal map of DANI-Net and 'DANI-Net *w* $\mathcal{L}_{\text{Si}}$' that keeps $\mathcal{L}_{\text{Si}}$ in all stages on Knight Fighting with partially non-occluding silhouettes on the sword (the top row of Fig. 13), and Nut Pp with predominantly non-occluding silhouettes on its base (the bottom row of Fig. 13). It can be observed that keeping $\mathcal{L}_{\text{Si}}$ in all stages does not significantly affect the normal estimation results for objects with partially occluding silhouettes like Knight Fighting, but it does harm the performance for objects with predominantly non-occluding silhouettes (MAE on normal estimation increases $1.16°$ on Nut Pp). This validates the effectiveness of our flexible strategy using $\mathcal{L}_{\text{Si}}$ on objects with non-occluding silhouettes.

Table 8. Quantitative comparison in terms of MAE of surface normal on DiLiGenT benchmark dataset [12] for different alternatives on silhouette loss. **Bold numbers** indicates the best results.

Method	Ball	Bear	Buddha	Cat	Cow	Goblet	Harvest	Pot1	Pot2	Reading	AVG
DANI-Net *w/o* $\mathcal{L}_{\text{Si}}$	1.68	4.22	9.44	5.30	5.79	8.12	14.18	7.43	5.53	8.89	7.06
DANI-Net *w/o* $\mathcal{L}_{\text{Si}}$-3	1.81	4.07	8.72	5.15	5.69	7.10	13.77	6.95	5.26	8.30	6.68
DANI-Net	**1.65**	**4.11**	**8.69**	**4.73**	**5.52**	**6.96**	**13.99**	**6.41**	**5.29**	**8.08**	**6.54**

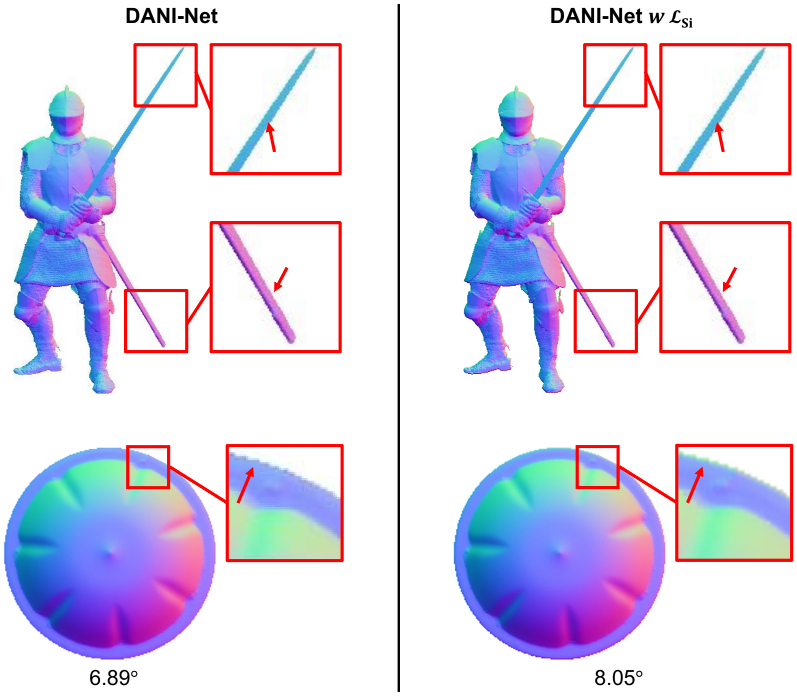


Figure 13. Visualization of the estimated normal map on Knight Fighting (top), and Nut Pom (bottom) from Light Stage Data Gallery [2] and DiLiGenT10^2 [11]. Sub-figures in the left (or right) column are normal maps estimated by DANI-Net (or 'DANI-Net *w* $\mathcal{L}_{\text{Si}}$') Red boxes highlight prominent regions for easy comparison. Red arrows point to the silhouette of objects. Numbers at the bottom indicate the MAE of normal estimation.

---

[2] It is easy to estimate the silhouette normal of those objects as $[0, 0, 1]$. We apply the same measurement in SCPS-NIR [7] to calculate the silhouette normal loss for a fair comparison.

9. Additional Results on DiLiGenT10^2 Dataset [11]

Quantitative results of normal estimation. Fig. 14 presents the comprehensive normal estimation results compared with CW20 [4], SCPS-NIR [7], and CNN-PS [5]. While DANI-Net outperforms current methods on Ball and Bunny or anisotropic group (*i.e.*, Al, Cu, and Steel), the average MAE of DANI-Net on 100 objects' normal estimation is slightly higher than CNN-PS (still lower than current UPS methods). Particularly, we find that DANI-Net is not competitive on objects of Nylon and Acrylic, and objects like Turbine and Propeller with complicated shapes in the 'anisotropic group'.

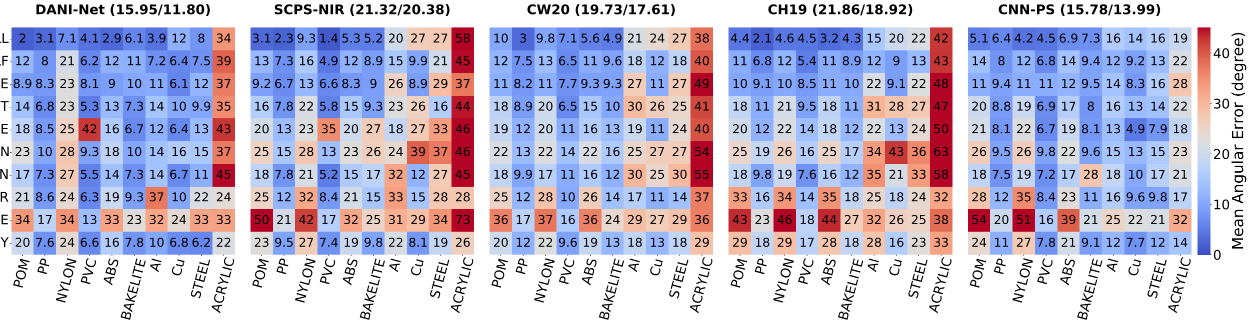


Figure 14. Normal estimation results on DiLiGenT10^2 Dataset [11] visualized by Shape-material error matrix [11]. Numbers in the bracket indicate each method's mean/median MAE for normal estimation.

Quantitative results of light calibration. Fig. 15 and Fig. 16 present the light calibration results, compared with CW20 [4], SCPS [3], and SCPS-NIR [7]. As can be observed in those figures, DANI-Net reaches the second-best performance on light direction calibration and the best on light intensity calibration. By cross-comparing the results of normal estimation, we find that the cases with high normal estimation MAE also result in particularly high MAE in light calibration.

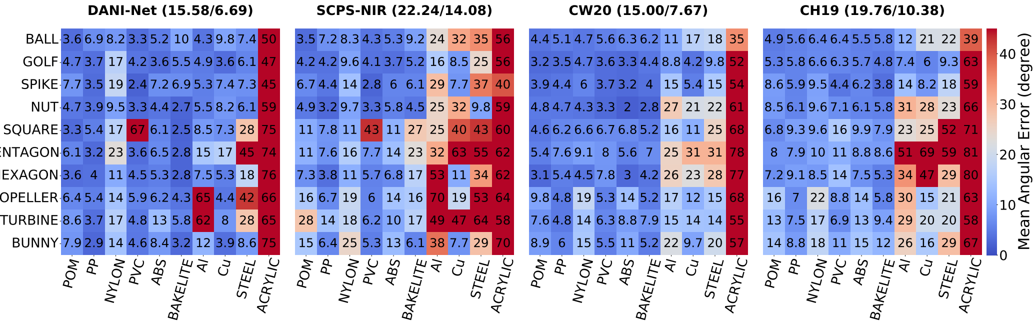


Figure 15. Light direction calibration results on DiLiGenT10^2 Dataset [11] visualized by Shape-material error matrix [11]. Numbers in the bracket indicate each method's mean/median MAE for light direction calibration.

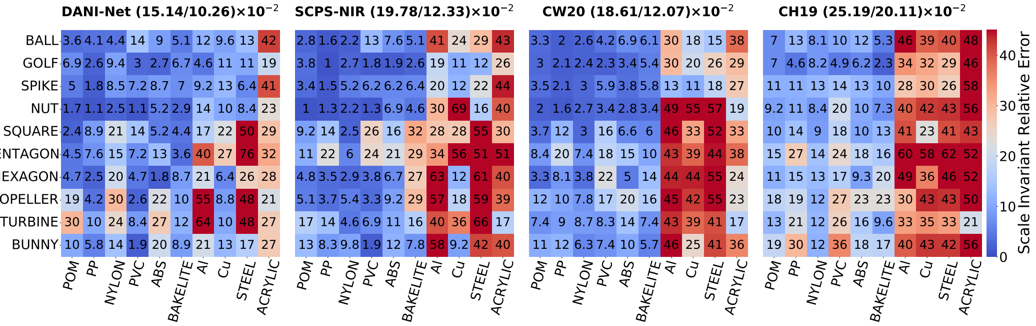
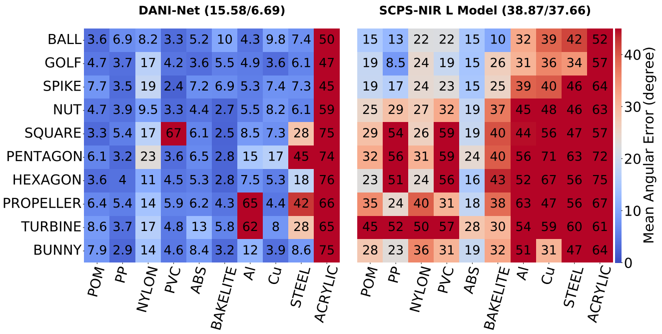


Figure 16. Light intensity calibration results on DiLiGenT10^2 Dataset [11] visualized by Shape-material error matrix [11]. The number in the matrix is scaled for visualization. Numbers in the bracket indicate each method's mean/median scale invariant relative error for light intensity calibration.

Analysis on light initialization. We first conjecture that inadequate light initialization is the main cause of inaccurate normal estimation. To validate that, we showcase the initial light's (calibrated by the pre-trained light model [7]) MAE in Fig. 17.



DANI-Net (15.58/6.69)

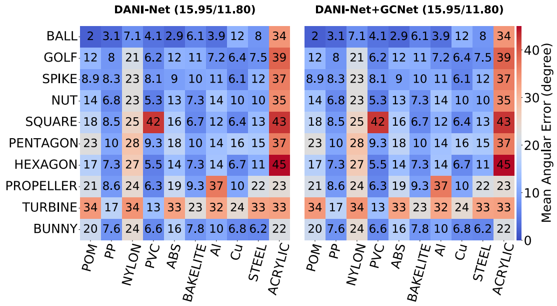
	POM	PP	NYLON	PVC	ABS	BAKELITE	Al	Cu	STEEL	ACRYLIC
BALL	3.6	6.9	8.2	3.3	5.2	10	4.3	9.8	7.4	50
GOLF	4.7	3.7	17	4.2	3.6	5.5	4.9	3.6	6.1	47
SPIKE	7.7	3.5	19	2.4	7.2	6.9	5.3	7.4	7.3	45
NUT	4.7	3.9	9.5	3.3	4.4	2.7	5.5	8.2	6.1	59
SQUARE	3.3	5.4	17	67	6.1	2.5	8.5	7.3	28	75
PENTAGON	6.1	3.2	23	3.6	6.5	2.8	15	17	45	74
HEXAGON	3.6	4	11	4.5	5.3	2.8	7.5	5.3	18	76
PROPELLER	6.4	5.4	14	5.9	6.2	4.3	65	4.4	42	66
TURBINE	8.6	3.7	17	4.8	13	5.8	62	8	28	65
BUNNY	7.9	2.9	14	4.6	8.4	3.2	12	3.9	8.6	75

SCPS-NIR L Model (38.87/37.66)

	POM	PP	NYLON	PVC	ABS	BAKELITE	Al	Cu	STEEL	ACRYLIC
BALL	15	13	22	22	15	10	32	39	42	52
GOLF	19	8.5	24	19	15	26	31	36	34	57
SPIKE	19	17	24	23	15	25	39	40	46	64
NUT	25	29	27	32	19	37	45	48	46	63
SQUARE	29	54	26	59	19	40	44	56	47	57
PENTAGON	32	56	31	59	24	40	56	71	63	72
HEXAGON	23	51	24	56	16	43	52	67	56	75
PROPELLER	35	24	40	31	18	38	63	47	56	67
TURBINE	45	52	50	57	28	30	54	59	60	61
BUNNY	28	23	36	31	19	32	51	31	47	64

Figure 17. Comparison between DANI-Net and the pre-trained light model of SCPS-NIR [7] on light calibration. Numbers in the bracket indicate each method's mean/median MAE for light direction calibration.

The results validate our conjecture, *i.e.*, the initial lights deviate significantly from the ground truth in the aforementioned cases. Therefore, initializing the light more accurately will be a reasonable strategy (*e.g.*, light calibrated by GCNet [4]). We compare DANI-Net with 'DANI-Net+GCNet', with the latter differing only in its light initialized by GCNet. As shown in Fig. 18, the average MAE of normal estimation (or light calibration) of 'DANI-Net+GCNet' decreases 0.77° (or 3.06°). Particularly, we observe a compelling enhancement for objects with complicated shapes (*i.e.*, PROPELLER, TURBINE), but a marginal improvement for objects made of NYLON and ACRYLIC. While objects made of ACRYLIC still have inaccurate light initialization leading to unsatisfactory light calibration and normal estimation[3], the results obtained for objects made of NYLON contradict our earlier conjecture. We further assume that the materials' exceptional properties may be the primary cause of the subpar performance.



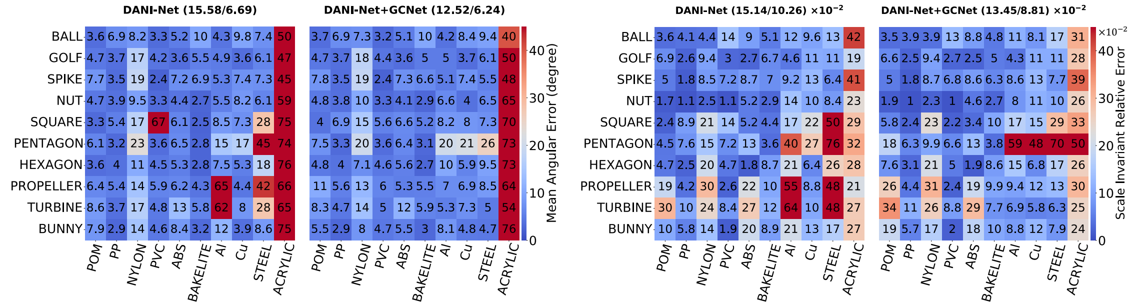
DANI-Net (15.95/11.80)

	POM	PP	NYLON	PVC	ABS	BAKELITE	Al	Cu	STEEL	ACRYLIC
BALL	2	3.1	7.1	4.1	2.9	6.1	3.9	12	8	34
GOLF	12	8	21	6.2	12	11	7.2	6.4	7.5	39
SPIKE	8.9	8.3	23	8.1	9	10	11	6.1	12	37
NUT	14	6.8	23	5.3	13	7.3	14	10	10	35
SQUARE	18	8.5	25	42	16	6.7	12	6.4	13	43
PENTAGON	23	10	28	9.3	18	10	14	16	15	37
HEXAGON	17	7.3	27	5.5	14	7.3	14	6.7	11	45
PROPELLER	21	8.6	24	6.3	19	9.3	37	10	22	23
TURBINE	34	17	34	13	33	23	32	24	33	33
BUNNY	20	7.6	24	6.6	16	7.8	10	6.8	6.2	22

DANI-Net+GCNet (15.95/11.80)

	POM	PP	NYLON	PVC	ABS	BAKELITE	Al	Cu	STEEL	ACRYLIC
BALL	2	3.1	7.1	4.1	2.9	6.1	3.9	12	8	34
GOLF	12	8	21	6.2	12	11	7.2	6.4	7.5	39
SPIKE	8.9	8.3	23	8.1	9	10	11	6.1	12	37
NUT	14	6.8	23	5.3	13	7.3	14	10	10	35
SQUARE	18	8.5	25	42	16	6.7	12	6.4	13	43
PENTAGON	23	10	28	9.3	18	10	14	16	15	37
HEXAGON	17	7.3	27	5.5	14	7.3	14	6.7	11	45
PROPELLER	21	8.6	24	6.3	19	9.3	37	10	22	23
TURBINE	34	17	34	13	33	23	32	24	33	33
BUNNY	20	7.6	24	6.6	16	7.8	10	6.8	6.2	22

(a) Normal Estimation



DANI-Net (15.58/6.69)

	POM	PP	NYLON	PVC	ABS	BAKELITE	Al	Cu	STEEL	ACRYLIC
BALL	3.6	6.9	8.2	3.3	5.2	10	4.3	9.8	7.4	50
GOLF	4.7	3.7	17	4.2	3.6	5.5	4.9	3.6	6.1	47
SPIKE	7.7	3.5	19	2.4	7.2	6.9	5.3	7.4	7.3	45
NUT	4.7	3.9	9.5	3.3	4.4	2.7	5.5	8.2	6.1	59
SQUARE	3.3	5.4	17	67	6.1	2.5	8.5	7.3	28	75
PENTAGON	6.1	3.2	23	3.6	6.5	2.8	15	17	45	74
HEXAGON	3.6	4	11	4.5	5.3	2.8	7.5	5.3	18	76
PROPELLER	6.4	5.4	14	5.9	6.2	4.3	65	4.4	42	66
TURBINE	8.6	3.7	17	4.8	13	5.8	62	8	28	65
BUNNY	7.9	2.9	14	4.6	8.4	3.2	12	3.9	8.6	75

DANI-Net+GCNet (12.52/6.24)

	POM	PP	NYLON	PVC	ABS	BAKELITE	Al	Cu	STEEL	ACRYLIC
BALL	3.7	6.9	7.3	3.2	5.1	10	4.2	8.4	9.4	40
GOLF	4.7	3.7	18	4.4	3.6	5	5	3.7	6.1	50
SPIKE	7.8	3.5	19	2.4	7.3	6.6	5.1	7.4	5.5	48
NUT	4.8	3.8	10	3.3	4.1	2.9	6.6	4	6.5	65
SQUARE	4	6.9	15	5.6	6.6	5.2	8.2	8	7.3	70
PENTAGON	7.5	3.3	20	3.6	6.4	3.1	20	21	26	73
HEXAGON	4.8	4	7.1	4.6	5.6	2.7	10	5.9	9.5	73
PROPELLER	11	5.6	13	6	5.3	5.5	7	6.9	8.5	64
TURBINE	8.3	4.7	14	5	12	5.9	5.3	7.3	5	54
BUNNY	5.5	2.9	8	4.7	5.5	3	8.1	4.8	4.7	76

DANI-Net (15.14/10.26) ×10⁻²

	POM	PP	NYLON	PVC	ABS	BAKELITE	Al	Cu	STEEL	ACRYLIC
BALL	3.6	4.1	4.4	14	9	5.1	12	9.6	13	42
GOLF	6.9	2.6	9.4	3	2.7	6.7	4.6	11	11	19
SPIKE	5	1.8	8.5	7.2	8.7	7	9.2	13	6.4	41
NUT	1.7	1.1	2.5	1.1	5.2	2.9	14	10	8.4	23
SQUARE	2.4	8.9	21	14	5.2	4.4	17	22	50	29
PENTAGON	4.5	7.6	15	7.2	13	3.6	40	27	76	32
HEXAGON	4.7	2.5	20	4.7	1.8	8.7	21	6.4	26	28
PROPELLER	19	4.2	30	2.6	22	10	55	8.8	48	21
TURBINE	30	10	24	8.4	27	12	64	10	48	27
BUNNY	10	5.8	14	1.9	20	8.9	21	13	17	27

DANI-Net+GCNet (13.45/8.81) ×10⁻²

	POM	PP	NYLON	PVC	ABS	BAKELITE	Al	Cu	STEEL	ACRYLIC
BALL	3.5	3.9	3.9	13	8.8	4.8	11	8.1	17	31
GOLF	6.6	2.5	9.4	2.7	2.5	5	4.3	11	11	28
SPIKE	5	1.8	8.7	6.8	8.6	6.3	8.6	13	7.7	39
NUT	1.9	1	2.3	1	4.6	2.7	8	11	10	26
SQUARE	5.8	2.4	23	2.2	3.4	10	17	15	29	33
PENTAGON	18	6.3	9.9	6.6	13	3.8	59	48	70	50
HEXAGON	7.6	3.1	21	5	1.9	8.6	15	6.8	17	26
PROPELLER	26	4.4	31	2.4	19	9.9	9.4	12	13	30
TURBINE	34	11	26	8.8	29	7.7	6.9	5.8	6.3	25
BUNNY	19	5.7	17	2	18	10	8.8	12	7.9	24

(b) Light Direction Calibration

(c) Light Intensity Calibration

Figure 18. Comparison between DANI-Net and DANI-Net+GCNet on normal estimation and light calibration. All of the results are visualized by the Shape-material error matrix [11]. Numbers in the top and bottom left brackets indicate each method's mean/median MAE for light direction calibration. Numbers in the bottom right bracket indicate each method's mean/median scale invariant relative error for light intensity calibration.

---

[3] This may mainly be caused by the sub-surface scattering of translucent materials.

Analysis on material properties. We check the observed images of Nylon Ball (shown in the right of Fig. 19) and find that some images are contaminated by unexpected artifacts (see the red square in the right sub-figure) regarded as the noise that harms DANI-Net's performance. A naive trick for noise removal is to exclude the pixel points with too high or too low intensity for a smooth intensity profile. However, since the specularity contains solid reflectance cues to solve UPS, we only remove the low-intensity pixel points (*i.e.*, the point with an intensity that is lower than the 25th percentile of overall pixels' intensities). We denote DANI-Net implementing this trick as 'DANI-Net *w rm*'. According to Table 9, we find this simple trick further improves DANI-Net's performance on Nylon and even remains effective on Acrylic objects because the sub-surface scattering of translucent materials can also be regarded as noise, shown in the left of Fig. 19. The MAE of the surface normal for Nylon (or Acrylic) objects decreases $5.18°$ (or $2.24°$). However, note that this trick cannot explicitly exclude the noise, we regard a robust method of modeling these materials as part of our future work.

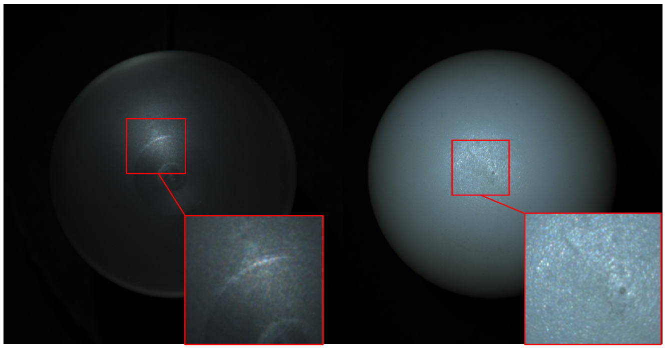


Figure 19. The observed image of Acrylic Ball (left) and Nylon Ball (right) from DiLiGenT10^2 Dataset [11]. The red squares indicate the regions that may harm the performance of DANI-Net.

Table 9. Quantitative comparison in terms of MAE of surface normal on DiLiGenT10^2 benchmark dataset [12]. **Bold numbers** indicates the best results.

Material	Method	Ball	Golf	Spike	Nut	Square	Pentagon	Hexagon	Propeller	Turbine	Bunny	AVG
Nylon	DANI-Net	7.06	20.98	22.63	22.79	25.33	28.3	26.96	24.00	**34.37**	23.82	23.62
	DANI-Net *w rm*	**5.23**	**13.40**	**9.61**	**16.22**	**20.43**	**24.10**	**19.18**	**22.62**	34.43	**19.17**	**18.44**
Acrylic	DANI-Net	**34.14**	39.13	36.63	35.25	42.70	36.90	**44.67**	**23.50**	33.03	21.89	34.78
	DANI-Net *w rm*	35.27	**25.50**	**35.10**	**33.38**	**36.98**	**36.16**	45.61	25.23	**32.90**	**19.27**	**32.54**

Validation of anisotropic material modeling. To validate the effectiveness of our anisotropic material modeling method, we visualize the rendered image, and 4 dominant ASG bases of 2 shapes (*i.e.*, BALL and BUNNY) made up of 3 anisotropic materials (*i.e.*, STEEL, CU, and AL respectively), compared with SCPS-NIR [7]'s rendered image and 4 dominant SG bases in Fig. 20. According to the result, our rendered images are more realistic than [7].

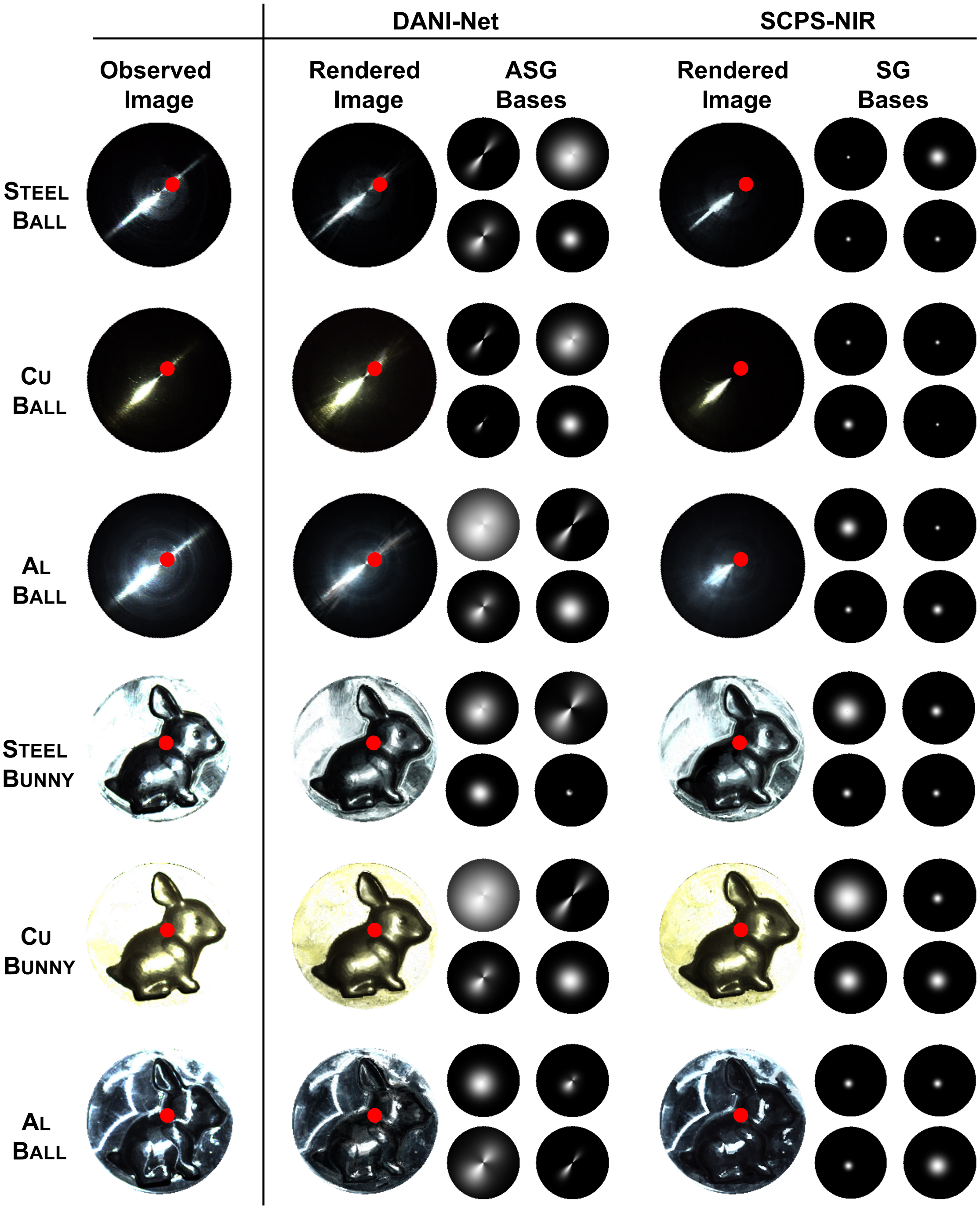


Figure 20. Visualization of the observed images (column 1), rendered images (column 2, 5), ASG bases (column 3, 4) and SG bases (column 6, 7) on BALL and BUNNY made up of STEEL, CU, and AL. The rendered images with anisotropic material modeling are generated by DANI-Net (column 2-4). The rendered images with isotropic material modeling are generated by SCPS-NIR [7] (column 5-7).

10. Results on APPLE & GOURD Dataset [1] and LIGHT STAGE DATA GALLERY dataset [2]

Preprocessing. For APPLE & GOURD, we scale the images into 0.5 of their original size for efficient training. For LIGHT STAGE DATA GALLERY, we use a preprocessing method that is different from [4, 7]. We follow [2] to implement gamma correction with an exponent as 2.2 on objects with LDR observed images (*i.e.* objects expect HELMET FRONT) and magnify the pixel intensity 5 times on objects except PLANT because the original pixel intensities on those objects are too low, leading to strong artifacts on the predicted normal map. We also downsample the PLANT's observed images to 0.5 of their original size for efficient training. The above preprocessing process is also applied in [7] for a fair comparison.

Qualitative and Quantitative Results. Table 10 shows the light calibration results of 3 objects from APPLE & GOURD Dataset [1] and 6 objects from LIGHT STAGE DATA GALLERY dataset [2]. As these two datasets don't release the ground truth normal map, we only provide the predicted normal map in Fig. 21 and Fig. 22 for qualitative analysis. The results show that DANI-Net can not only estimate the surface normal reasonably but also calibrate the light accurately (we have the state-of-the-art light calibration results), illustrating that DANI-Net is free from the deviation of different experimental setups.

Table 10. Quantitative comparison in terms of MAE of light direction and scale-invariant error of intensity on DILIGENT benchmark dataset [12]. **Bold numbers** and underlined numbers indicate the best and the second-best results, respectively.

Model	APPLE		GOURD1		GOURD2		PLANT		HELMET LEFT		HELMET FRONT		KNIGHT KNEELING		KNIGHT STANDING		KNIGHT FIGHTING		AVG	
	dir.	int.	dir.	int.	dir.	int.	dir.	int.	dir.	int.	dir.	int.	dir.	int.	dir.	int.	dir.	int.	dir.	int.
PF14 [10]	6.68	0.109	21.23	0.096	25.87	0.329	20.56	0.227	25.40	0.576	81.60	0.133	46.69	9.805	33.81	1.311	69.50	1.137	36.82	1.52
CW20 [4]	10.91	0.094	4.29	0.042	7.13	**0.199**	10.49	0.154	5.33	**0.096**	6.22	0.183	14.41	**0.181**	**5.31**	0.198	13.42	0.168	8.61	0.146
SCPS-NIR [7]	**1.87**	0.016	2.34	**0.027**	2.01	0.233	10.52	0.150	3.96	0.123	**3.31**	0.138	6.72	0.211	9.09	0.185	11.82	0.299	5.74	0.154
DANI-Net	2.40	**0.014**	**1.50**	**0.027**	**0.97**	0.227	**5.76**	**0.096**	**3.90**	0.122	4.69	**0.094**	**6.43**	0.196	8.88	**0.159**	**6.12**	**0.147**	**4.74**	**0.129**

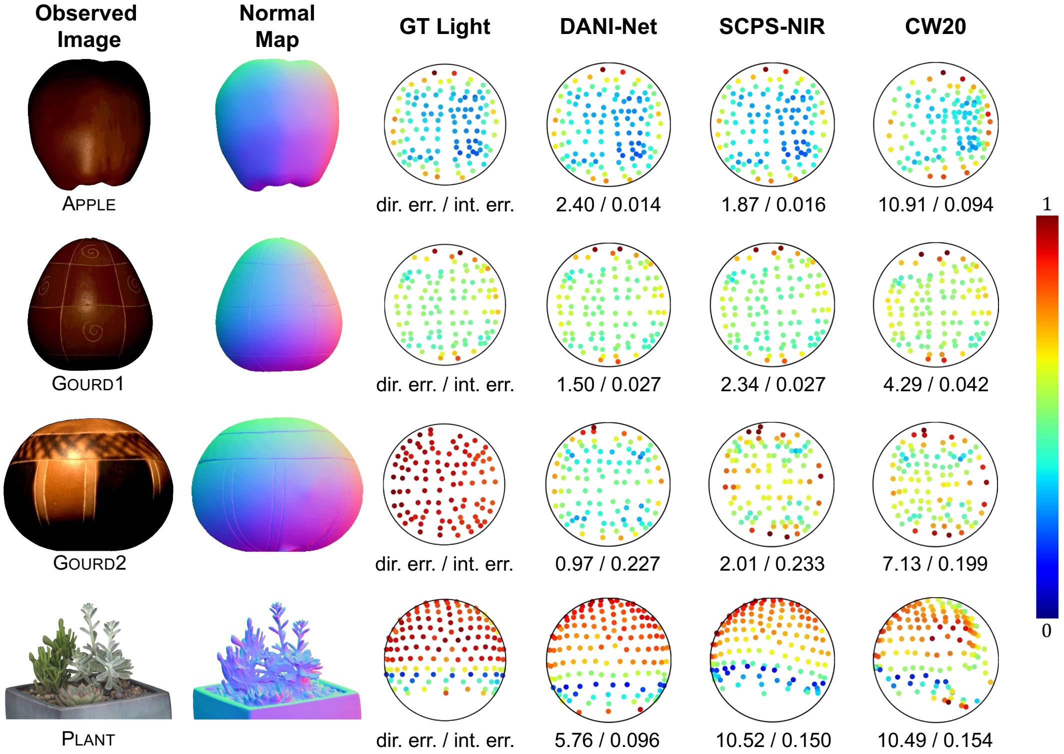


Figure 21. The visual quality comparison among DANI-Net, SCPS-NIR [7], and CW20 [4] on APPLE, GOURD1, and GOURD2 from APPLE & GOURD [1] and PLANT from LIGHT STAGE DATA GALLERY [2] in terms of normal map (column 2) and light map (column 3 - 6). Numbers indicate the MAE (for light directions).

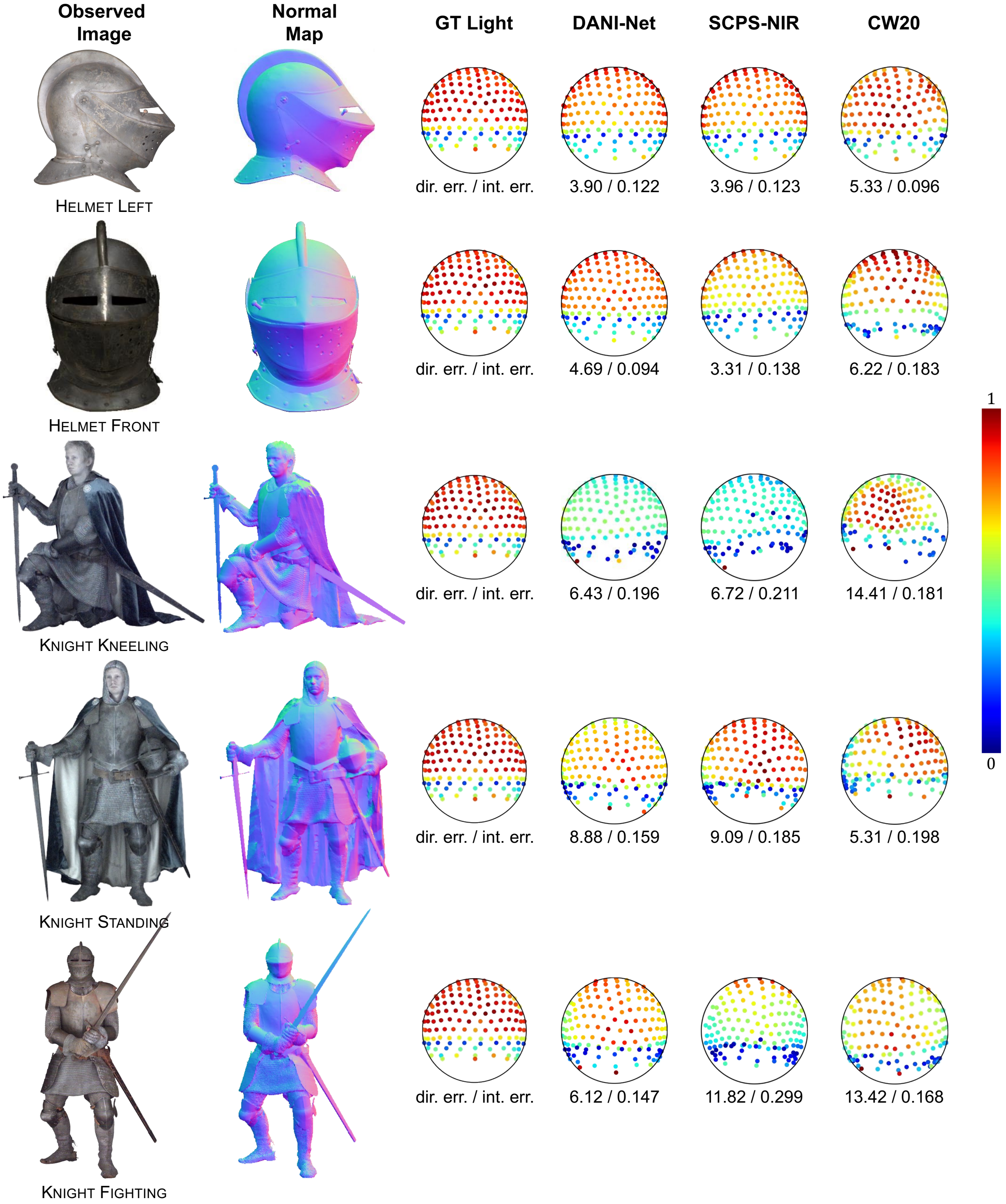


Figure 22. The visual quality comparison among DANI-Net, SCPS-NIR [7], and CW20 [4] on HELMET LEFT, HELMET FRONT, KNIGHT KNEELING, KNIGHT STANDING and KNIGHT FIGHTING from LIGHT STAGE DATA GALLERY [2] in terms of normal map (column 2) and light map (column 3 - 6). Numbers indicate the MAE (for light directions).

References

[1] Neil Alldrin, Todd Zickler, and David Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *Proc. Computer Vision and Pattern Recognition (CVPR)*, 2008. 1, 14

[2] Charles-Félix Chabert, Per Einarsson, Andrew Jones, Bruce Lamond, Wan-Chun Ma, Sebastian Sylwan, Tim Hawkins, and Paul Debevec. Relighting human locomotion with flowed reflectance fields. In *EGSR*, 2006. 1, 9, 14, 15

[3] Guanying Chen, Kai Han, Boxin Shi, Yasuyuki Matsushita, and Kwan-Yee K Wong. Self-calibrating deep photometric stereo networks. In *Proc. Computer Vision and Pattern Recognition (CVPR)*, 2019. 10

[4] Guanying Chen, Michael Waechter, Boxin Shi, Kwan-Yee K Wong, and Yasuyuki Matsushita. What is learned in deep uncalibrated photometric stereo? In *Proc. European Conference on Computer Vision (ECCV)*, 2020. 1, 3, 4, 5, 6, 10, 11, 14, 15

[5] Satoshi Ikehata. Cnn-ps: Cnn-based photometric stereo for general non-convex surfaces. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–18, 2018. 10

[6] Junxuan Li and Hongdong Li. Neural reflectance for shape recovery with shadow handling. In *Proc. Computer Vision and Pattern Recognition (CVPR)*, 2022. 1, 3, 4, 5, 6, 7

[7] Junxuan Li and Hongdong Li. Self-calibrating photometric stereo by neural inverse rendering. In *Proc. European Conference on Computer Vision (ECCV)*, 2022. 1, 2, 3, 4, 5, 6, 9, 10, 11, 13, 14, 15

[8] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European conference on computer vision*, pages 405–421. Springer, 2020. 2

[9] Diego Nehab, Szymon Rusinkiewicz, James Davis, and Ravi Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. In *ACM transactions on graphics (TOG)*, 2005. 7

[10] Thoma Papadhimitri and Paolo Favaro. A closed-form, consistent and robust solution to uncalibrated photometric stereo via local diffuse reflectance maxima. In *International Journal of Computer Vision (IJCV)*, 2014. 14

[11] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12581–12590, 2022. 1, 9, 10, 11, 12

[12] 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 *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2016. 1, 3, 4, 5, 6, 7, 8, 9, 12, 14