Reveal Object in Lensless Photography via Region Gaze and Amplification

Xiangjun Yin, Huihui Yue.

Nanyang Technological University

Motivation

Task: We propose RGANet for COD in lensless imaging, which uses spatial-frequency RGMs to localize objects and a local attention-based RA to enhance region details.

Challenges:

- Lensless imaging lacks traditional visual features, making it challenging to extract task-relevant information;
- Data complexity increases training difficulty, especially for denoising and key feature retention;
- The inherent complexity of the COD task further amplifies these challenges.

Motivation:

- Enhancing COD in lensless imaging requires reducing semantic clutter and capturing fine details.
- Frequency cues filter irrelevant features, while spatial proximity boosts detail perception, together improving object recognition.

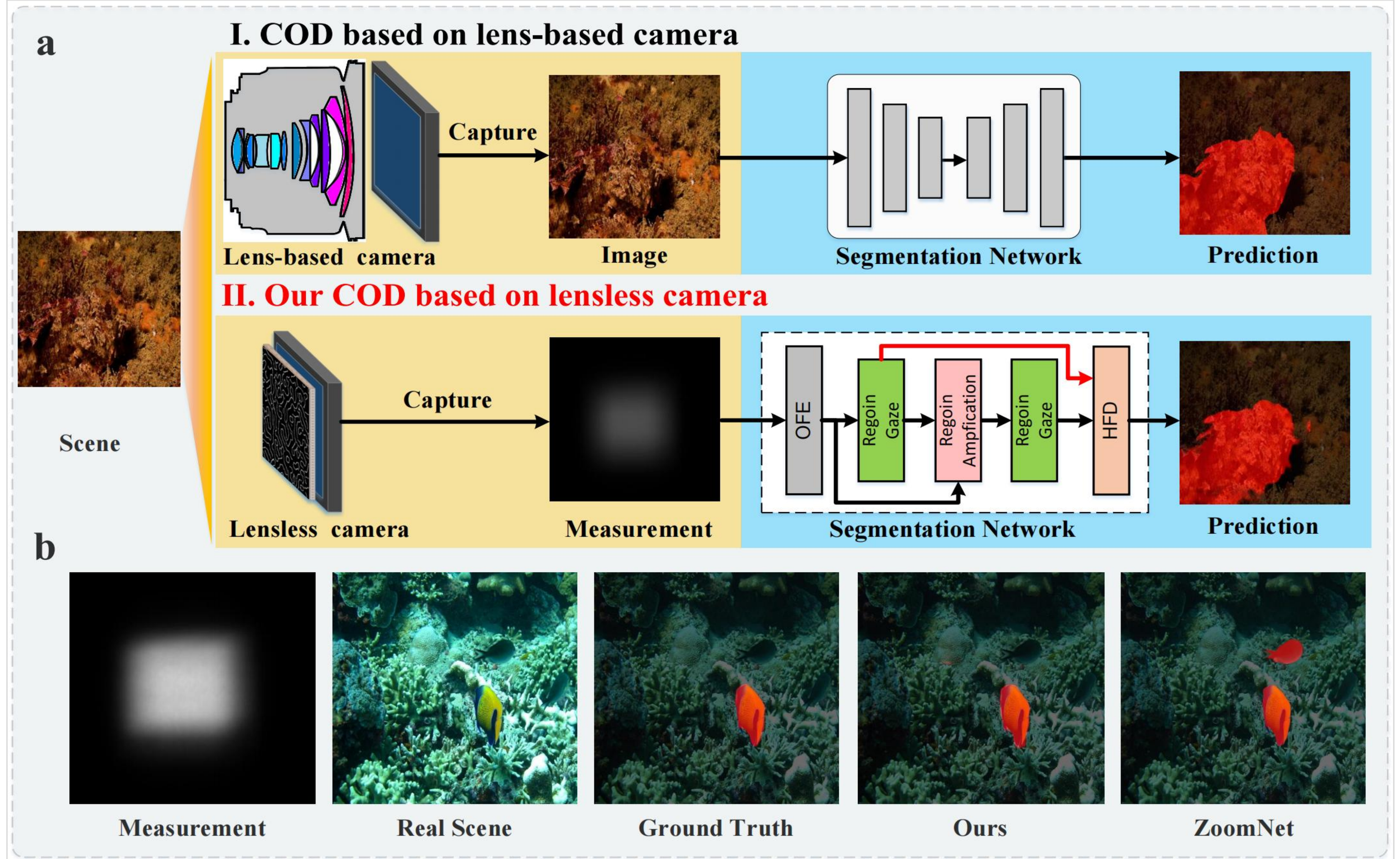


Fig 1 Motivations and Comparison Results

Method

Overview: Our RGANet includes an OFE for feature extraction, two RGMs to mine spatial-frequency cues, an RA to enhance object details, and an HFD for refinement. This enables accurate object detection from lensless data.

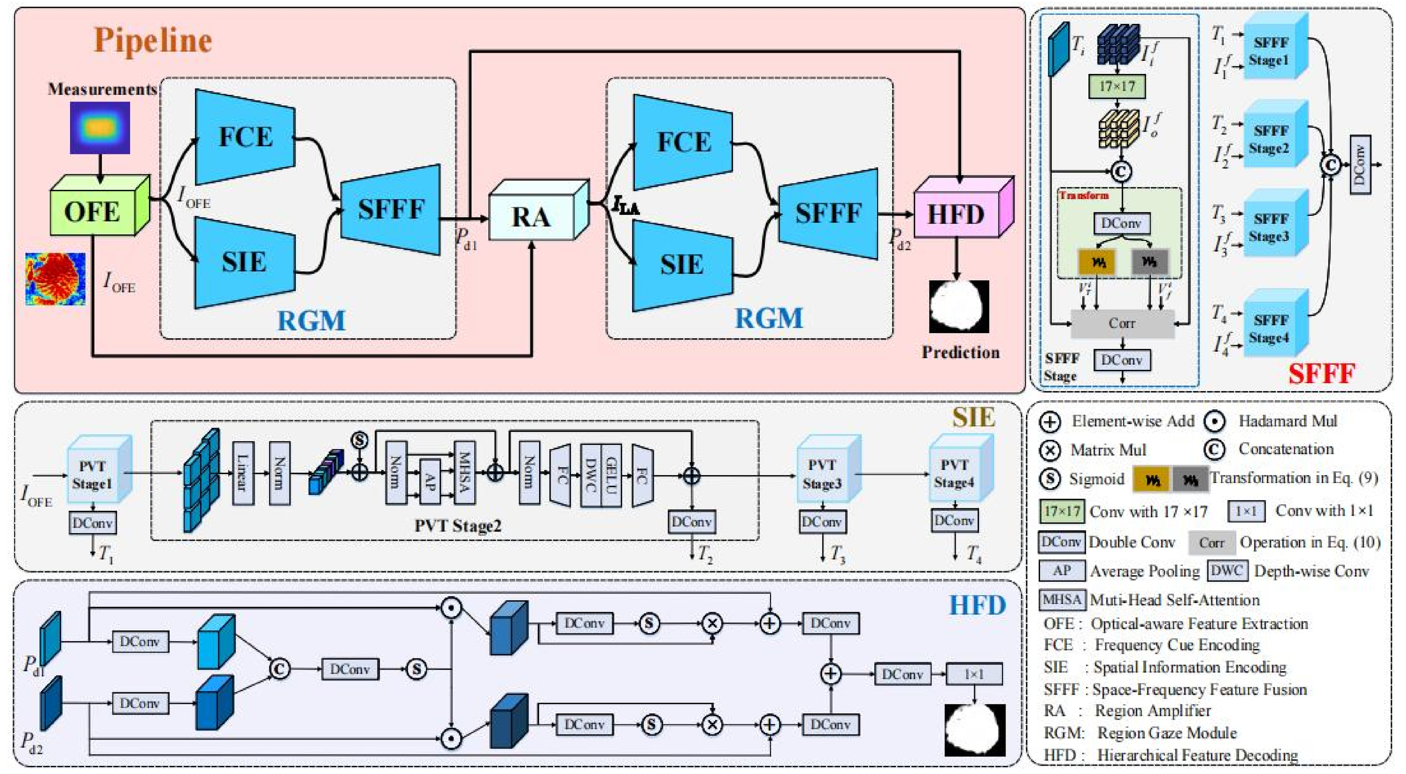


Fig 2 The pipeline of RGANet

OFE: OFE with a Wiener filtering mechanism as

$$I_{\mathrm{OFE}} = \mathcal{F}^{-1}\left(\frac{\mathrm{Conj}\left(\mathcal{F}(A_\theta)\right)}{K_\theta + |\mathcal{F}(A_\theta)|^2} \odot \mathcal{F}(Y)\right)$$

RGM: RGM that learns spatial and frequency features collaboratively by SIE, FCE, and SFFF.

$$\begin{cases} (F^{L}_{i,j})^c = F^c_{i,j}(m,n), & |m - o_m| \le r, |n - o_n| \le r \\ (F^{H}_{i,j})^c = F^c_{i,j}(m,n), & \text{otherwise} \end{cases}$$

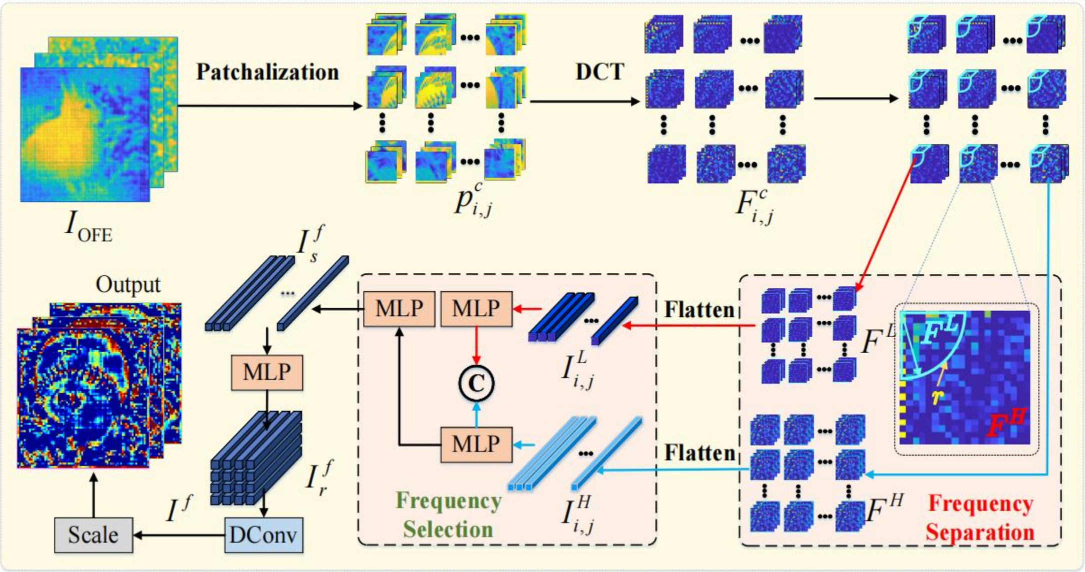


Fig 3 The structure of FCE

It includes DCT, frequency separation, and selection to gather frequency cues.

RA: RA compresses background and amplify concealed objects by generating an attention map from first RGM output and magnifying objects based on this map:

$$M_x(n) = \sum_{s=1}^{n} \max_{1\le t\le W} M_{t,s}$$

$$M_y(n) = \sum_{t=1}^{n} \max_{1\le s\le H} M_{t,s},$$

$$\mathcal{Q}\left(I_{\mathrm{OFE}}, M\right)_{t,s} = \left(I_{\mathrm{OFE}}\right)_{M_x^{-1}(t), M_y^{-1}(s)}$$

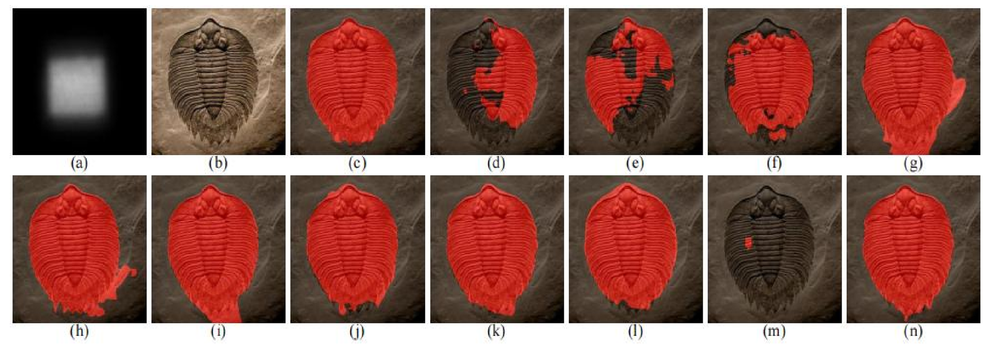
$$I_{\mathrm{RA}} = \mathcal{Q}\left(I_{\mathrm{OFE}}, M\right)$$

Loss Functions: We combine the weighted BCE loss and weighted IoU loss for effective training.

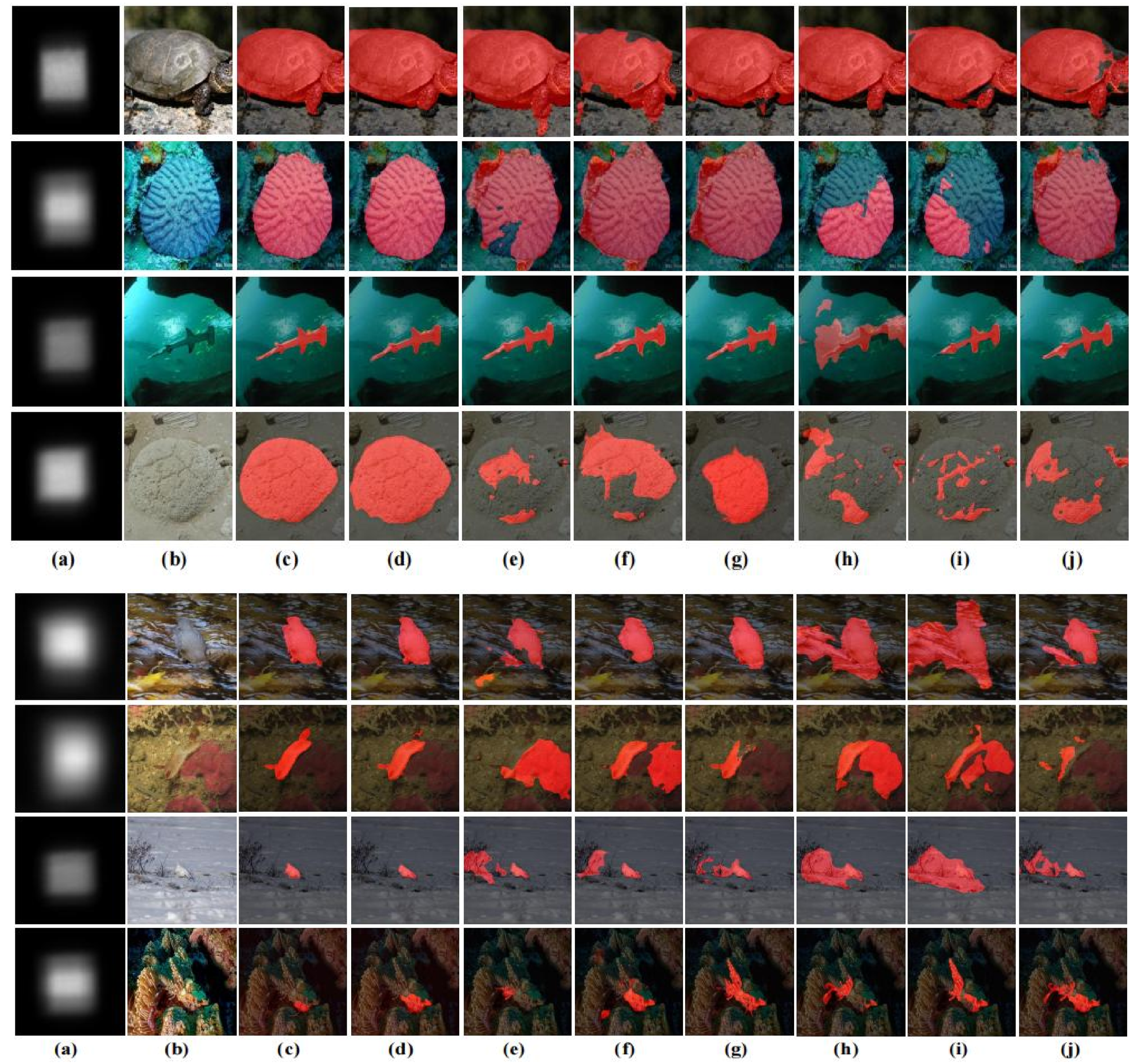
$$L_{\mathrm{All}} = L_{\mathrm{s}}(P_{\mathrm{d1}}, P_{\mathrm{gt}}) + L_{\mathrm{s}}(P_{\mathrm{d2}}, P_{\mathrm{gt}}) + L_{\mathrm{s}}(P_{\mathrm{final}}, P_{\mathrm{gt}}) \quad L_{\mathrm{s}} = L_{\mathrm{wBCE}} + L_{\mathrm{wIOU}}$$

Experiments

ID	OFE	1-st RGM			RA	2-nd RGM			HFD	Test-Easy		Test-Hard	
		FCE	SIE	SFFF		FCE	SIE	SFFF		$F^w_\beta \uparrow$	$\mathcal{M} \downarrow$	$F^w_\beta \uparrow$	$\mathcal{M} \downarrow$
#1	✓	✓	✓	✓						0.509	0.259	0.468	0.306
#2	✓		✓	✓			✓	✓		0.624	0.163	0.511	0.182
#3	✓		✓	✓	✓		✓	✓		0.651	0.139	0.557	0.158
#4	✓		✓	✓	✓		✓	✓	✓	0.682	0.134	0.564	0.153
#5	✓	✓	✓	✓	✓		✓	✓	✓	0.721	0.122	0.596	0.147
#6	✓		✓	✓	✓	✓	✓	✓	✓	0.718	0.125	0.592	0.145
#7	✓	✓	✓	✓		✓	✓	✓		0.729	0.116	0.601	0.142
#8	✓	✓	✓	✓	✓	✓	✓	✓		0.795	0.087	0.672	0.116
#9	✓	✓	✓	✓		✓	✓	✓	✓	0.756	0.106	0.617	0.134
#10		✓	✓	✓	✓	✓	✓	✓	✓	0.423	0.382	0.392	0.361
#11	✓				✓	✓	✓	✓	✓	0.631	0.157	0.539	0.162
Ours	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.815	0.079	0.705	0.098

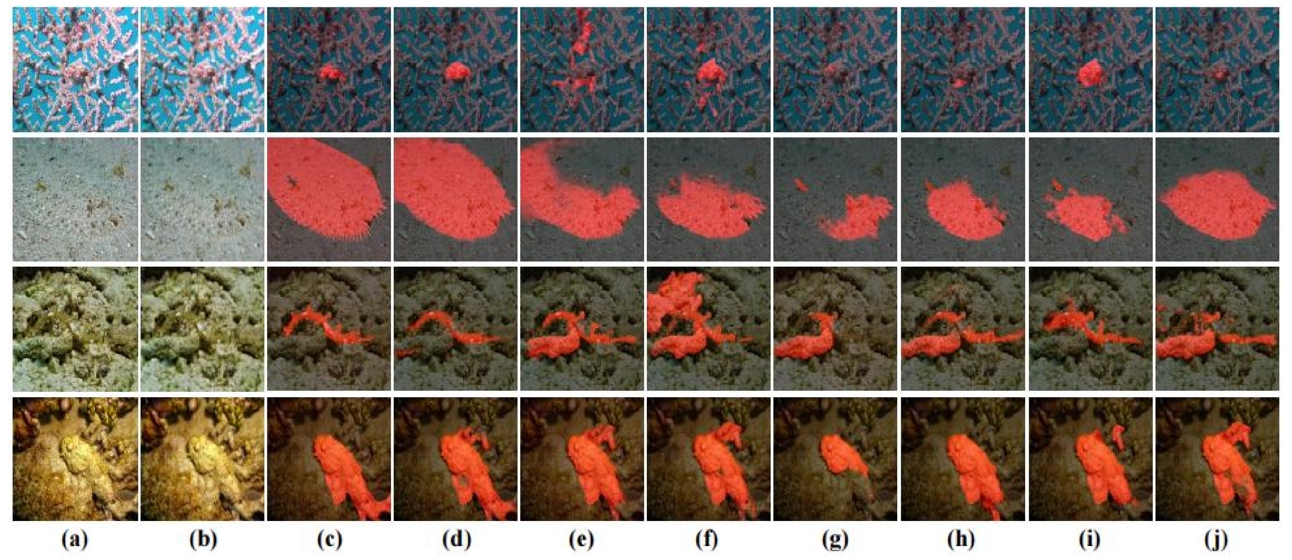


Ablation study on different configurations (d)-(m) correspond to IDs #1-#10, (n) is RGANet, and (c) is label map for lensless imaging measurement (a) and underlying scene (b)



Method	FLOPs (G)	#Param (M)	Test-Easy				Test-Hard			
			$F^w_\beta \uparrow$	$\mathcal{M} \downarrow$	$E_\xi \uparrow$	$S_\alpha \uparrow$	$F^w_\beta \uparrow$	$\mathcal{M} \downarrow$	$E_\xi \uparrow$	$S_\alpha \uparrow$
EyeCoD	84.37	26.92	0.712	0.131	0.819	0.791	0.563	0.162	0.745	0.710
LLI_T	44.35	17.23	0.743	0.110	0.832	0.802	0.527	0.167	0.741	0.651
LOINet	6.42	25.31	0.762	0.103	0.853	0.821	0.624	0.122	0.779	0.733
MSCAF-Net	63.04	30.32	0.697	0.131	0.812	0.788	0.563	0.161	0.790	0.710
OCENet	13.32	55.01	0.623	0.163	0.851	0.769	0.511	0.182	0.811	0.709
ZoomNet	39.41	32.58	0.714	0.126	0.821	0.782	0.619	0.121	0.804	0.717
Ours	48.62	39.45	0.815	0.079	0.896	0.834	0.705	0.098	0.845	0.770

We compare our RGANet with lensless inference-based methods (EyeCoD, LLI_T, LOINet) and state-of-the-art COD methods (MSCAF-Net, OCENet, ZoomNet). All models are retrained using open-source codes and a consistent OFE module for fair evaluation. Results on the Test-Easy and Test-Hard datasets, highlighting our method's ability to infer more complete object structures.



Method	FLOPs (G)	#Param (M)	Test-Easy				Test-Hard			
			$F^w_\beta \uparrow$	$\mathcal{M} \downarrow$	$E_\xi \uparrow$	$S_\alpha \uparrow$	$F^w_\beta \uparrow$	$\mathcal{M} \downarrow$	$E_\xi \uparrow$	$S_\alpha \uparrow$
FlatNet + EyeCoD	204.27	86.32	0.810	0.085	0.823	0.807	0.708	0.091	0.832	0.794
FlatNet + LLI_T	164.25	76.63	0.836	0.063	0.887	0.859	0.729	0.075	0.847	0.835
FlatNet + LOINet	126.32	84.71	0.843	0.054	0.897	0.868	0.751	0.063	0.866	0.827
FlatNet + MSCAF-Net	182.94	89.72	0.831	0.071	0.889	0.851	0.737	0.078	0.856	0.804
FlatNet + OCENet	133.22	114.41	0.829	0.057	0.876	0.853	0.741	0.071	0.852	0.816
FlatNet + ZoomNet	159.31	91.98	0.847	0.051	0.903	0.871	0.752	0.059	0.869	0.831
FlatNet + Ours	48.62	39.45	0.815	0.079	0.896	0.834	0.705	0.098	0.845	0.770

Results for the detection-after-reconstruction strategy with 10% improvment compared to direct COD methods, despite higher computational cost, validating the potential of direct COD in lensless imaging.