CVPR JUNE 3-7, 2026 DENVER COLORADO

SCE-Depth: A Spherical Compound Eye Framework for Wide FOV Depth Estimation

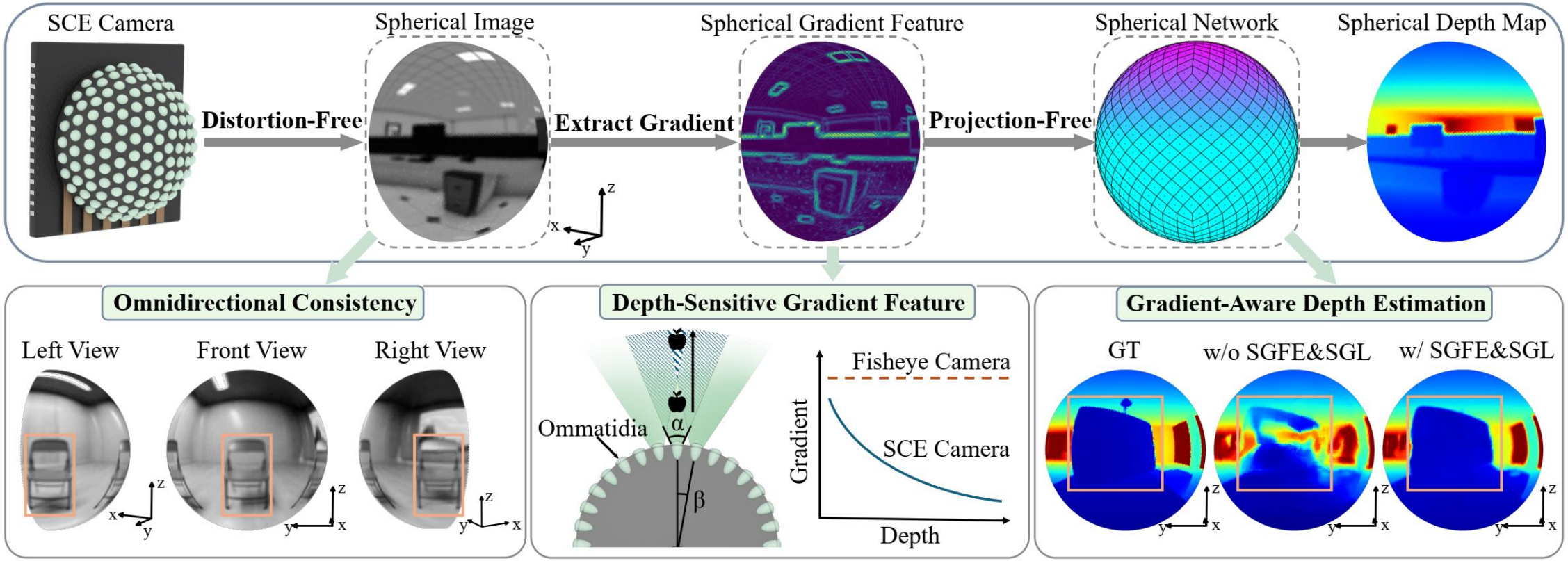
Yi Zhu[1,2,†]; Hao Xiong[1,†]; Lin Xiao[1]; Ranfeng Shi[1]; Qinying Gu[2]; Leilei Gu[1,2,*]
[1]Shanghai Jiao Tong University, [2]Shanghai Artificial Intelligence Laboratory

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

Overview

Native spherical sensing + Spherical gradient feature improves wide-FOV depth estimation.



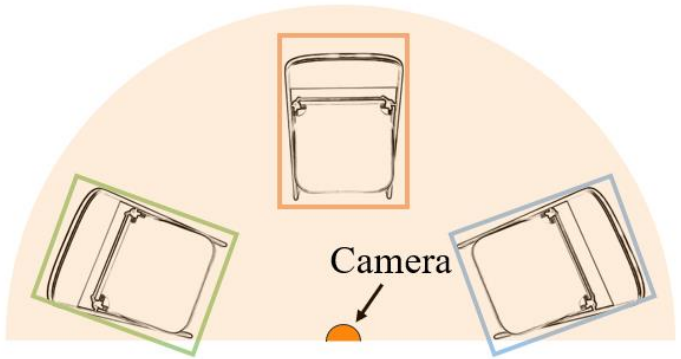
Motivation: Modality Misalignment

Why ordinary fisheye pipelines are not ideal for spherical data

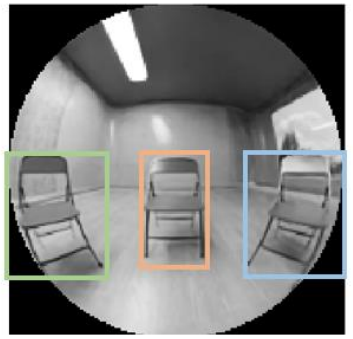
Problem

- Robots, drones and AR devices need compact wide FOV depth perception.
- Fisheye projection maps a sphere to a plane, creating severe barrel distortion.
- Planar CNNs and Transformers learn on geometry that is misaligned with the sensor.

Goal: align the imaging geometry, network representation, and learning objective.

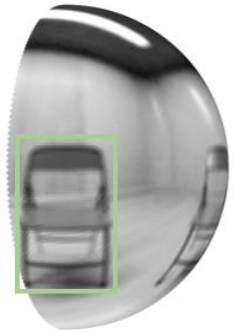


Scene Setup

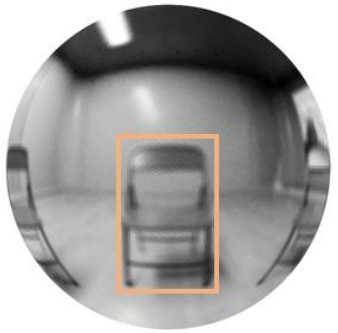


Fisheye

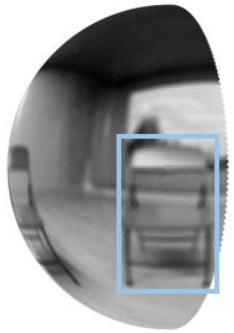
Barrel distortion



SCE Left View



SCE Front View

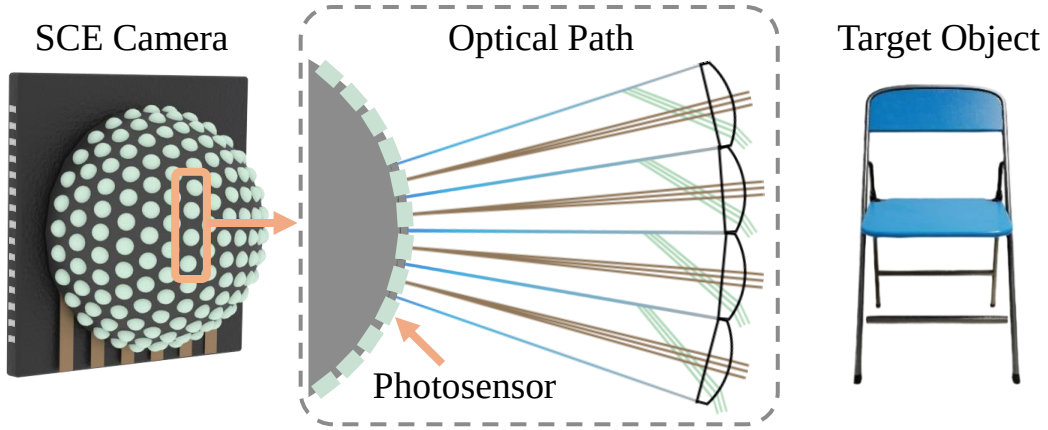


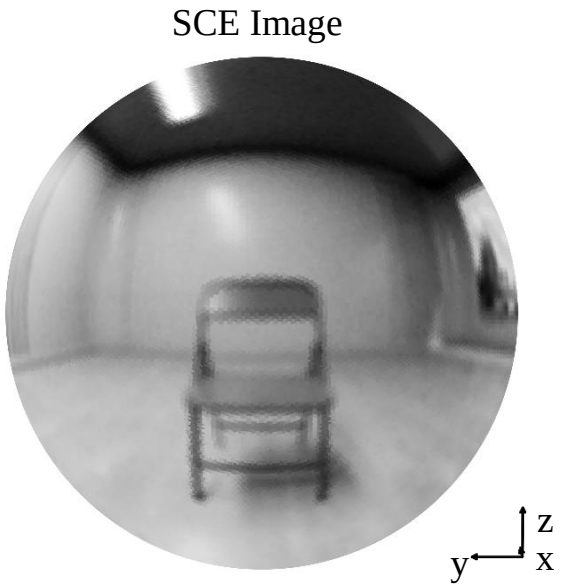
SCE Right View

No Distortion

Key Idea 1: Native Spherical Sensing

From ommatidia to a hemispherical image

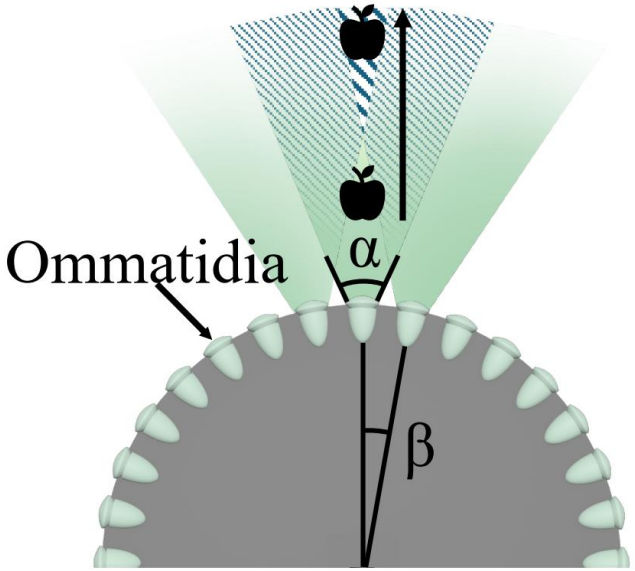


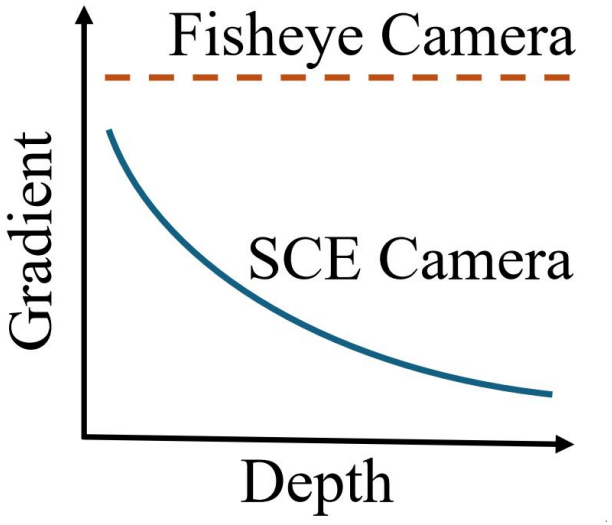


Spherical Compound Eye Structure

- Ommatidia are arranged on a hemispherical surface.
- Each ommatidium comprises a lens and a single-pixel photosensor, with optical isolation between adjacent units to eliminate cross-talk.
- Every ommatidium integrates light over its own FOV into one value (intensity).
- **All ommatidial measurements form a hemispherical image rather than a planar image.**

Key Idea 2: Depth-Sensitive Gradient Features





Depth Cues

- When the **ommatidial FOV α** is larger than the **angular spacing β**, adjacent ommatidia overlap.
- At different depths, an object edge contributes differently to neighboring ommatidia.
- The inter-ommatidial intensity difference $\Delta I(d)$ decreases as depth increases.

Depth-Gradient Formal Model

Adjacent ommatidia act as a single-pixel stereo analogue, where Δ*I(d)* replaces conventional disparity.

Ommatidial intensity $I_i(d)$

$$I_i(d) = \frac{1}{\alpha}\left[\int_{-\alpha/2}^{\theta_i(d)} L_1\, d\theta + \int_{\theta_i(d)}^{\alpha/2} L_2\, d\theta\right] = \frac{L_1 + L_2}{2} + \frac{L_1 - L_2}{\alpha}\,\theta_i(d)\ , i \in \{L, R\} \tag{1}$$

Inter-ommatidial gradient $\Delta I(d)$

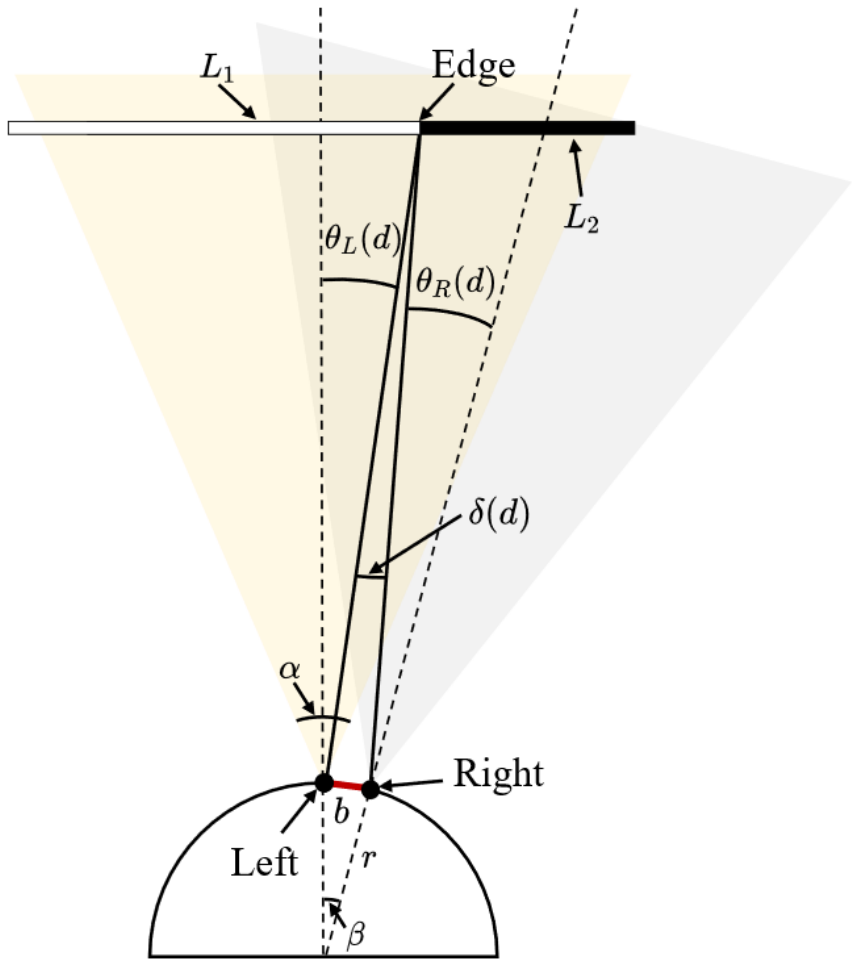
$$\Delta I(d) = |I_L(d) - I_R(d)| = \left|\frac{L_1 - L_2}{\alpha}\left(\theta_L(d) - \theta_R(d)\right)\right| \tag{2}$$

$$\theta_L(d) - \theta_R(d) = \beta + \delta(d) \tag{3}$$

$$\delta(d) \approx \frac{b}{d} \approx \frac{r\beta}{d} \quad (d \gg b) \tag{4}$$

$$\Delta I(d) = \left|\frac{L_1 - L_2}{\alpha}\left(\beta + \frac{r\beta}{d}\right)\right|$$

$$\Rightarrow \frac{\partial \Delta I(d)}{\partial d} = -\frac{|L_1 - L_2|\, r\beta}{\alpha} \cdot \frac{1}{d^2} \tag{5}$$

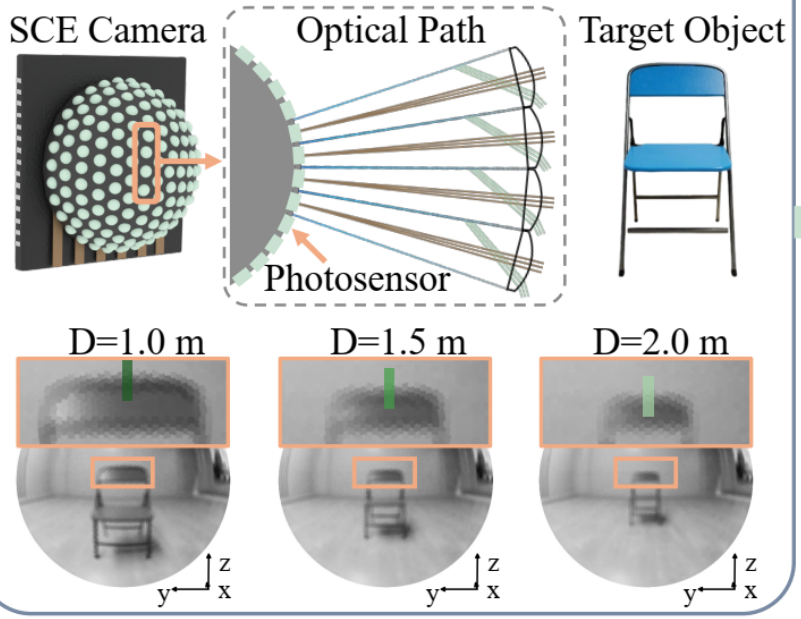


Δ*I(d)* decreases as depth increases.

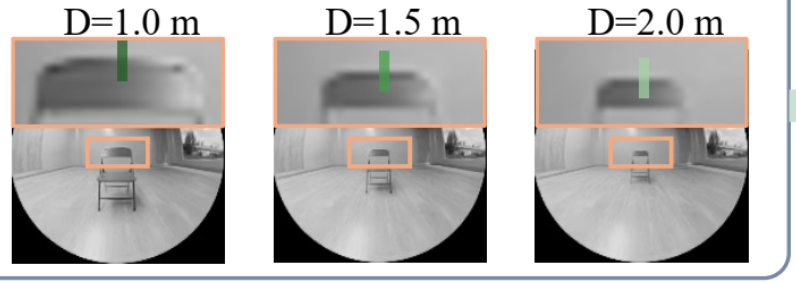
Simulation and Prototype Validation

Simulated SCE Camera

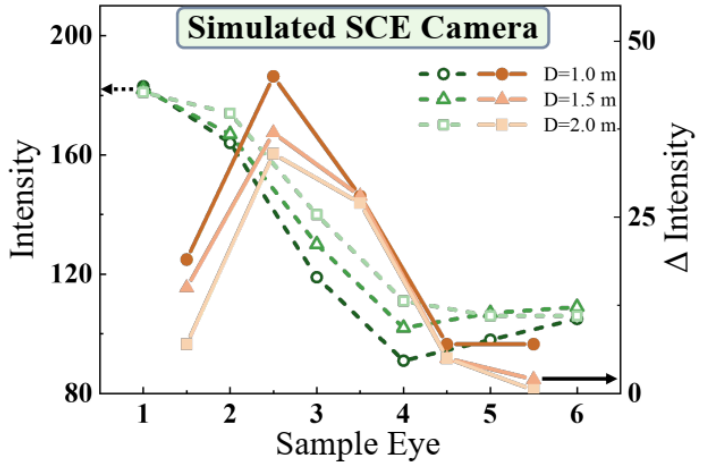
Depth-Dependent Blur in SCE Imaging

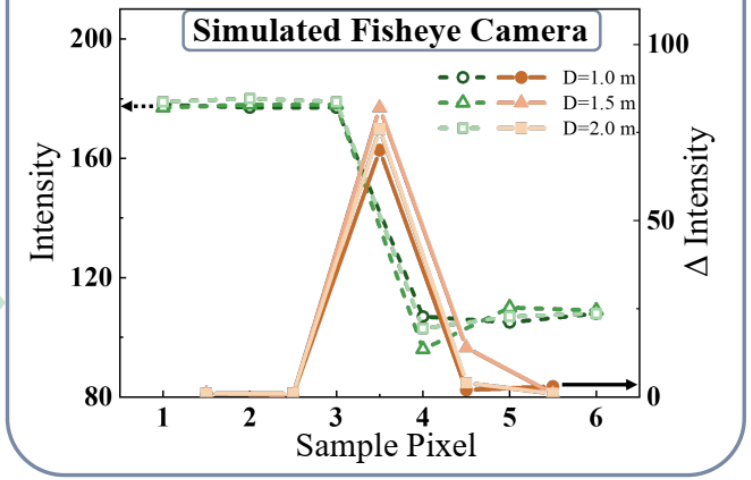


Depth-Independent Blur in Fisheye Imaging

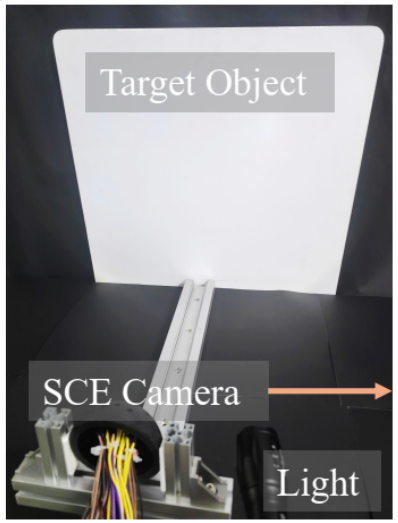


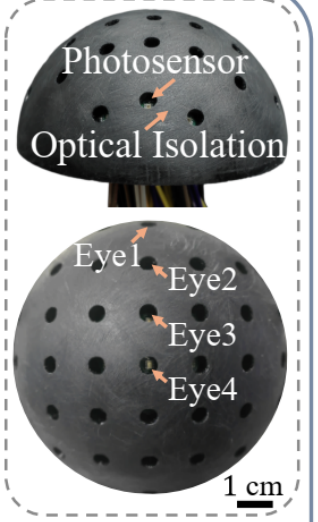
SCE Δ Intensity Decreases with Depth

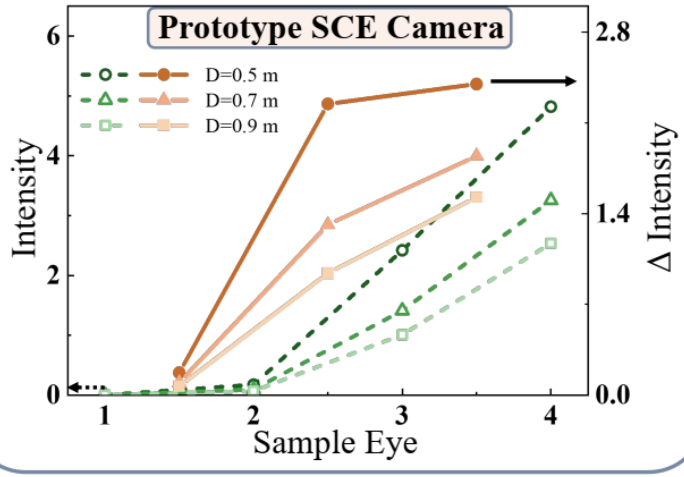




Prototype SCE Camera

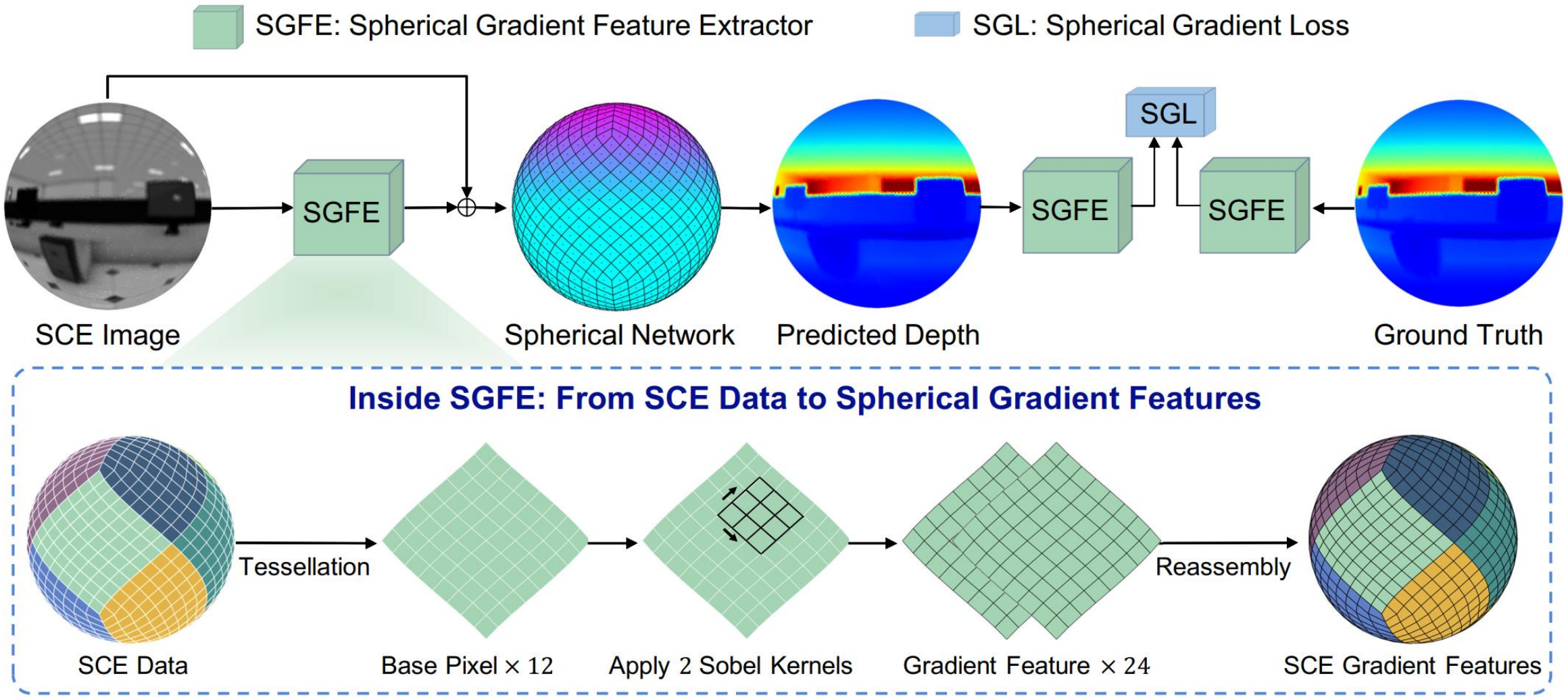






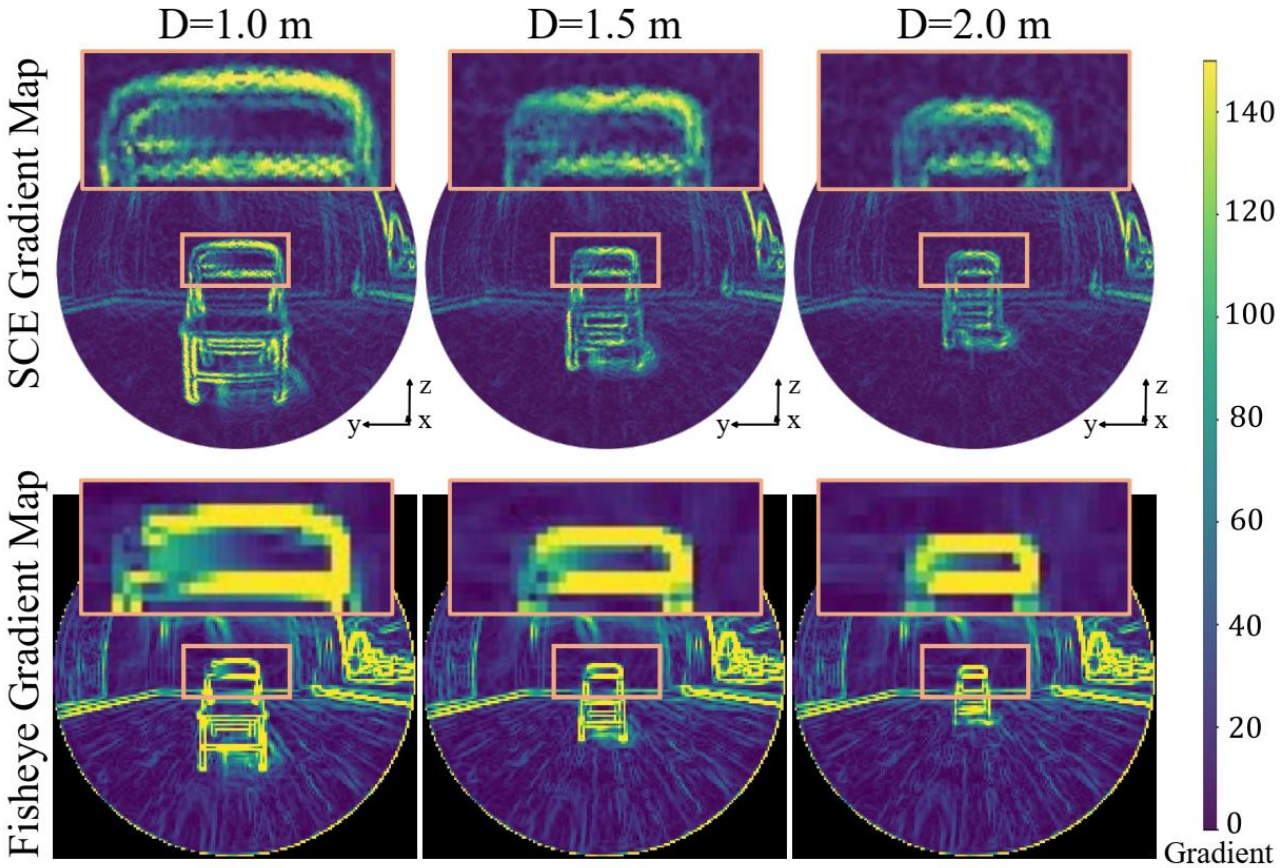
Overlapping FOVs create the depth-sensitive gradient features.

SCE-Depth Framework



$$\mathcal{L}_{\mathrm{SG}} = \mathrm{BerHu}(\nabla_\theta, \nabla_\theta^*) + \mathrm{BerHu}(\nabla_\phi, \nabla_\phi^*)$$

Visualization of the SGFE



SGFE Effectiveness

- SCE gradient maps exhibit clear attenuation as depth increases.
- Fisheye gradient maps remain nearly depth-invariant under the same Sobel operator.
- This confirms that SGFE captures depth-sensitive cues specific to SCE imaging geometry.

Datasets

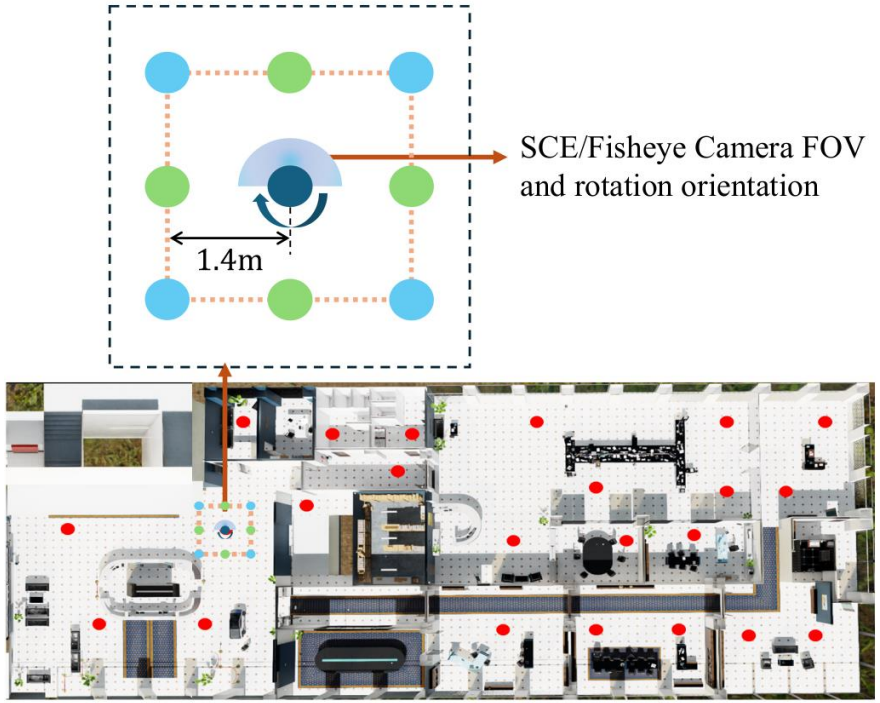
CompoundDepth and paired fisheye data under identical scenes

CompoundDepth Datasets

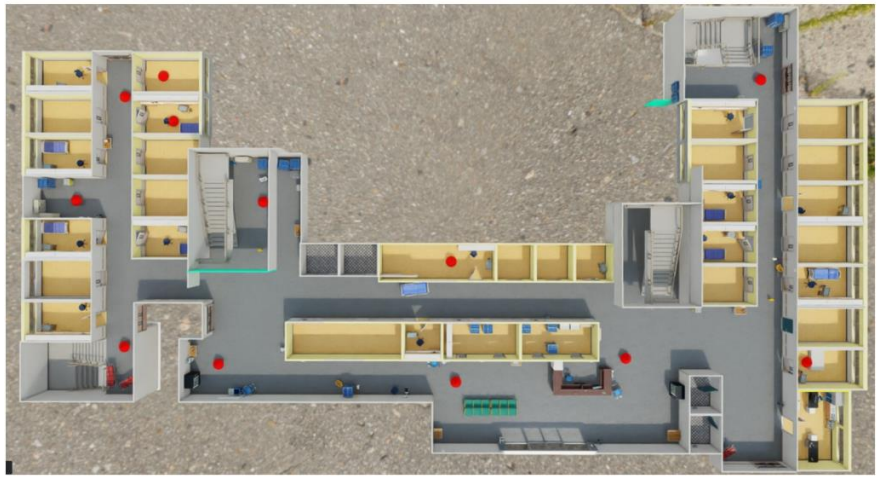
20,609
individual ommatidia per SCE image

5,456
spherical image-depth samples

4 FOVs
α = 1°, 3°, 5°, 7°



(a) Scene of Office



(b) Scene of Hospital

SCE-Depth Design Choices

Both the sensor representation and gradient operators matter

α = 3°

best clarity–depth trade-off

FOV(α)	Model	RMSE ↓	Abs Rel ↓	δ_1 ↑	δ_2 ↑
1°	HealSwin	0.383	0.104	92.08	96.15
	SUFormer	0.513	0.114	89.79	94.25
	Ours	0.279	0.061	94.90	97.83
3°	HealSwin	0.350	0.082	92.76	96.99
	SUFormer	0.429	0.105	90.31	94.76
	Ours	**0.256**	**0.050**	**95.74**	**98.48**
5°	HealSwin	0.376	0.107	90.85	95.80
	SUFormer	0.404	0.115	90.40	94.85
	Ours	0.289	0.074	93.75	97.34
7°	HealSwin	0.393	0.110	90.31	95.53
	SUFormer	0.386	0.114	90.95	94.91
	Ours	0.284	0.072	93.47	97.31

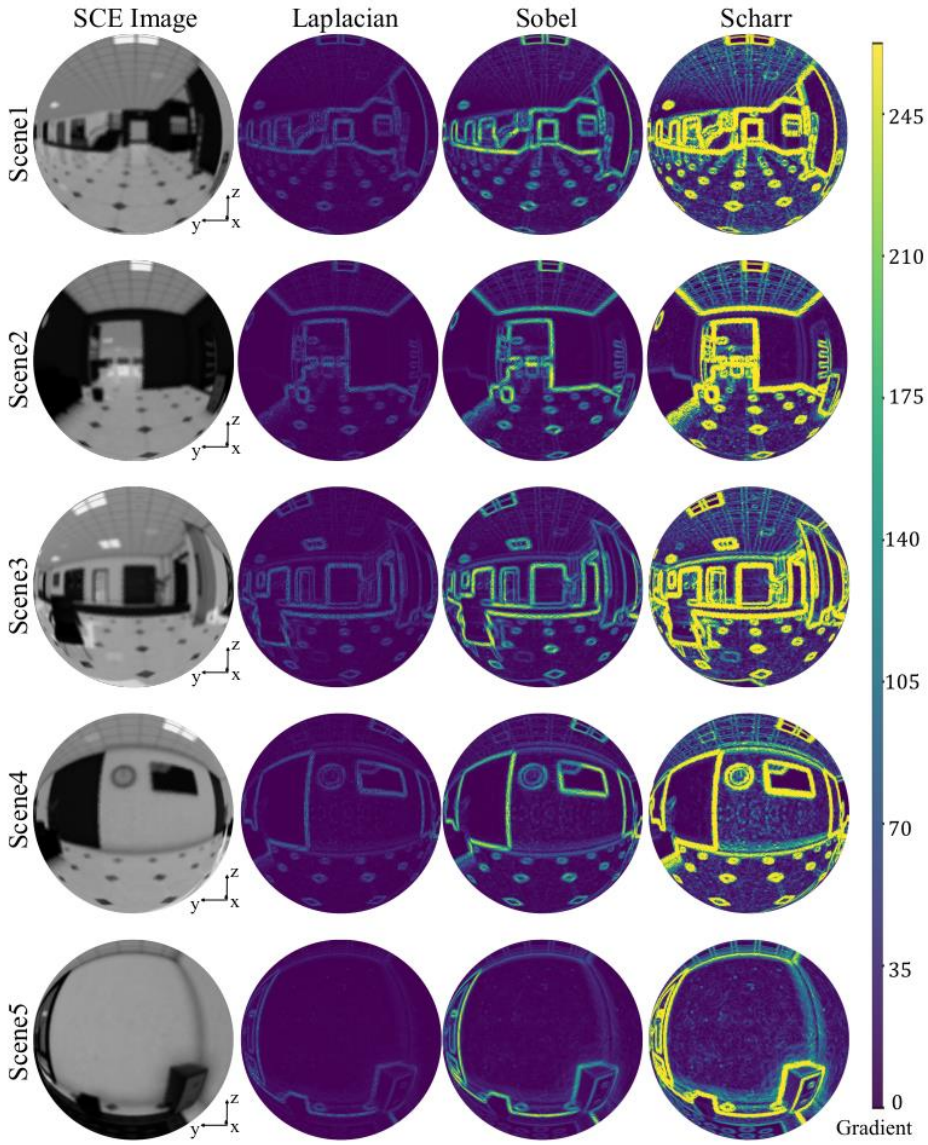
Table 1: Comparison of different ommatidial FOV.

Sobel

best gradient operator

Operator	RMSE ↓	Abs Rel ↓	δ_1 ↑	δ_2 ↑
Laplacian	0.295	0.068	94.30	97.71
Scharr	0.258	0.059	95.18	98.01
Sobel	**0.256**	**0.050**	**95.74**	**98.48**

Table 2: Comparison of different gradient operators.



Quantitative Results

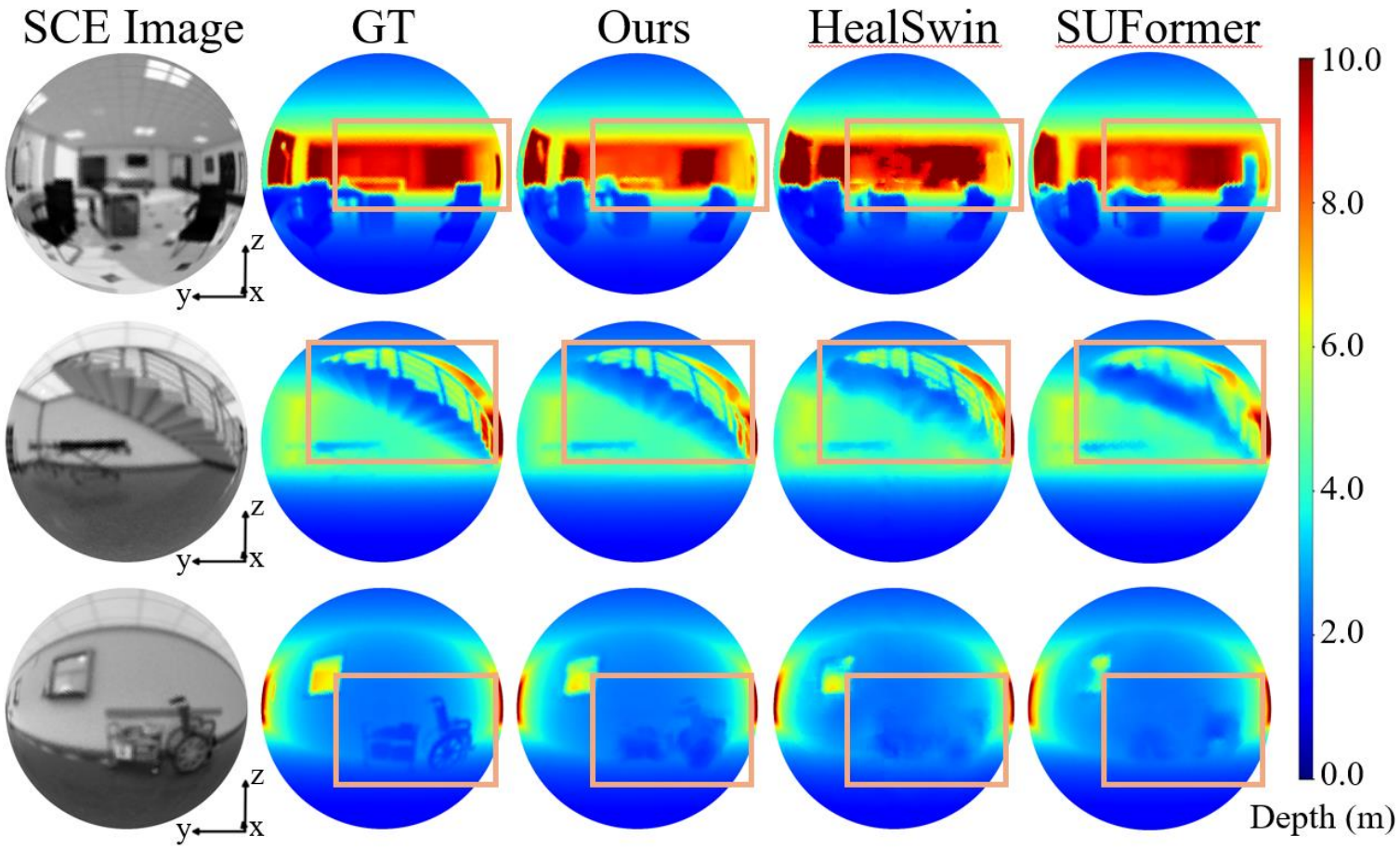
Model	#Paras(M)	FisheyeDepth				CompoundDepth ($\alpha = 3°$)			
		RMSE ↓	Abs Rel ↓	δ_1 ↑	δ_2 ↑	RMSE ↓	Abs Rel ↓	δ_1 ↑	δ_2 ↑
Swin [25]	41.3	0.324	0.095	91.68	96.31	0.376	0.113	90.80	95.74
DepthAnythingv2 [35]	24.8	0.335	0.096	91.38	96.20	0.313	0.061	94.59	98.04
PanDA [4]	24.8	0.354	0.070	93.64	97.55	0.283	0.064	94.35	97.86
DepthAnyCamera [15]	208.9	0.292	**0.064**	93.59	96.79	0.266	0.055	94.89	97.84
UGSCNN [17]	1.3	0.889	10.310	78.00	83.64	0.866	6.266	82.06	87.91
HealSwin [5]	41.3	0.361	0.106	91.83	96.28	0.350	0.082	92.76	96.99
SUFormer [3]	14.9	0.518	0.128	88.53	93.27	0.429	0.105	90.31	94.76
Ours	41.3	**0.287**	0.067	**94.68**	**97.66**	**0.256**	**0.050**	**95.74**	**98.48**

Table 3: Depth estimation performance on FisheyeDepth and CompoundDepth datasets.

SCE-Depth achieves the best depth accuracy on CompoundDepth.

Qualitative Results

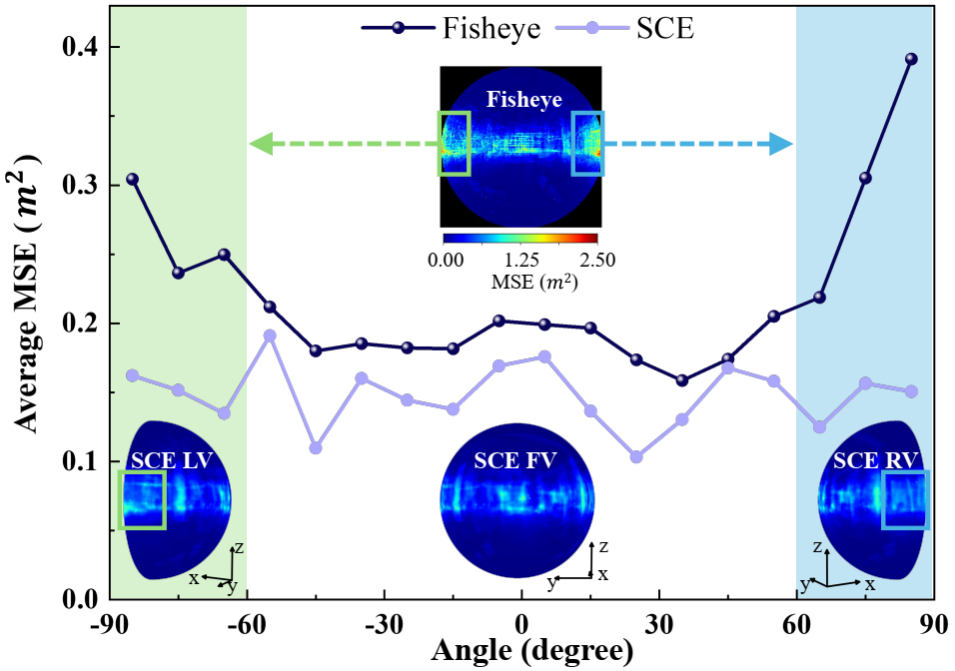
Sharper geometry and better edge/detail recovery



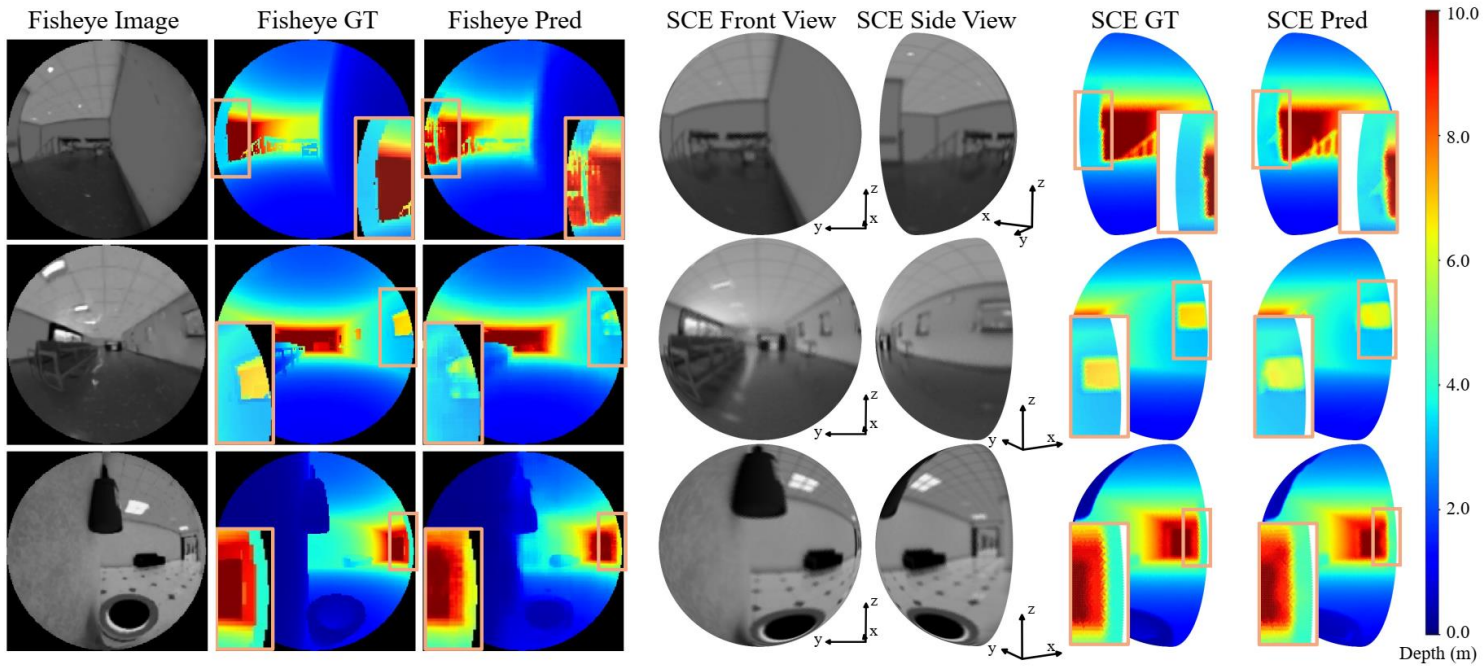
Omnidirectional Consistency

Native spherical imaging avoids fisheye distortion-induced errors

Quantitative Results



Qualitative Results



Generalization and Ablation

Model	SynWoodScape		Stanford2D3D	
	RMSE ↓	Abs Rel ↓	RMSE ↓	Abs Rel ↓
HealSwin	7.896	0.425	0.497	0.143
SUFormer	10.376	0.519	0.432	**0.071**
Ours	**7.606**	**0.297**	**0.418**	0.114

Table 4: Depth estimation performance on SynWoodScape and Stanford2D3D datasets.

SCE	SGL	SGFE	RMSE ↓	Abs Rel ↓	δ_1 ↑	δ_2 ↑
✗	✗	✗	0.361	0.106	91.83	96.28
✓	✗	✗	0.350	0.082	92.76	96.99
✓	✓	✗	0.281	0.064	94.35	97.65
✓	✓	✓	**0.256**	**0.050**	**95.74**	**98.48**

Table 5: Ablation study.

CVPR JUNE 3-7, 2026 DENVER COLORADO

Code & Datasets

SCE-Depth: A Spherical Compound Eye Framework for Wide FOV Depth Estimation

Yi Zhu[1,2,†]; Hao Xiong[1,†]; Lin Xiao[1]; Ranfeng Shi[1]; Qinying Gu[2]; Leilei Gu[1,2,*]
[1]Shanghai Jiao Tong University, [2]Shanghai Artificial Intelligence Laboratory

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

上海人工智能实验室 Shanghai Artificial Intelligence Laboratory