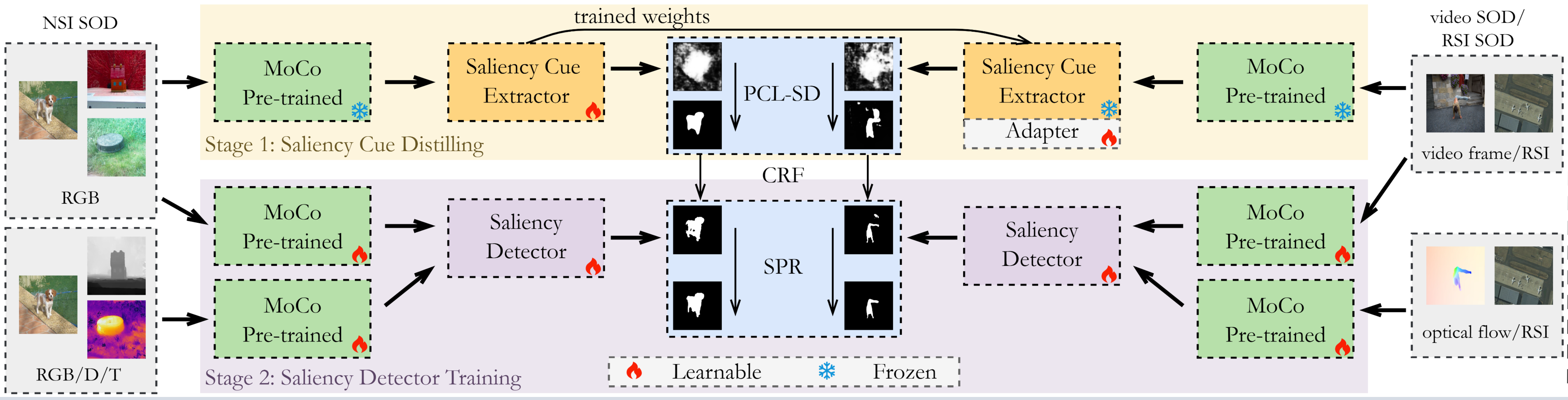
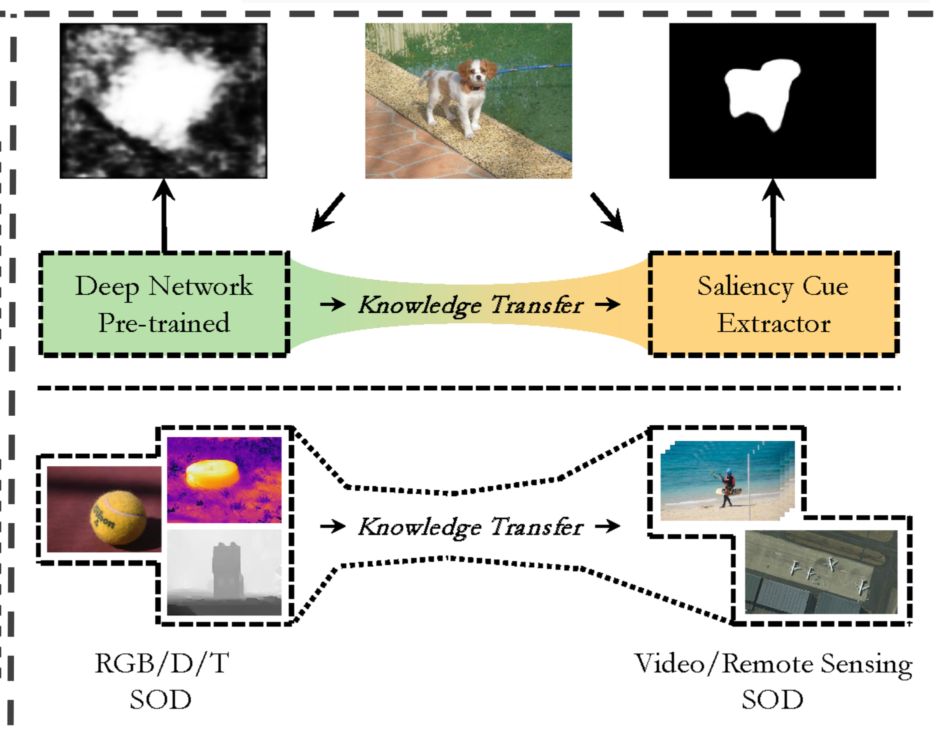
Unified Unsupervised Salient Object Detection via Knowledge Transfer

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Yao Yuan, Wutao Liu, Pan Gao*, Qun Dai and Jie Qin*
Nanjing University of Aeronautics and Astronautics

IJCAI JEJU 2024





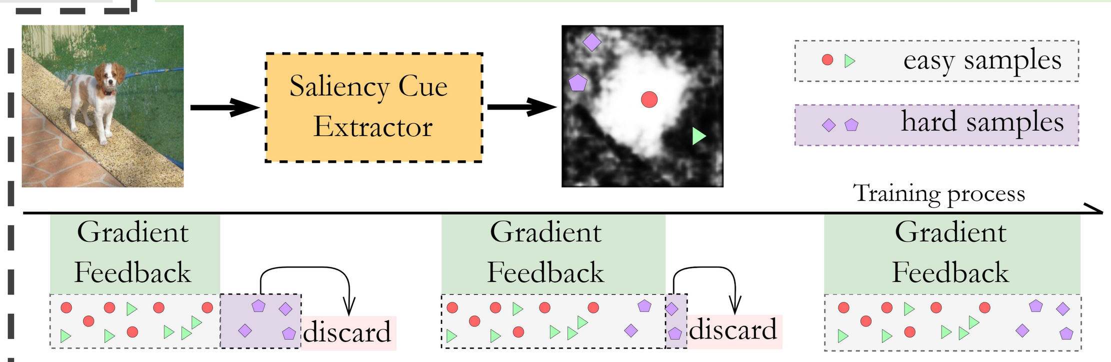
We propose a two-stage framework for unified USOD tasks. In stage 1, we train a saliency cue extractor (SCE) to transfer saliency knowledge from a pre-trained deep network. In stage 2, we utilize the obtained saliency cues as initial pseudo-labels to train a saliency detector (SD). We train our base model on Natural Still Image datasets and subsequently transfer the model to non-NSI SOD tasks such as video SOD and RSI SOD.

Challenges for Unsupervised Salient Object Detection (USOD):

1. **Obtaining initial saliency cues** is crucial for unsupervised SOD. Cues obtained via handcuffed methods are of limited quality, while those based on deep learning suffer from training instability issues.
2. **Enhancing pseudo-label quality** can effectively improve model performance, an area largely unexplored by current USOD methods.
3. **Improving method generalization** across multiple SOD tasks such as RGB SOD and RGB-D SOD remains a challenge, as existing methods are mostly limited to single-task scenarios.

Main Contributions of the proposed method:

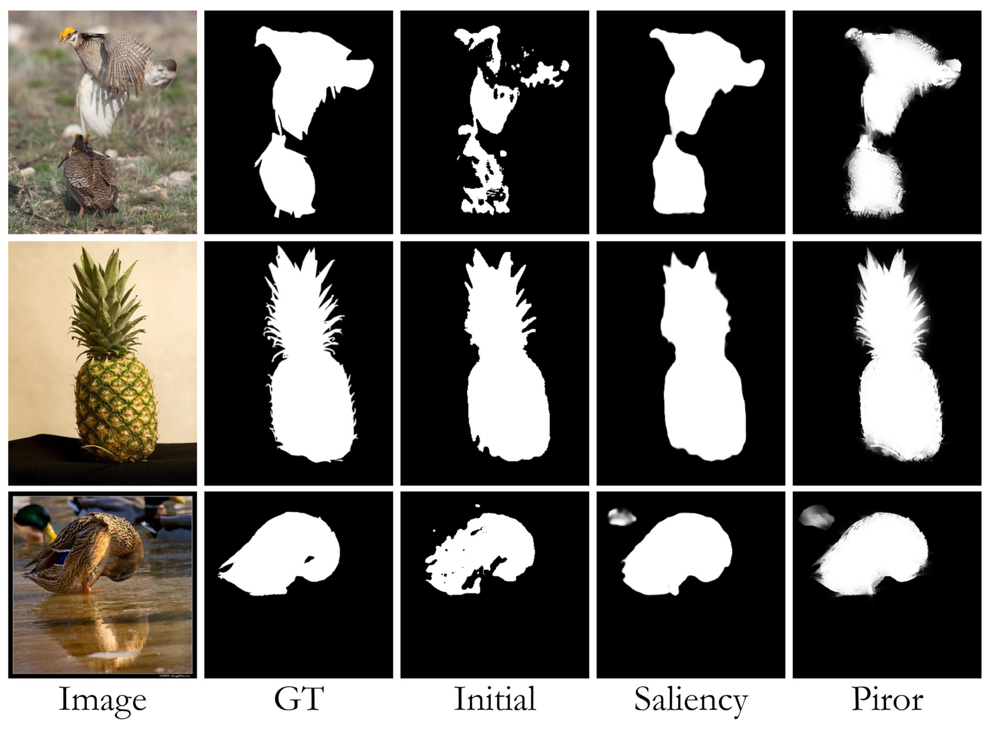
1. We propose a **Progressive Curriculum Learning-based Saliency Distilling** (PCL-SD) mechanism to extract saliency cues from a pre-trained deep network. This mechanism starts with easy samples and progressively moves towards harder ones, effectively mitigates the accumulation of errors, leading to a more stable and robust training process. The following formula shows how we use the mask matrix M to remove difficult samples.



$$M(i) = \begin{cases} 0 & if\,|S(i) - 0.5| < p, \\ 1 & otherwise \end{cases}$$

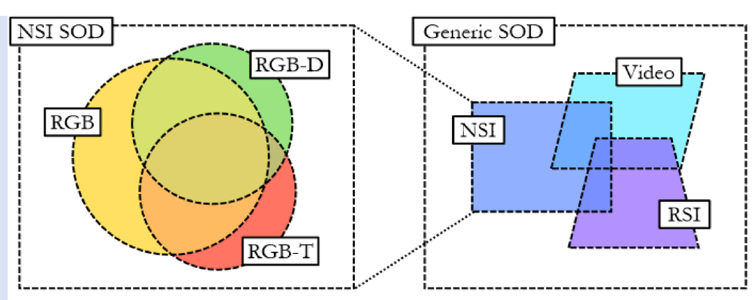
$$L_{pcl-sd} = 0.5 - \frac{1}{N}\sum_{i}^{N} |M(i) \odot S(i) - 0.5|$$

2. We design a **Self-rectify Pseudo-label Refinement** (SPR) mechanism to improve the quality of pseudo-labels. This mechanism combines prior rectification from a real-time pixel refiner and posterior rectification from the saliency prediction of model, gradually improving the quality of pseudo labels during the training process, demonstrating strong self-supervised performance. The formula on the right shows our process of updating pseudo labels G_{ref}.



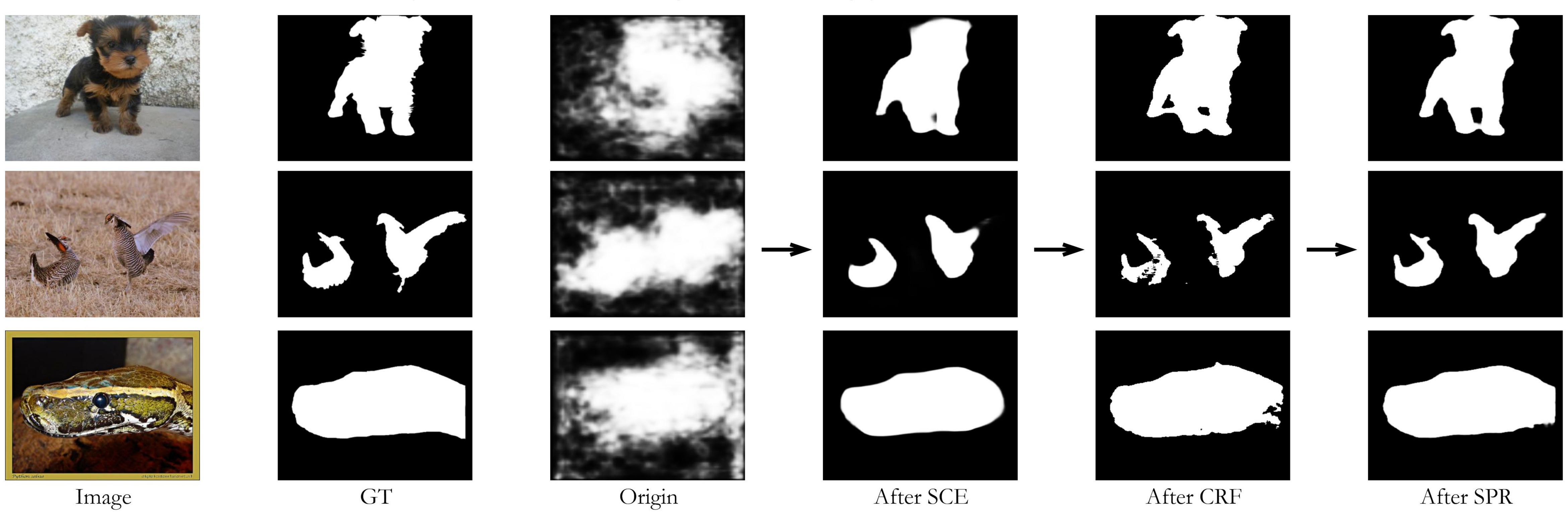
$$G_{ref} = \lambda_1 R_{pri} + \lambda_2 R_{post} + \lambda_3 G_{pre}$$

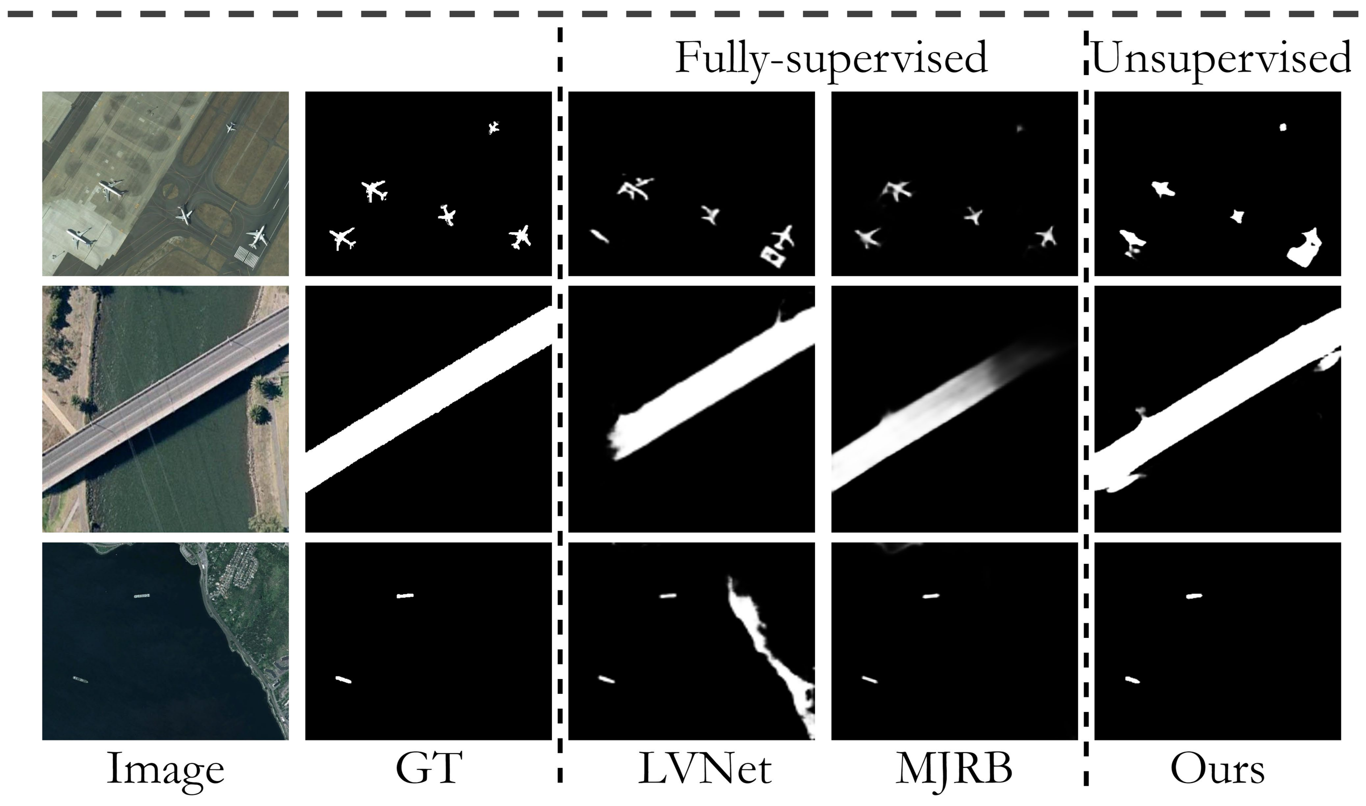
3. We devise an **adapter-tuning** method to transfer the acquired saliency knowledge, leveraging shared knowledge to attain superior transferring performance on the target tasks. Specifically, we employed an Adapter structure to fine-tune the deep layers of the model, allowing it to avoid experiencing performance degradation.



In contrast to NSI SOD, Generic SOD tasks exhibit a lower degree of shared knowledge.

The process of obtaining and refining pseudo labels.

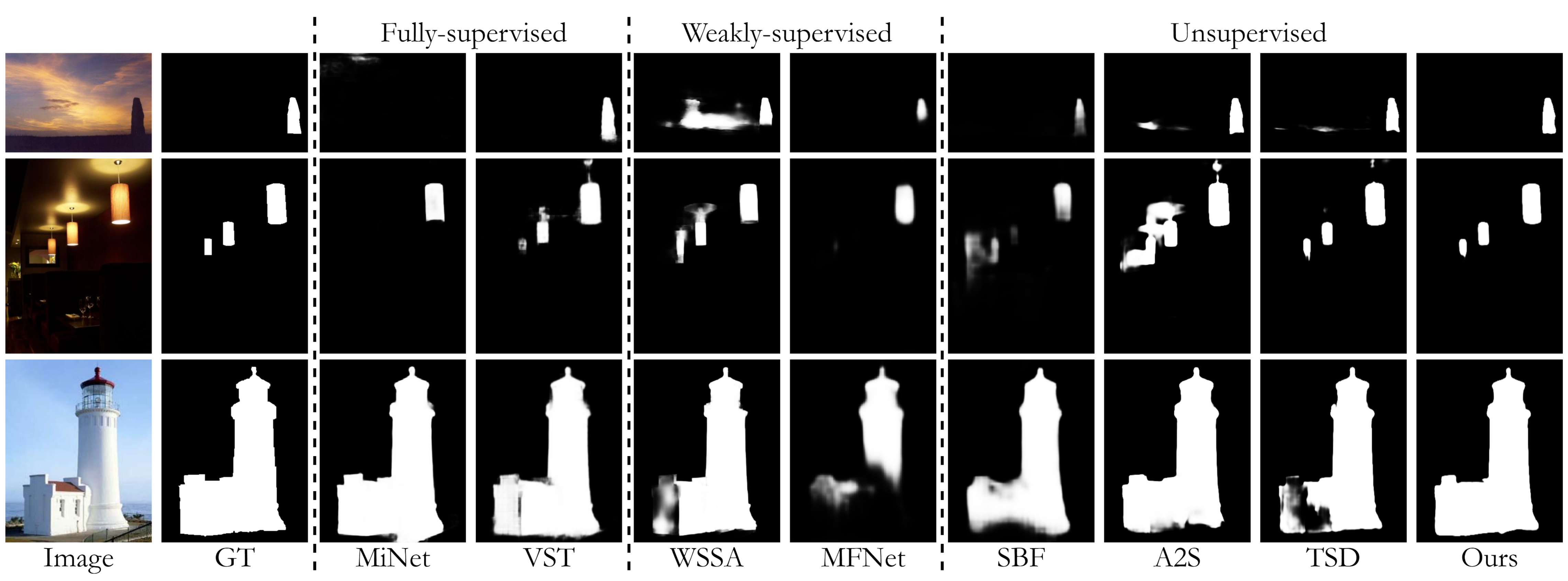




Visual comparison on RSI SOD datasets.

Quantative and Visual comparison on RGB SOD datasets.

dataset			DUT-O			DUTS-TE			ECSSD			HKU-IS			PASCAL-S		
Method	Year	Sup.	$M\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$M\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$M\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$M\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$M\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$
MINet	2020	F	**.055**	**.756**	**.873**	**.037**	**.828**	**.917**	**.033**	**.924**	.953	**.028**	**.908**	**.961**	.064	**.842**	.899
VST	2021	F	.058	**.756**	.872	**.037**	.818	.916	**.033**	.92	**.957**	.029	.9	.96	**.061**	.829	**.902**
WSSA	2020	W	**.068**	**.703**	**.845**	**.062**	**.742**	**.869**	**.047**	**.860**	**.932**	**.059**	**.870**	**.917**	**.096**	**.785**	**.855**
MFNet	2021	W	.098	.621	.784	.079	.693	.832	.058	.839	.919	.084	.844	.889	.115	.756	.824
EDNS	2020	U	.076	.682	.821	.065	.735	.847	.068	.872	.906	.046	.874	.933	.097	.801	.846
SelfMask	2022	U	.078	.668	.815	.063	.714	.848	.058	.856	.920	.053	.819	.915	.087	.774	.856
DCFD	2022	U	.070	.710	.837	.064	.764	.855	.059	.888	.915	.042	.889	.935	.090	.795	.860
TSD	2023	U	**.061**	.745	.863	.047	.810	.901	.044	.916	.938	.037	.902	.947	.074	.830	.882
STC	2023	U	.068	.753	.852	.052	.809	.891	.050	.903	.935	.041	.891	.942	.076	.827	.881
Ours$_{t.s.}$	-	U	.063	.749	.864	**.046**	.814	**.906**	**.038**	.922	.95	.034	.905	.953	**.068**	.841	.898
Ours	-	U	.062	**.759**	**.868**	.047	**.816**	**.906**	**.038**	**.923**	**.951**	**.033**	**.908**	**.954**	.069	**.844**	**.899**



Ablation studies on SPR and supervision strtagegy.

Refine Settings			RGB		RGB-D		RGB-T	
G_{res}	R_{pri}	R_{post}	$M\downarrow$	$F_\beta\uparrow$	$M\downarrow$	$F_\beta\uparrow$	$M\downarrow$	$F_\beta\uparrow$
✓	✗	✗	.04	.918	.064	.825	.03	.923
✓	✗	✓	.034	.925	.048	.868	.022	.951
✓	✓	✓	**.033**	**.927**	**.047**	**.87**	**.020**	**.953**

Loss Settings	RGB		DUTS-TE		NLPR	
	$M\downarrow$	$F_\beta\uparrow$	$M\downarrow$	$F_\beta\uparrow$	$M\downarrow$	$F_\beta\uparrow$
w/o PCL-SD	**.044**	.895	.077	.7	.05	.757
w/ PCL-SD	**.044**	**.896**	**.074**	**.713**	**.047**	**.77**
$\mathcal{L}_{bce}$	.034	.924	.050	.784	.033	.84
$\mathcal{L}_{iou}$	.034	.926	.049	.806	.029	.866
$\mathcal{L}_{iou}+\mathcal{L}_{bce}$	**.033**	**.928**	.049	.799	.032	.851
$\mathcal{L}_{iou}+\mathcal{L}_{ms}$	**.033**	.927	**.047**	**.816**	**.028**	**.871**

To get more information, you can

Check code or **Contact me**