Supplemental Material: Simultaneous Acquisition of High Quality RGB image and Polarization Information using a Sparse Polarization Sensor

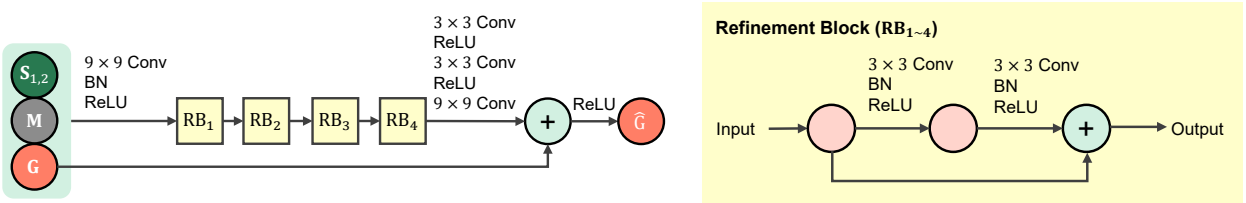
Teppei Kurita Yuhi Kondo Legong Sun Yusuke Moriuchi
Sony Group Corporation
{Teppei.Kurita, Yuhi.Kondo, Legong.Sun, Yusuke.Moriuchi}@sony.com
https://github.com/sony/polar-densification

This supplement is outlined as follows. References to the main study (Section, Equations, Figures and Tables) are highlighted in blue. Section 1 describes the details of the network architecture in Sec. 3 of the main study. Section 2 describes the details of our independently acquired real-world and generated synthetic datasets in Sec 3.4 of the main study. Section 3 describes the details of the experiments discussed in Sec. 4 of the main study. This includes simulation data generation and demosaicing methods as well as detailed quantitative and qualitative experimental results.

1. Network architecture details

1.1. RGB refinement network (RGBRN)

Figure 1 shows the details of the RGBRN discussed in Sec. 3.3 of the main study. The RGB image, stokes vector, and polarization pixel masks are combined in the direction of the input channel. The network consists of four refinement blocks, which are primarily based on full convolution without resolution reduction to learn the difference from the input RGB image.



(a) Network overview. (b) Refinement block.

Figure 1. **Details of RGB refinement network (RGBRN).** (a) Network overview, and (b) refinement block.

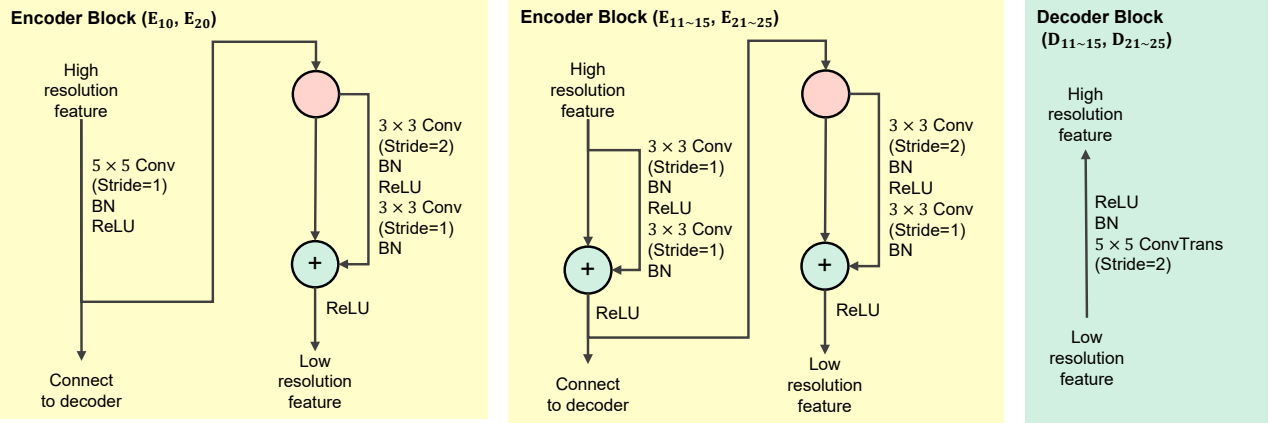
1.2. Polarization compensation network (PCN)

In this section, the details of the PCN discussed in Sec. 3.3 and Fig. 6 of the main study are described in detail. Each encoder and decoder block in the network is organized as shown in Fig. 2. The encoder block first performs convolution to generate the features to be connected to the decoder and then performs convolution with a stride set to 2 to generate low-resolution features. The decoder block generates high-resolution features via transposed convolution with a stride set to 2. Moreover, although omitted in the main study, the polarization information $\hat{\mathbf{S}}_{1,2}$ is obtained as the final output by blending the first and second outputs, $\hat{\mathbf{S}}^{1\mathrm{st}}_{1,2}$ and $\hat{\mathbf{S}}^{2\mathrm{nd}}_{1,2}$, with their respective confidence levels, $\hat{\mathbf{C}}^{1\mathrm{st}}$ and $\hat{\mathbf{C}}^{2\mathrm{nd}}$, using Eqn. 1, as follows.

$$\hat{\mathbf{S}}_{1,2} = \frac{e^{\hat{\mathbf{C}}^{1\mathrm{st}}} \cdot \hat{\mathbf{S}}^{1\mathrm{st}}_{1,2} + e^{\hat{\mathbf{C}}^{2nd}} \cdot \hat{\mathbf{S}}^{2\mathrm{nd}}_{1,2}}{e^{\hat{\mathbf{C}}^{1\mathrm{st}}} + e^{\hat{\mathbf{C}}^{2\mathrm{nd}}}}. \tag{1}$$

2. Dataset details

This section discusses the datasets in Sec. 3.4 of the main study in detail. A comparison with other published datasets that are more detailed than Tab. 2 in the main study is presented in Tab. 1. Herein, the vertical and horizontal resolution



(a) Encoder block (E_{10}, E_{20}). (b) Encoder block ($E_{11\sim15}$, $E_{21\sim25}$). (c) Decoder block.

Figure 2. **Details of the polarization compensation network (PCN).** (a) Encoder block (E_{10}, E_{20}). (b) Encoder block ($E_{11\sim15}$, $E_{21\sim25}$). (c) Decoder block ($D_{11\sim15}$, $D_{21\sim25}$).

(Resolution), type of scene (Level), and type of accompanying ground truth data (Ground truth) have been compared. All of our datasets consist of scenes, and the depth, surface normals, reflectance and segmentation masks of the synthetic datasets were simultaneously generated, as shown in Fig 3.

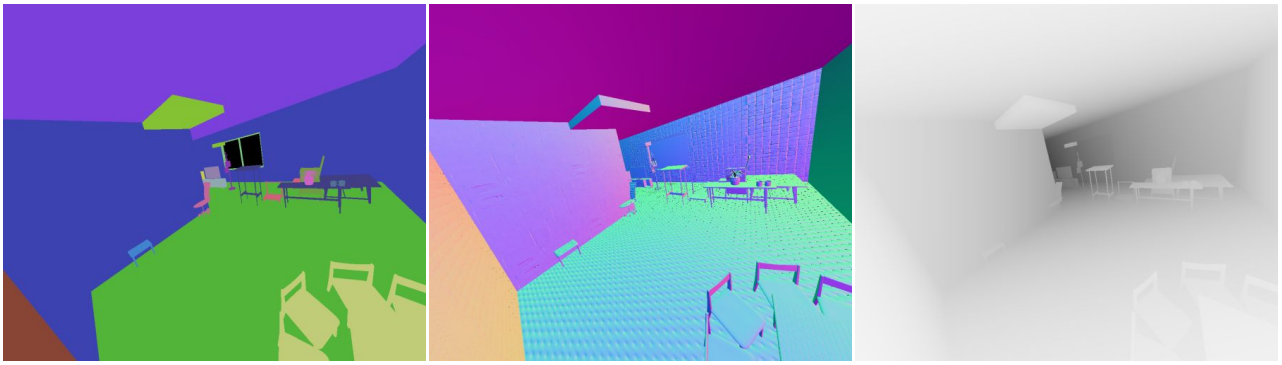
Further, we have described the details of the synthetic dataset. Herein, a polarization-reflective ray-tracing renderer was implemented and used. As mentioned in the main study, the floorplans, camera positions, and objects were procedurally generated. Furthermore, 110 floorplans and 100 data for each floorplan were generated. The camera field of view (FOV) was randomly selected from 30° to 120°, the F value was randomly selected from 0.8 to 8, and the camera roll was randomly selected from $-15°$ to 15°. Half of the scenes were set up as day and the other half as night. Additionally, an outdoor environment map was set up as the light source for the day scenes. The number of SPP (sample per pixel) was set to 100, and the maximum number of ray bounces was set to 7. The NVIDIA OptiX Denoiser [2] was used to remove the noise and improve the quality of RGB images. It takes approximately 8 s to render 1.3 M (1216×1024) data on an NVIDIA RTX3090 GPU, and 11000 data can be generated in approximately 1 d.

We acquired real-world datasets by two methods: using a polarization camera and turning a polarizer. Figure 4 shows examples of each dataset. Using a polarization camera to acquire images is easy; however, the resolution of the polarization camera is limited and the data quality is not high because of the demosaicing artifacts, as shown in Fig. 4(a). If a polarizer is rotated in front of the RGB camera, high-quality data equivalent to that of a normal high-resolution camera can be acquired, as shown in Fig. 4 (b); however, large amounts of data cannot be acquired rapidly because obtaining one image takes several minutes.

Table 1. Comparison between different polarization datasets

Dataset	Level	Collection	Size	Resolution	Ground truth
Ba [1]	Object	Polarization Camera	263	1224×1024	RGB, Polarization, Surface Normal
Lei [4]	Scene	Polarization Camera	522	1224×1024	RGB, Polarization, Surface Normal, Depth
Ono [5]	Scene	Polarization Camera	69	2448×2048	RGB, Polarization
	Scene	Polarizer Rotation	13	2080×2080	RGB, Polarization
Ours	Scene	Synthetic	11000	1216×1024 & 768×576	RGB, Polarization ($\mathbf{S}_{1,2}$), Surface Normal, Depth, Reflectance, Segmentation Mask
	Scene	Polarization Camera	811	2448×2048	RGB, Polarization
	Scene	Polarizer Rotation	238	5472×3648	RGB, Polarization

(a) RGB image (b) $\mathbf{S}_1$ component (c) Reflectance



(d) Segmentation mask (d) Surface normal (e) Depth

Figure 3. **Details of the synthetic dataset.** (a) RGB image ($\mathbf{S}_0$ component), (b) $\mathbf{S}_1$ component, (c) reflectance, (d) segmentation mask, (d) surface normal, and (e) depth.

(a) Polarization camera (b) Polarizer rotation

Figure 4. **Details of real-world datasets.** (a) Polarization camera, and (b) polarizer rotation.

3. Experiment details

3.1. Dataset and implementation details

This section discusses the simulation method used to generate the raw data for the sparse polarization sensor in each dataset that was used in Sec. 4.1 of the main study.

Raw image generation from synthetic dataset: The synthetic dataset consists of RGB data and the $\mathbf{S}_{1,2}$ ground truth, as shown in Fig 5 (a). The RGB image and $\mathbf{S}_{1,2}$ are grayed and the four-polarization angle image is calculated to reflect the sensitivity difference. Each pixel from the RGB image and the generated four-polarization angle image are selected to generate raw data.

Raw image generation from polarization camera dataset: A conventional polarization camera produces a raw image, as shown in Fig. 5 (b). Demosaicing is performed to generate an RGB image and a four-polarization angle image. Subsequently, the four-polarization angle image is multiplied by the sensitivity difference gain. Finally, each pixel from the RGB image and the generated four-polarization angle image is selected to generate the raw data.

Raw image generation from polarizer rotation dataset: A polarizer is placed on the entire surface of a regular RGB camera and rotated to obtain an RGB image with four-polarization angles, as shown in Fig. 5 (c). A non-polarized RGB image is generated by averaging the RGB values of the four-polarization angles. The RGB images of each of the four-polarization angles are grayed to produce a one-channel four-polarization angle image. Next, the four-polarization angle image is multiplied by the sensitivity difference gain. Raw data is generated by selecting each pixel from the RGB image and the generated four-polarization angle image. Finally, as the raw image generated here is used for evaluation purposes, it is noised according to the noise model of the sensor.

3.2. Demosaicing

This section details the demosaicing process implemented to obtain RGB and four-polarization angle images in Sec. 3.3 of the main study. Demosaicing is a method that considers the pixels near the target pixel and their spatial frequency [3, 5]. For sparse polarization sensors, the same process is used for unpolarized pixels to generate RGB values, and four-polarization angle pixels are interpolated in the range of 2×2 to generate the sparse stokes vector, as shown in Fig. 6.

3.3. Additional assessment results

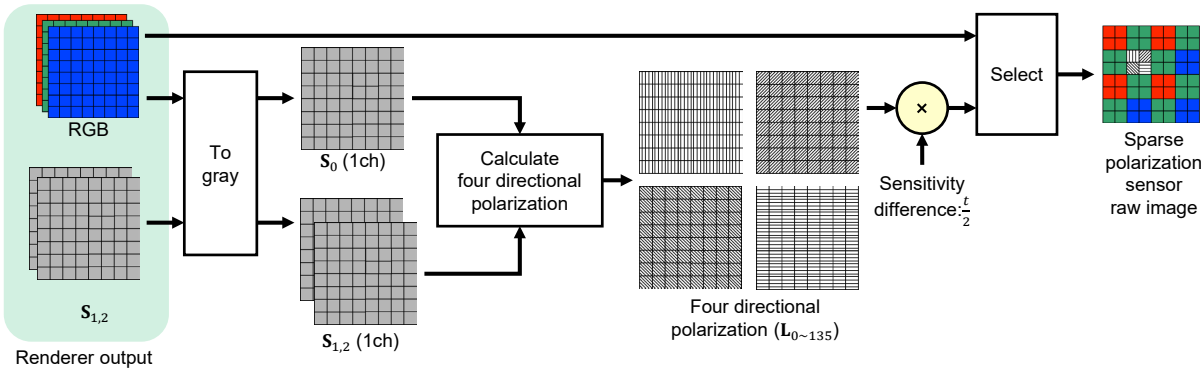
This section discusses the additional results of the evaluation in Sec 4.2 of the main study.

Ablation study: Table 2 presents a comparison of the ablation study in Tab 4 of the main study for different percentages of polarization pixels. The results show that the higher the ratio of polarization pixels, the more effective FTB and AFA are. Conversely, when the percentage of polarized pixels is low ($r = 1/64$), FTB and AFA are not as effective. This may be because the extremely low number of polarization pixels limits the generation of valid features for completion, thereby limiting the effectiveness of FTB and AFA. Therefore, we did not implement FTB and AFA when the percentage of polarized pixels was very low.

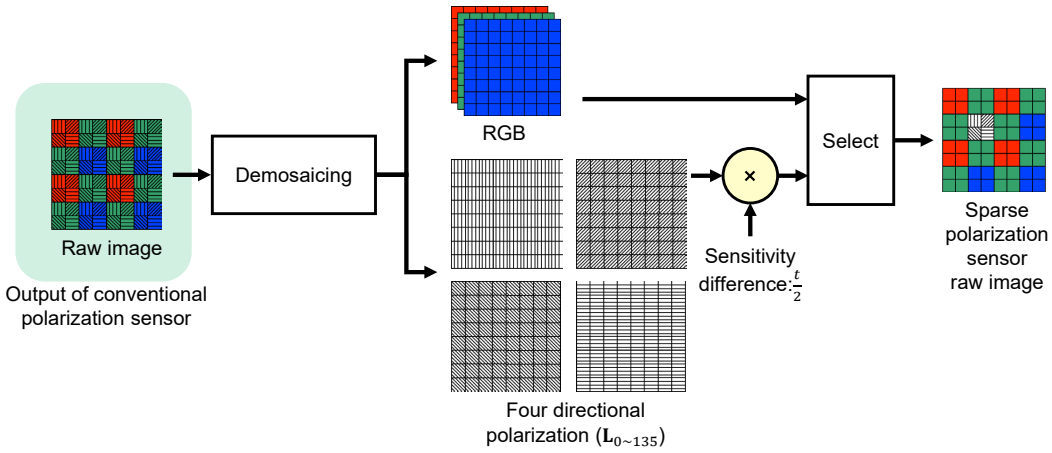
Comparison with depth completion and upsampling networks: Table 3 presents a comparison of the proposed method with other networks discussed in Tab. 5 of the main study for different percentages of polarization pixels. Furthermore, the computation time has been additionally described. The results show that the smaller the percentage of polarization pixels, the more effective our method is. When the percentage of polarization pixels is high, although NLSPN may perform slightly better than our method for some indices, it requires more computation time. Figure 7 and 8 compare the results of our method with other methods (conventional polarization sensors, basic methods (Eqn. 2 and Eqn. 3 of the main study), and other networks). The effectiveness of our method can be qualitatively confirmed.

Comparison between the synthetic and real-world datasets: Table 4 presents a comparison of the datasets discussed in Tab. 6 of the main study with different percentages of polarized pixels. The trend of the parameters is the same for different percentages of polarized pixels.

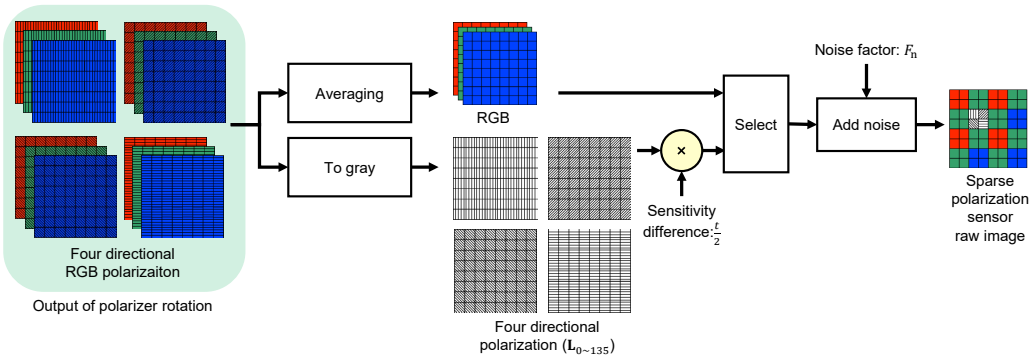
Additional qualitative evaluation results: The qualitative evaluation results of our method in various scenes, which could not be described in the main study due to space limitations, are shown in Fig. 9,10,11,12,13,14,15,16,17, and 18. The effectiveness of our method for conventional polarization sensors can be confirmed.



(a) Raw image generation for synthetic dataset



(b) Raw image generation for polarization camera dataset



(c) Raw image generation for polarizer rotation dataset

Figure 5. **Details of raw image generation for the:** (a) synthetic, (b) polarization camera, and (c) polarizer rotation datasets.

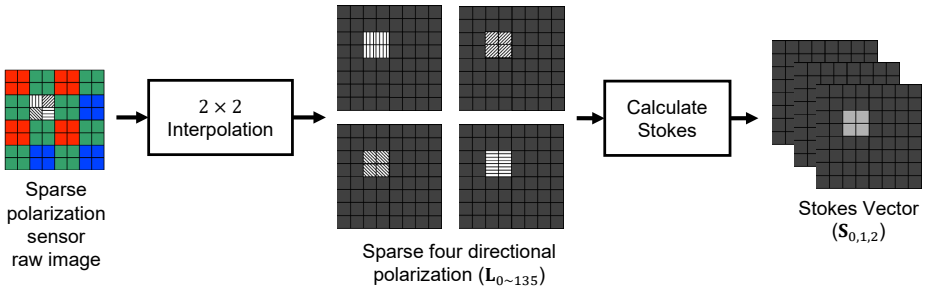


Figure 6. **Interpolation of polarization pixels.** Pixel values for four polarization angles are interpolated in a 2×2 region where polarization pixels are present.

Table 2. **Ablation study.** Additional results.

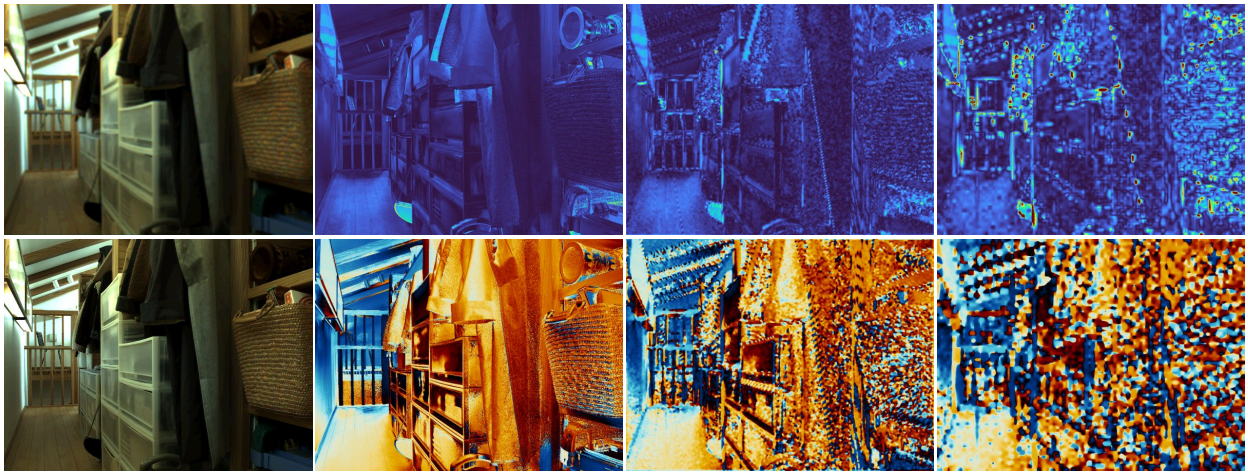
Polarization sensor	r	Operation	$\mathbf{S}_{0,1,2}$ RMSE ↓ [$\times10^{-3}$]	$\mathbf{S}_{1,2}$ RMSE ↓ [$\times10^{-3}$]	DoLP PSNR ↑ [dB]	AoLP Error ↓ [°]
Conventional	0	Baseline (four polar comp.)	12.830	11.068	17.32	24.99
		+ SNA ($\mathbf{S}_{1,2}$ comp.)	14.563	4.101	26.54	12.97
		+ RGBRN	6.931	3.909	26.10	14.77
		+ FTB	**6.915**	3.865	26.92	13.06
		+ AFA	**6.915**	**3.854**	**27.22**	**12.35**
Sparse	$\frac{1}{4}$	Baseline (four polar comp.)	10.948	10.574	16.57	31.70
		+ SNA ($\mathbf{S}_{1,2}$ comp.)	10.328	4.046	26.94	**12.31**
		+ RGBRN	4.974	3.979	26.75	13.43
		+ FTB	4.952	3.952	26.81	13.14
		+ AFA	**4.881**	**3.824**	**27.41**	12.36
	$\frac{1}{16}$	Baseline (four polar comp.)	9.099	8.952	18.76	29.05
		+ SNA ($\mathbf{S}_{1,2}$ comp.)	8.568	4.304	26.25	**13.30**
		+ RGBRN	4.752	4.216	26.45	13.50
		+ FTB	4.727	4.186	26.42	13.61
		+ AFA	**4.707**	**4.151**	**26.48**	13.95
	$\frac{1}{64}$	Baseline (four polar comp.)	8.746	8.794	17.27	32.98
		+ SNA ($\mathbf{S}_{1,2}$ comp.)	8.110	4.952	24.80	16.02
		+ RGBRN	5.033	**4.791**	**25.19**	**15.56**
		+ FTB	5.080	4.853	25.01	16.04
		+ AFA	**5.032**	4.801	24.85	17.39

Table 3. **Comparison of the results obtained after replacing PCN with depth completion and upsampling networks in our network architecture.** Additional results.

Polarization sensor	r	Method	$\mathbf{S}_{0,1,2}$ RMSE ↓ $[\times 10^{-3}]$	$\mathbf{S}_{1,2}$ RMSE ↓ $[\times 10^{-3}]$	DoLP PSNR ↑ [dB]	AoLP Error ↓ [°]	Runtime [ms]
Conventional	0	UNet	6.915	4.013	26.68	12.97	40.8
		U2Net	7.120	4.277	25.11	15.42	160.6
		FDSR	7.032	4.223	26.81	**12.06**	35.6
		GuideNet	6.907	4.048	25.77	15.04	61.8
		NLSPN	**6.894**	3.879	26.37	14.73	172.8
		Ours	6.915	**3.854**	**27.22**	12.35	50.9
Sparse	$\frac{1}{4}$	UNet	4.986	4.013	26.46	13.91	41.0
		U2Net	5.179	4.306	25.12	16.17	162.1
		FDSR	5.191	4.389	25.87	14.02	35.7
		GuideNet	4.984	4.021	25.79	14.33	71.3
		NLSPN	**4.848**	**3.788**	27.40	13.13	172.9
		Ours	4.881	3.824	**27.41**	**12.36**	50.6
	$\frac{1}{16}$	UNet	4.974	4.568	25.00	16.31	41.0
		U2Net	5.537	5.224	23.90	19.11	158.8
		FDSR	5.128	4.837	25.12	15.48	36.0
		GuideNet	4.859	4.390	25.59	15.62	68.7
		NLSPN	4.905	4.470	23.97	20.20	172.8
		Ours	**4.707**	**4.151**	**26.48**	**13.95**	50.9
	$\frac{1}{64}$	UNet	6.526	6.872	21.96	31.74	41.1
		U2Net	5.981	6.108	21.92	25.34	163.7
		FDSR	5.570	5.615	23.84	18.74	35.6
		GuideNet	6.241	6.255	16.32	33.34	67.4
		NLSPN	6.402	6.564	21.05	26.48	173.1
		Ours	**5.033**	**4.791**	**25.19**	**15.56**	50.5

Table 4. **Comparison between real-world (R) and synthetic (S) datasets.** Additional results.

Polarization sensor	r	Data	Train size	$\mathbf{S}_{0,1,2}$ RMSE ↓ $[\times 10^{-3}]$	$\mathbf{S}_{1,2}$ RMSE ↓ $[\times 10^{-3}]$	DoLP PSNR ↑ [dB]	AoLP Error ↓ [°]
Conventional	0	R	729	21.578	4.731	25.60	13.88
		S	729	8.836	5.177	24.47	15.62
		S	10000	7.093	4.055	26.69	**12.29**
		R+S	10729	**6.915**	**3.854**	**27.22**	12.35
Sparse	$\frac{1}{4}$	R	729	11.162	5.240	25.40	14.79
		S	729	6.346	5.342	24.05	15.55
		S	10000	5.137	4.093	26.58	13.11
		R+S	10729	**4.881**	**3.824**	**27.41**	**12.36**
	$\frac{1}{16}$	R	729	7.257	5.809	24.71	15.76
		S	729	6.304	5.889	23.32	18.38
		S	10000	4.975	4.471	25.81	14.41
		R+S	10729	**4.707**	**4.151**	**26.48**	**13.95**
	$\frac{1}{64}$	R	729	6.762	6.400	23.84	17.86
		S	729	6.718	6.629	22.43	21.11
		S	10000	5.431	5.305	24.57	16.35
		R+S	10729	**5.033**	**4.791**	**25.19**	**15.56**

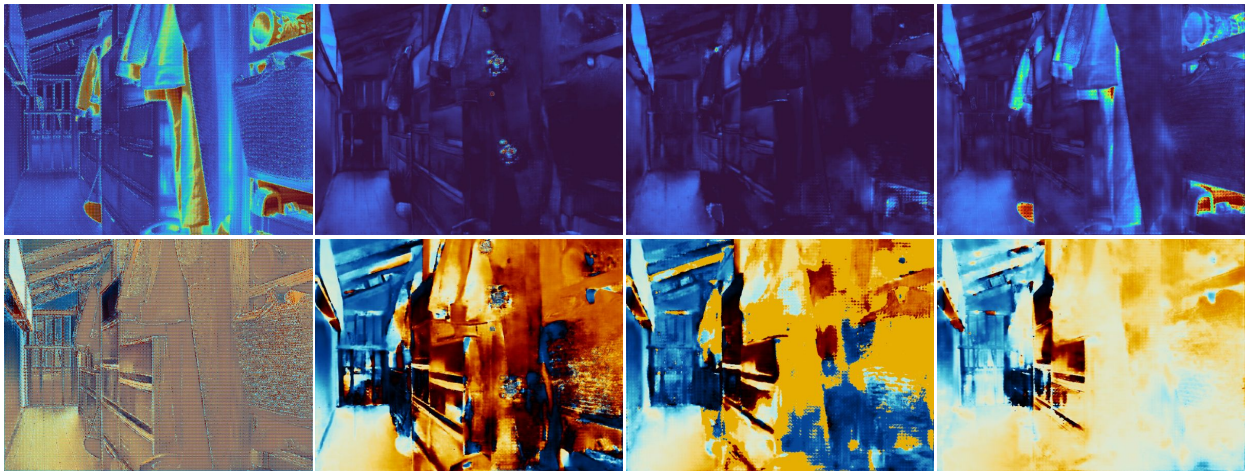


(Top) Conventional RGB
(Bottom) Our RGB

(a) Ground truth

(b) Conventional

(c) Sparse + Bilinear

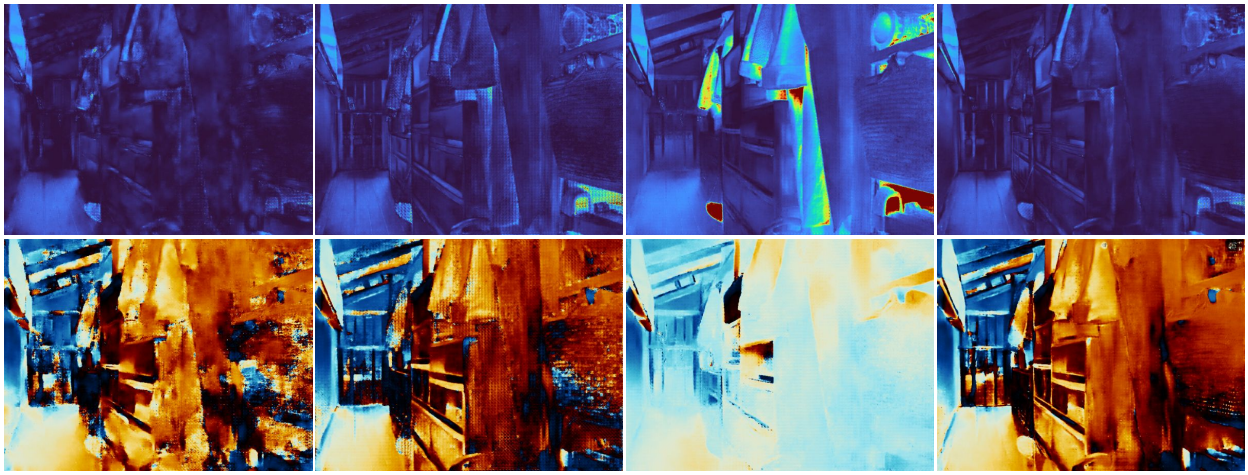


(d) Sparse + Eqn. 2 with PCN (Ours)

(e) Sparse + Eqn. 3 with PCN (Ours)

(f) Sparse + SNA with UNet

(g) Sparse + SNA with U2Net



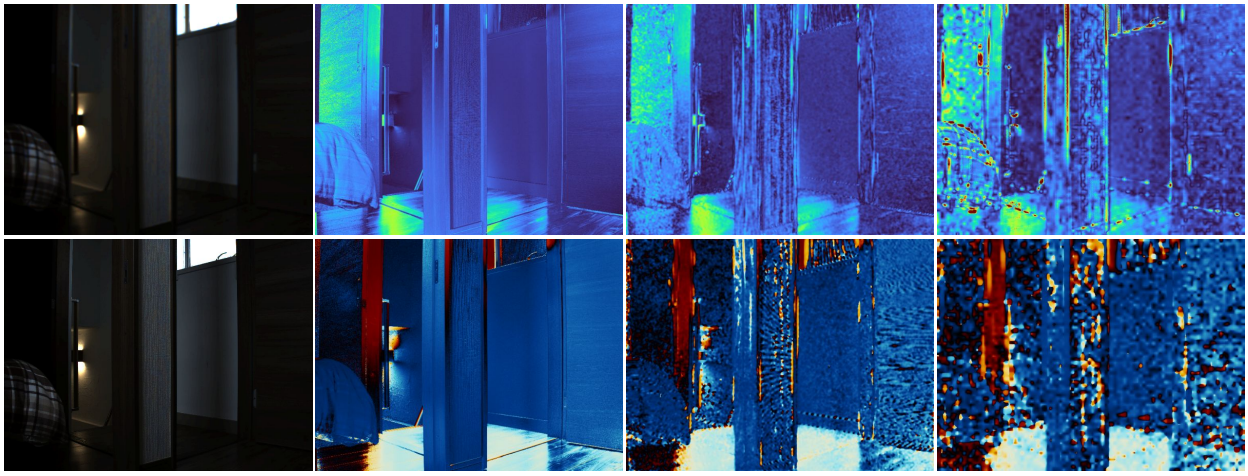
(h) Sparse + SNA with FDSR

(i) Sparse + SNA with GuideNet

(j) Sparse + SNA with NLSPN

(k) Sparse + SNA with PCN (Ours)

Figure 7. **Scene1: Comparison with other methods.** Evaluation at r = 1/16. The top is DoLP and the bottom is AoLP.

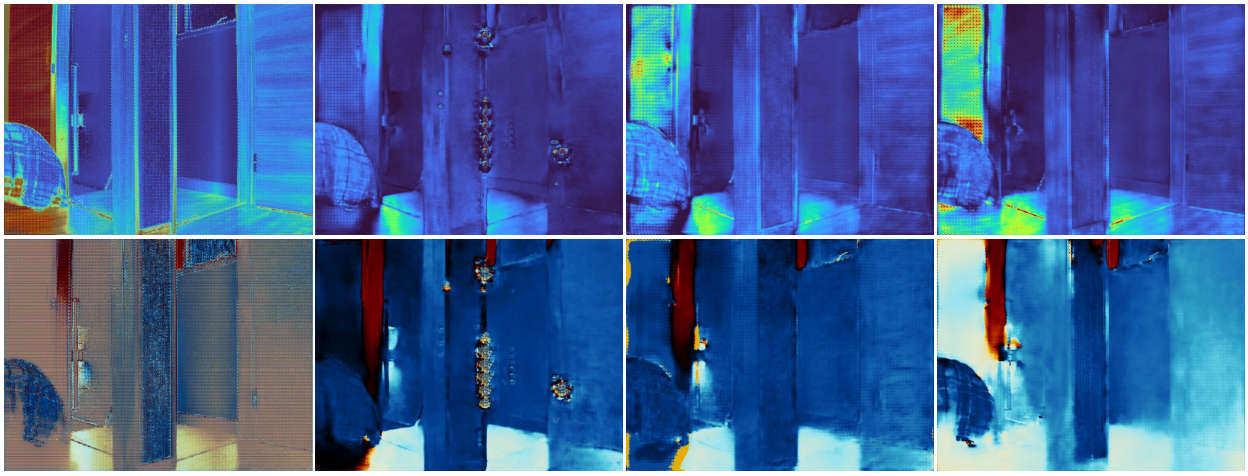


(Top) Conventional RGB
(Bottom) Our RGB

(a) Ground truth

(b) Conventional

(c) Sparse + Bilinear

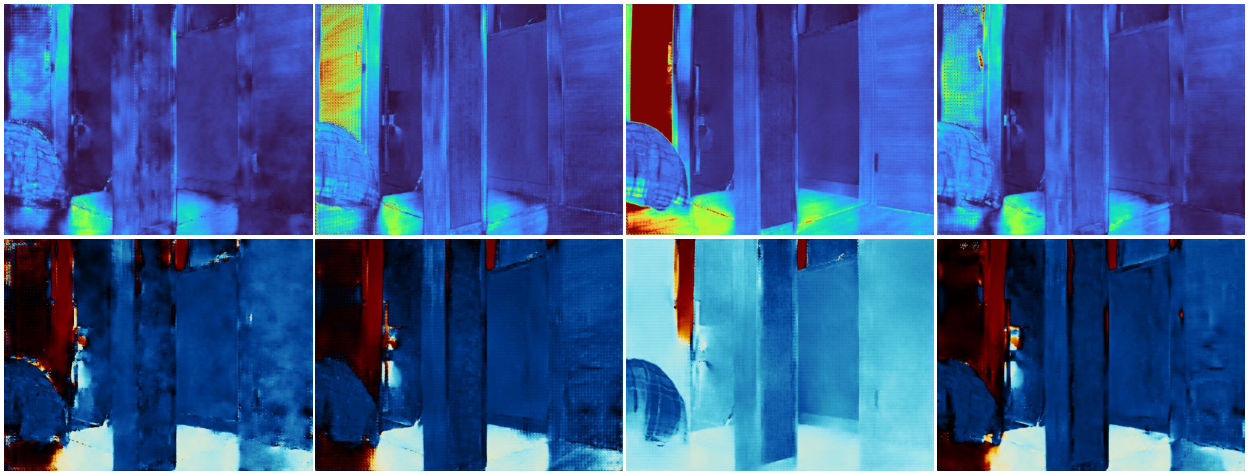


(d) Sparse + Eqn. 2 with PCN (Ours)

(e) Sparse + Eqn. 3 with PCN (Ours)

(f) Sparse + SNA with UNet

(g) Sparse + SNA with U2Net



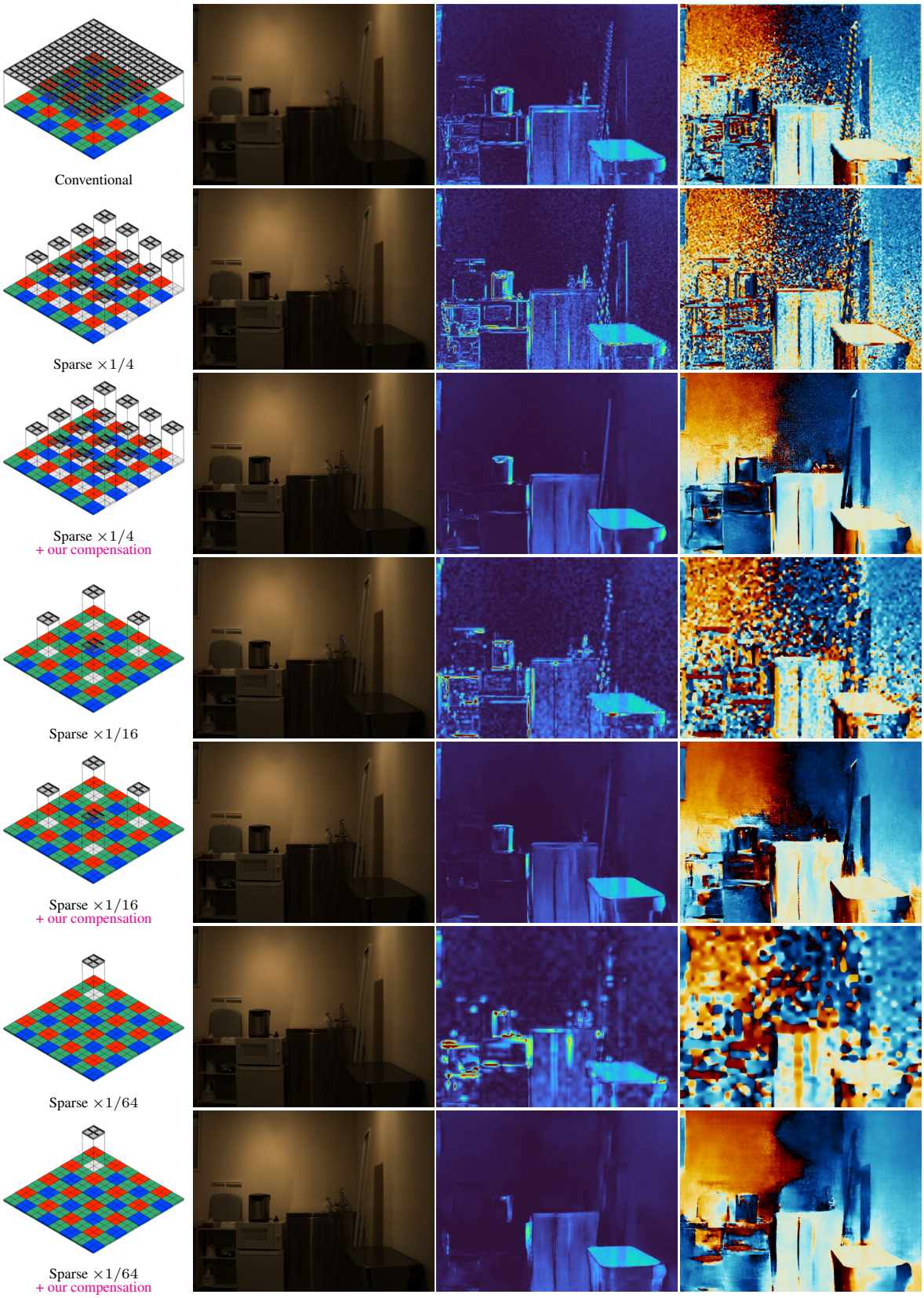
(h) Sparse + SNA with FDSR

(i) Sparse + SNA with GuideNet

(j) Sparse + SNA with NLSPN

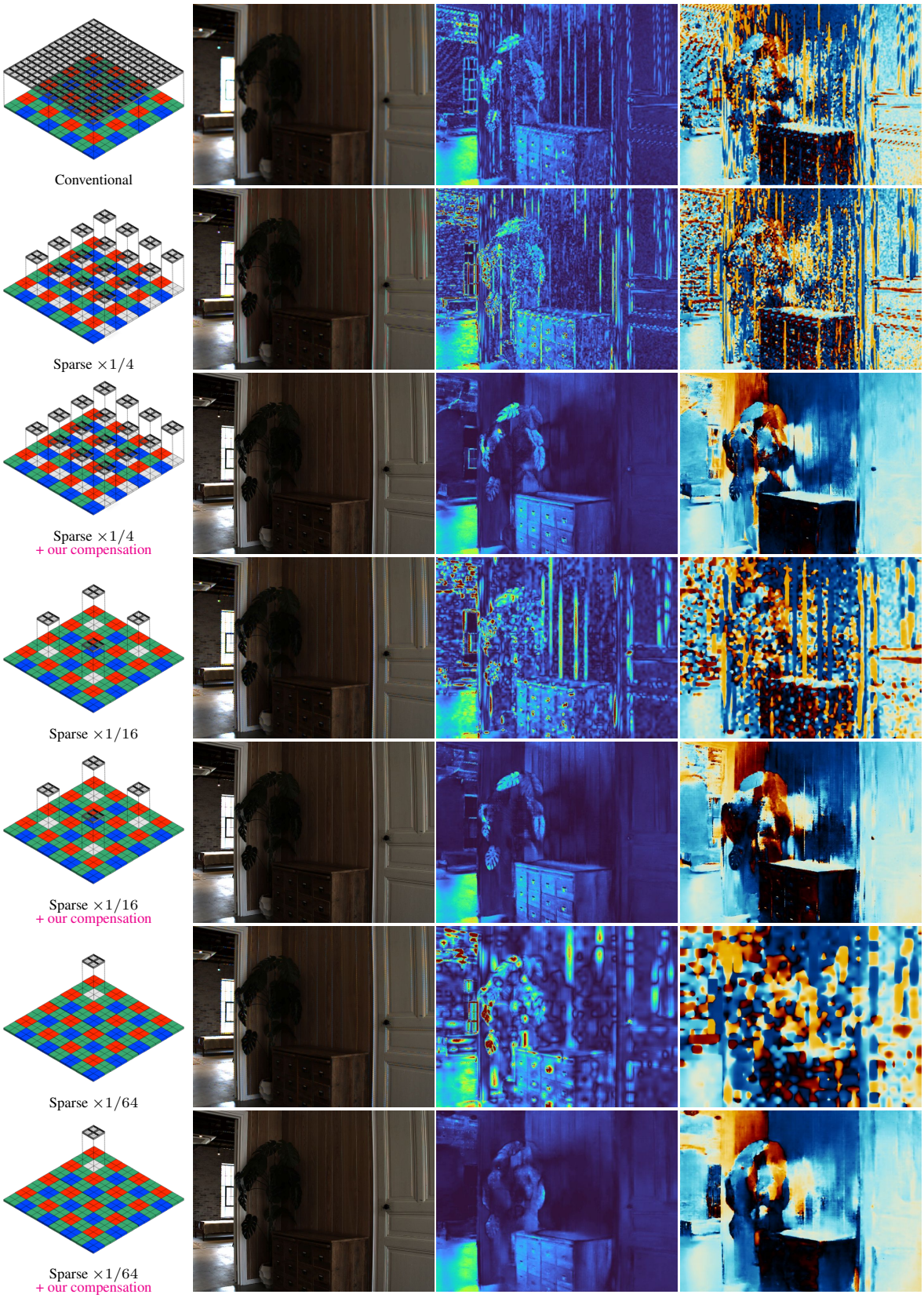
(k) Sparse + SNA with PCN (Ours)

Figure 8. **Scene2: Comparison with other methods.** Evaluation at r = 1/16. The top is DoLP and the bottom is AoLP.



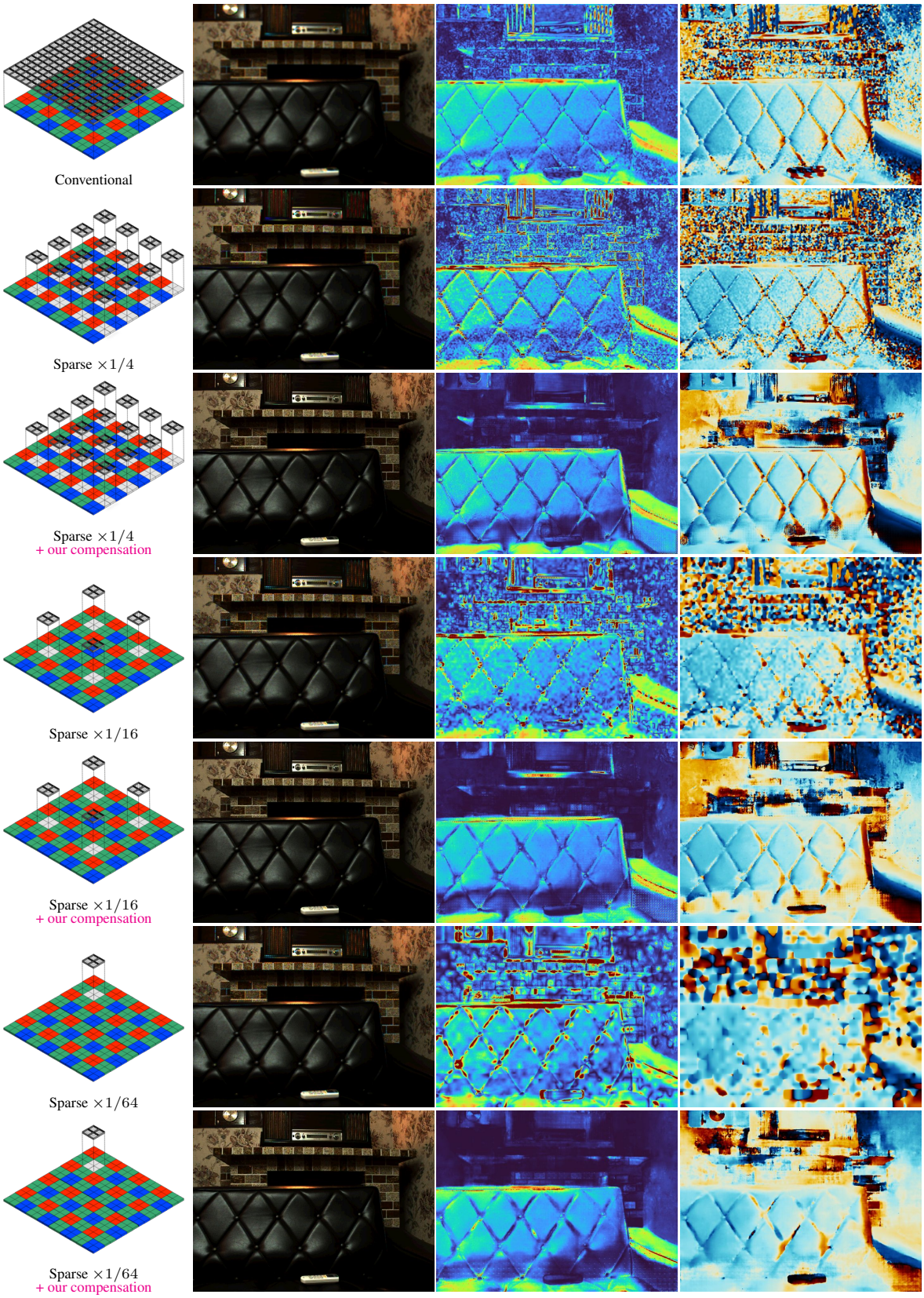
(a) Sensor array (b) RGB (c) DoLP (d) AoLP

Figure 9. **Scene1: Comprehensive qualitative evaluation of conventional and sparse polarization sensors.** The effectiveness of our compensation is qualitatively demonstrated.



(a) Sensor array (b) RGB (c) DoLP (d) AoLP

Figure 10. **Scene2: Comprehensive qualitative evaluation of conventional and sparse polarization sensors.** The effectiveness of our compensation is qualitatively demonstrated.



(a) Sensor array (b) RGB (c) DoLP (d) AoLP

Figure 11. **Scene3: Comprehensive qualitative evaluation of conventional and sparse polarization sensors.** The effectiveness of our compensation is qualitatively demonstrated.

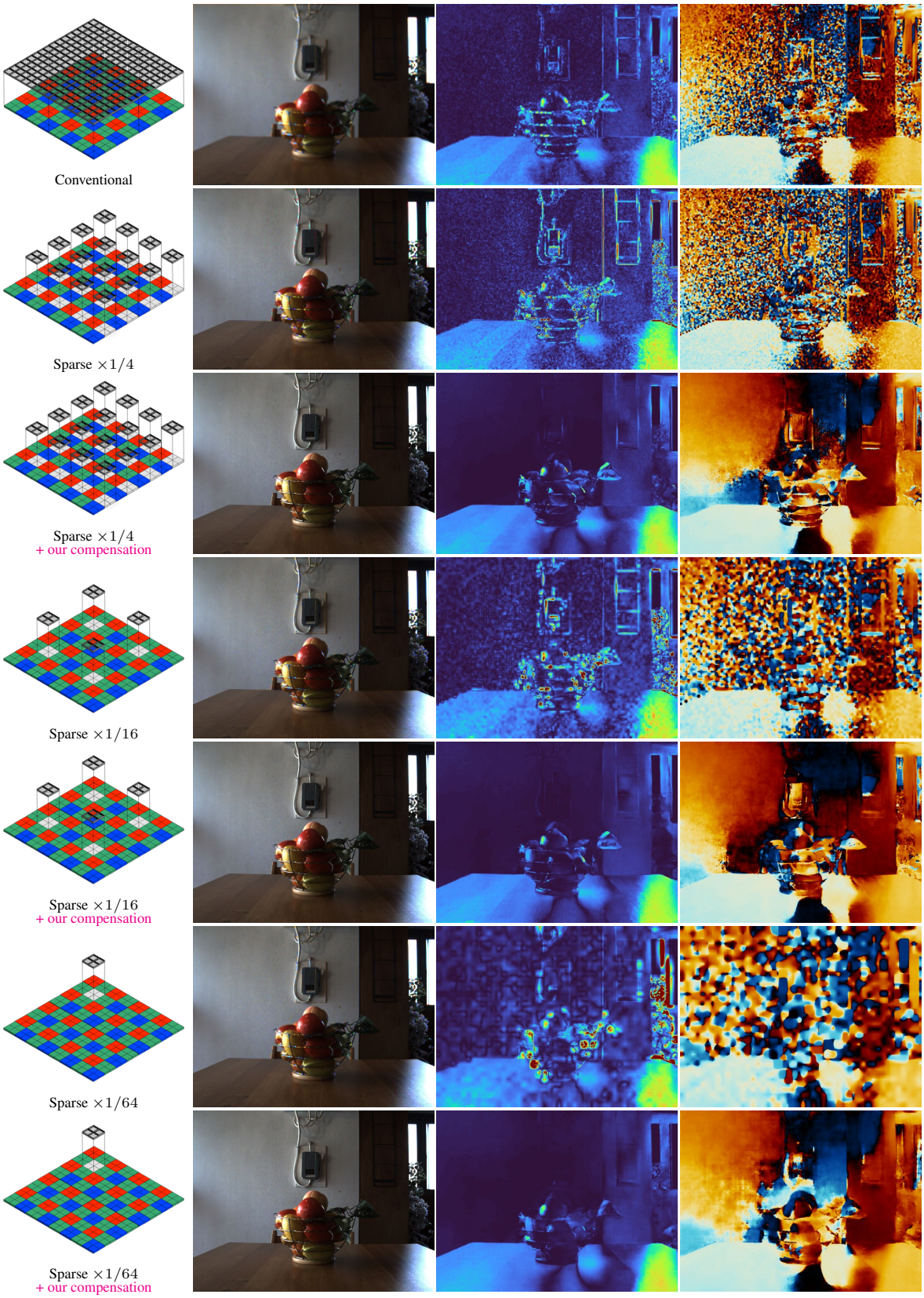
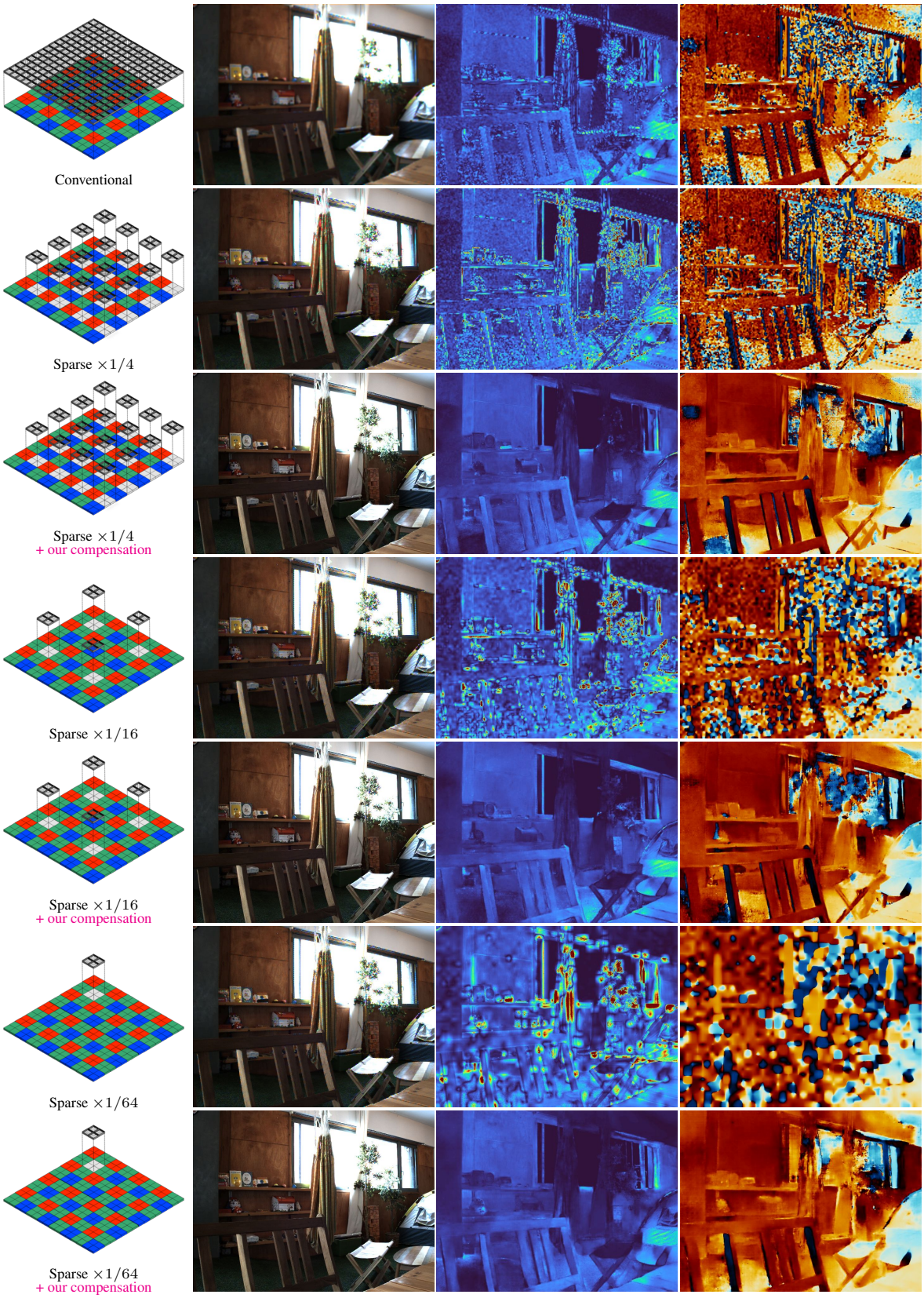
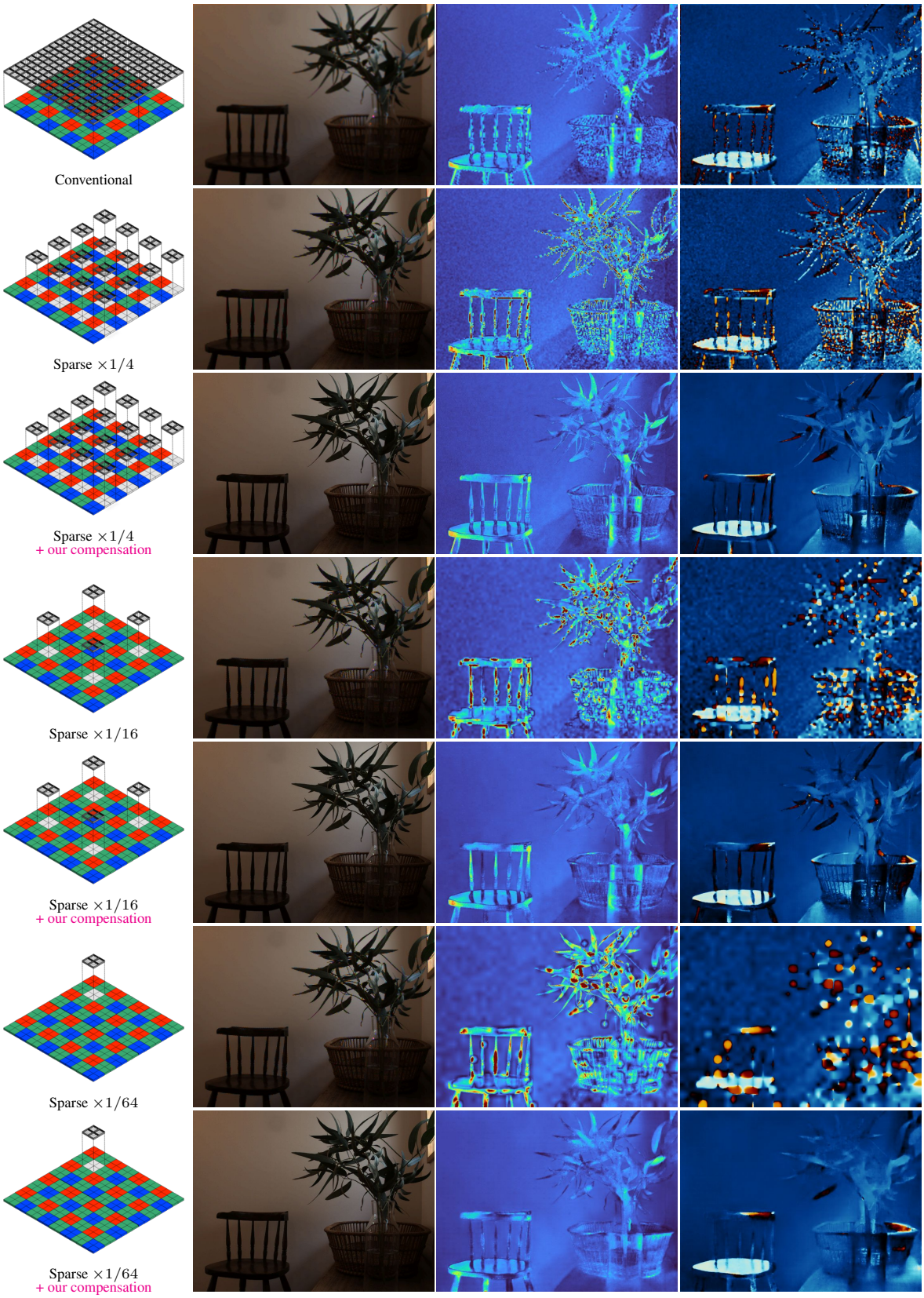


Figure 12. **Scene4: Comprehensive qualitative evaluation of conventional and sparse polarization sensors.** The effectiveness of our compensation is qualitatively demonstrated.



(a) Sensor array (b) RGB (c) DoLP (d) AoLP

Figure 13. **Scene5: Comprehensive qualitative evaluation of conventional and sparse polarization sensors.** The effectiveness of our compensation is qualitatively demonstrated.



(a) Sensor array (b) RGB (c) DoLP (d) AoLP

Figure 14. **Scene6: Comprehensive qualitative evaluation of conventional and sparse polarization sensors.** The effectiveness of our compensation is qualitatively demonstrated.

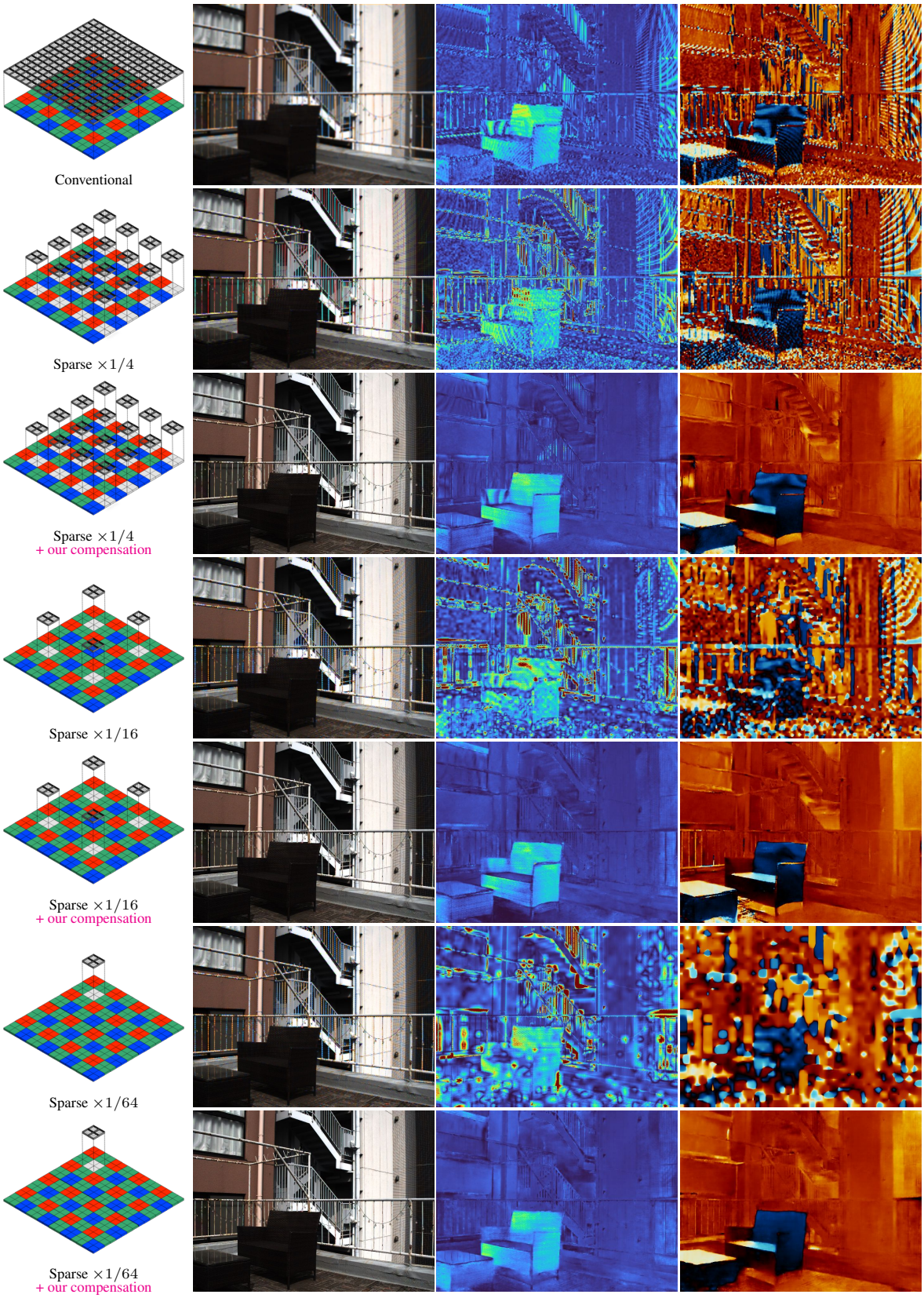
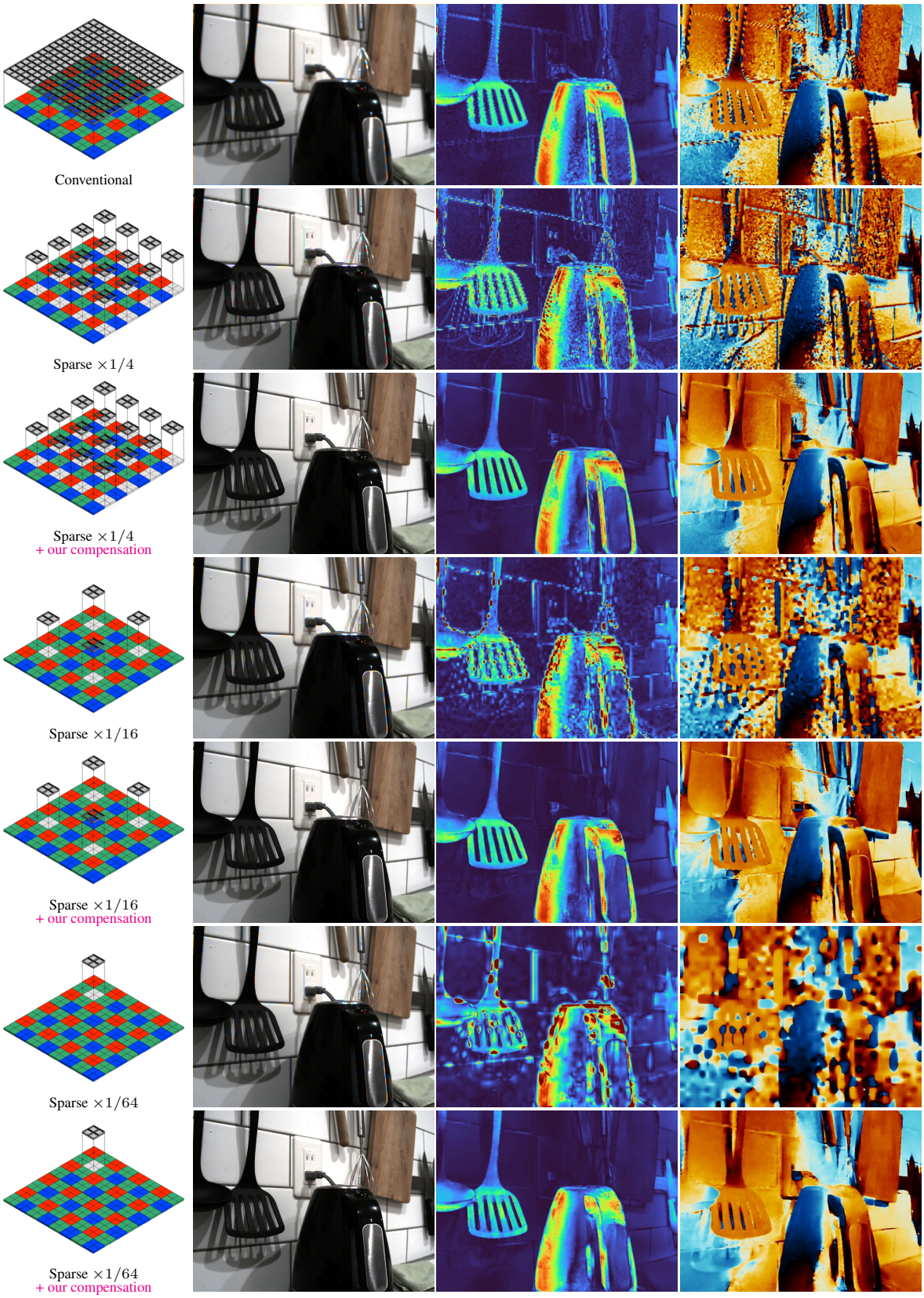
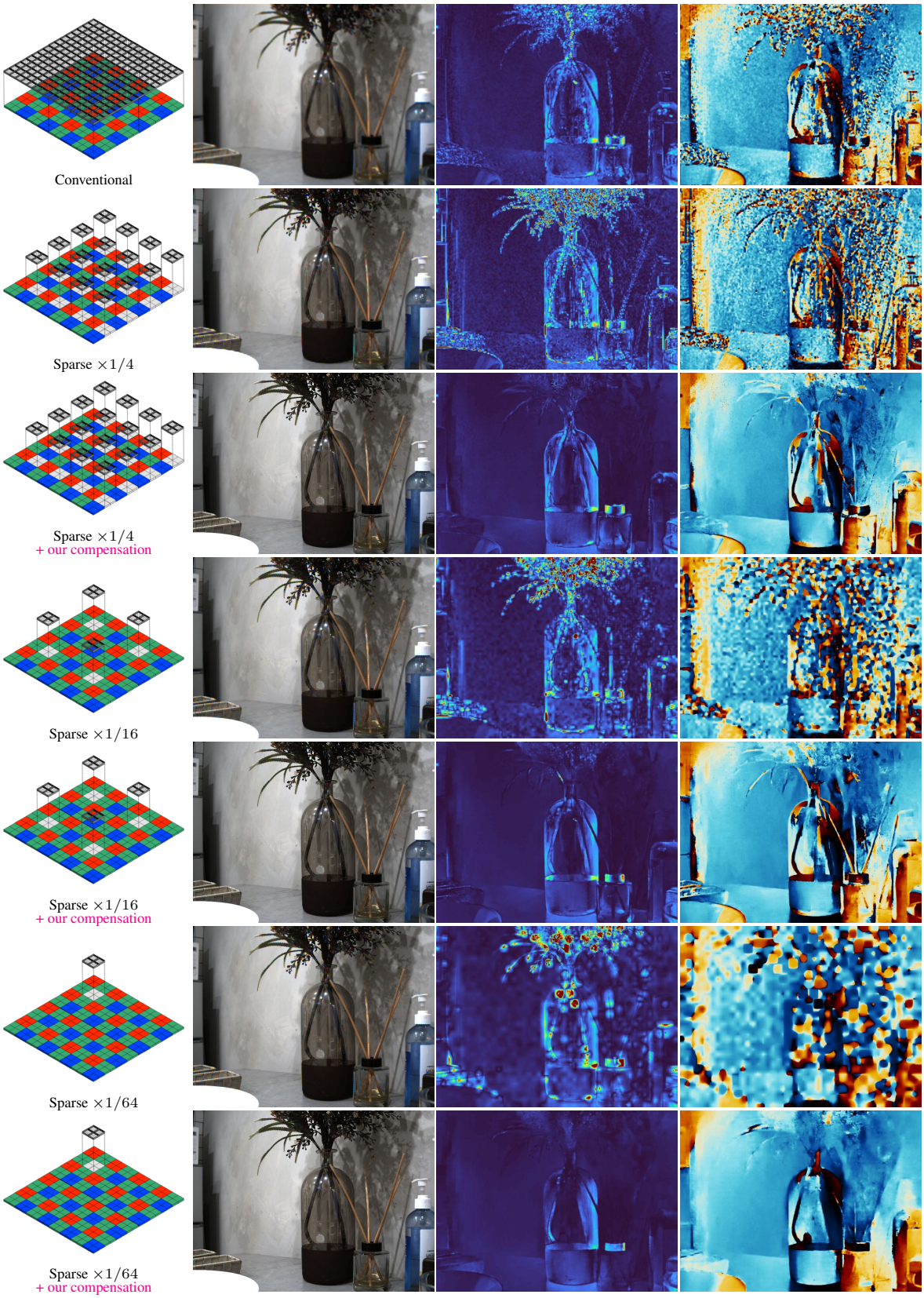


Figure 15. **Scene7: Comprehensive qualitative evaluation of conventional and sparse polarization sensors.** The effectiveness of our compensation is qualitatively demonstrated.



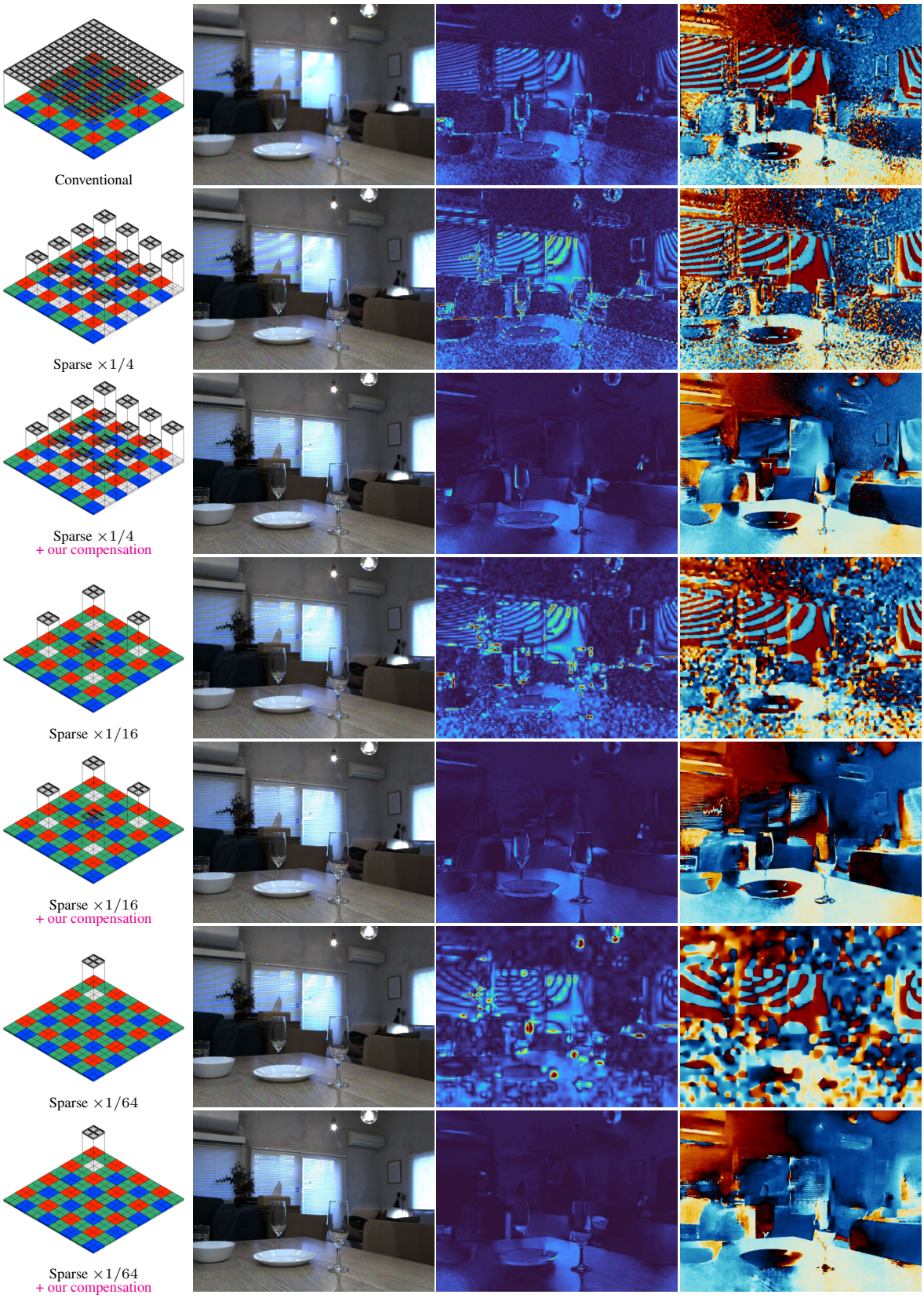
(a) Sensor array (b) RGB (c) DoLP (d) AoLP

Figure 16. **Scene8: Comprehensive qualitative evaluation of conventional and sparse polarization sensors.** The effectiveness of our compensation is qualitatively demonstrated.



(a) Sensor array (b) RGB (c) DoLP (d) AoLP

Figure 17. **Scene9: Comprehensive qualitative evaluation of conventional and sparse polarization sensors.** The effectiveness of our compensation is qualitatively demonstrated.



(a) Sensor array (b) RGB (c) DoLP (d) AoLP

Figure 18. **Scene10: Comprehensive qualitative evaluation of conventional and sparse polarization sensors.** The effectiveness of our compensation is qualitatively demonstrated.

References

[1] Yunhao Ba, Alex Gilbert, Franklin Wang, Jinfa Yang, Rui Chen, Yiqin Wang, Lei Yan, Boxin Shi, and Achuta Kadambi. Deep shape from polarization. In European Conference on Computer Vision (ECCV), pages 554–571. Springer, 2020. 2

[2] Chakravarty R Alla Chaitanya, Anton S Kaplanyan, Christoph Schied, Marco Salvi, Aaron Lefohn, Derek Nowrouzezahrai, and Timo Aila. Interactive reconstruction of monte carlo image sequences using a recurrent denoising autoencoder. ACM Transactions on Graphics (TOG), 36(4):1–12, 2017. 2

[3] Teppei Kurita, Shun Kaizu, Yuhi Kondo, Yasutaka Hirasawa, and Ying Lu. Image processing device and image processing method, Oct. 29 2019. US Patent 10,460,422. 4

[4] Chenyang Lei, Chenyang Qi, Jiaxin Xie, Na Fan, Vladlen Koltun, and Qifeng Chen. Shape from polarization for complex scenes in the wild. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 12632–12641, 2022. 2

[5] Taishi Ono, Yuhi Kondo, Legong Sun, Teppei Kurita, and Yusuke Moriuchi. Degree-of-linear-polarization-based color constancy. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 19740–19749, June 2022. 2, 4