Milano 2024

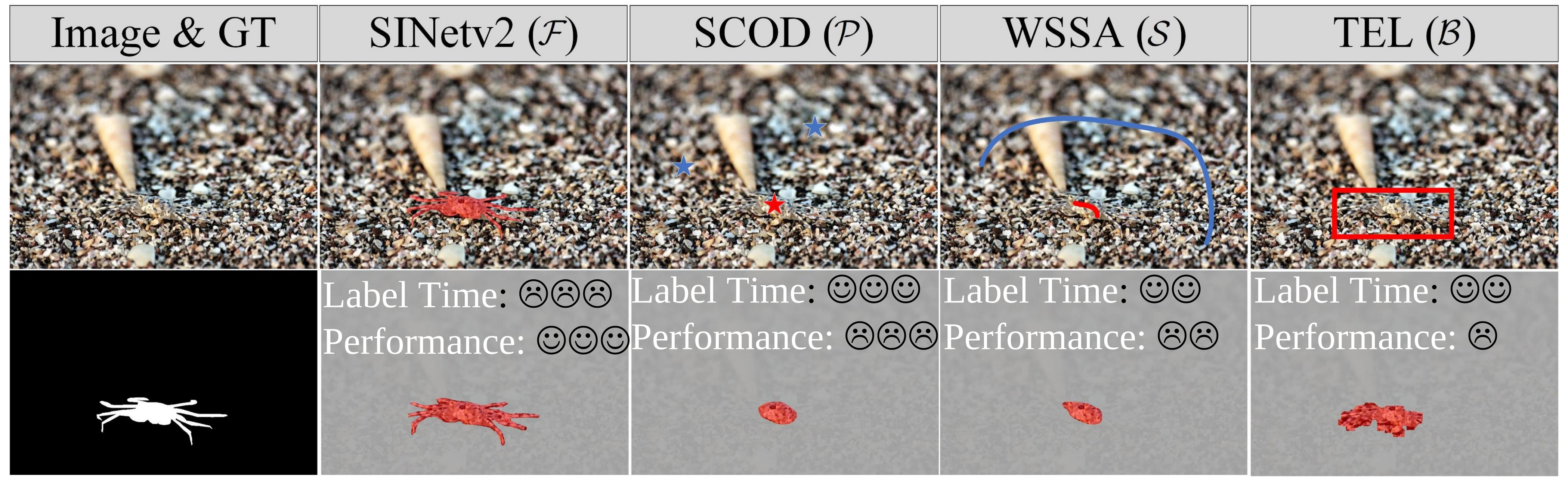
Learning Camouflaged Object Detection from Noisy Pseudo Label

Jin Zhang[1] Ruiheng Zhang[1] Yanjiao Shi[2] Zhe Cao[1] Nian Liu[3] Fahad Shahbaz Khan[3,4]

[1]Beijing Institute of Technology, China [2]Shanghai Institute of Technology, China [3]Mohamed bin Zayed University of Artificial Intelligence, UAE [4]Linköping University,Linköping, Sweden

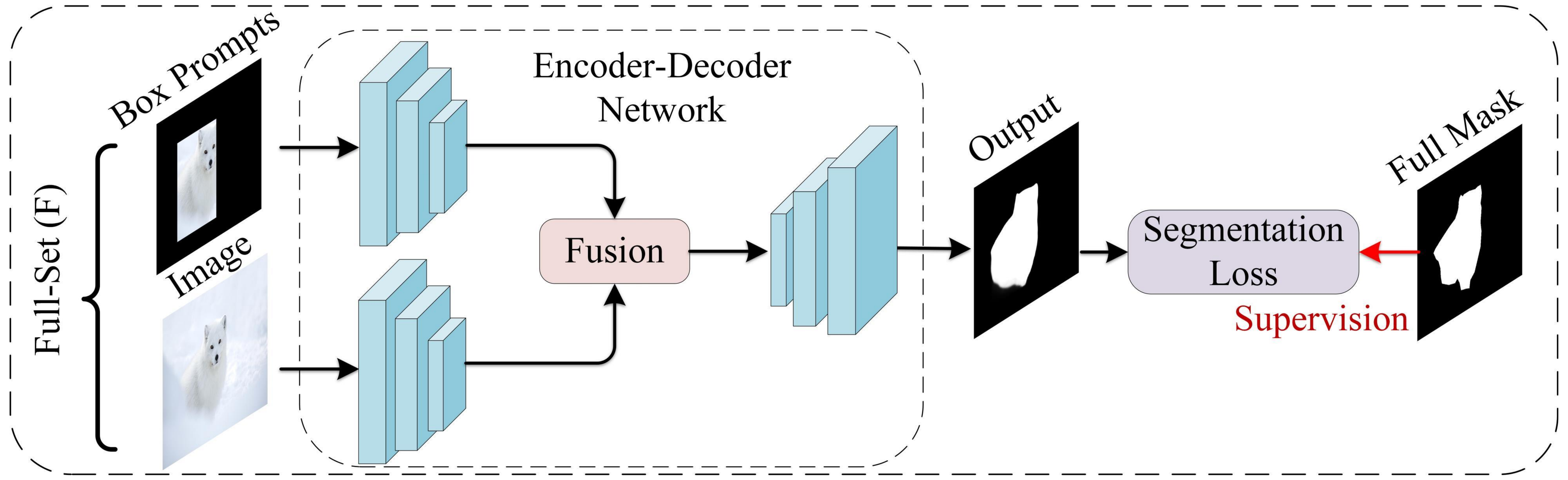
ABSTRACT & INTRODUCTION

In this paper, we aim to address the time-consuming and labor-intensive nature of dataset annotation, as well as the issue of noisy pixels in pseudo-labels.



Annotation is time-consuming, and weakly supervised labels often yield poor results. So how can we achieve good results with minimal annotation?

We can annotate only a small amount of data and use box prompts to generate high-quality pseudo-labels for the remaining images. Models can be trained using these annotations and the pseudo-label.

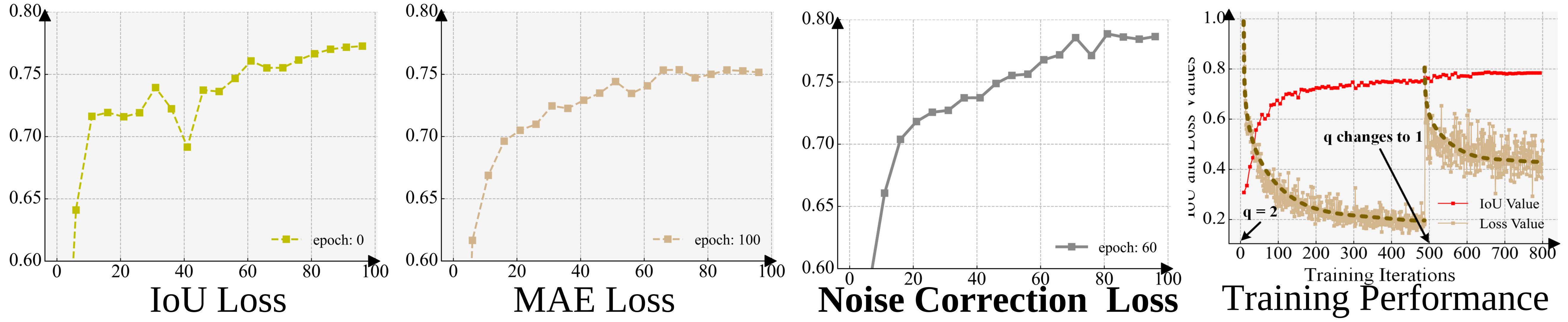


Even with box prompts, some predictions remain poor, and **such pseudo-labels can negatively impact the model's gradient learning, leading to incorrect optimization directions.**

METHOD & ALGORITHM

Let $\{x_t, gt\}$ be a pair of image and its noisy label. For any loss functions, the risk gradient of any model $m(\cdot)$ can be divided as

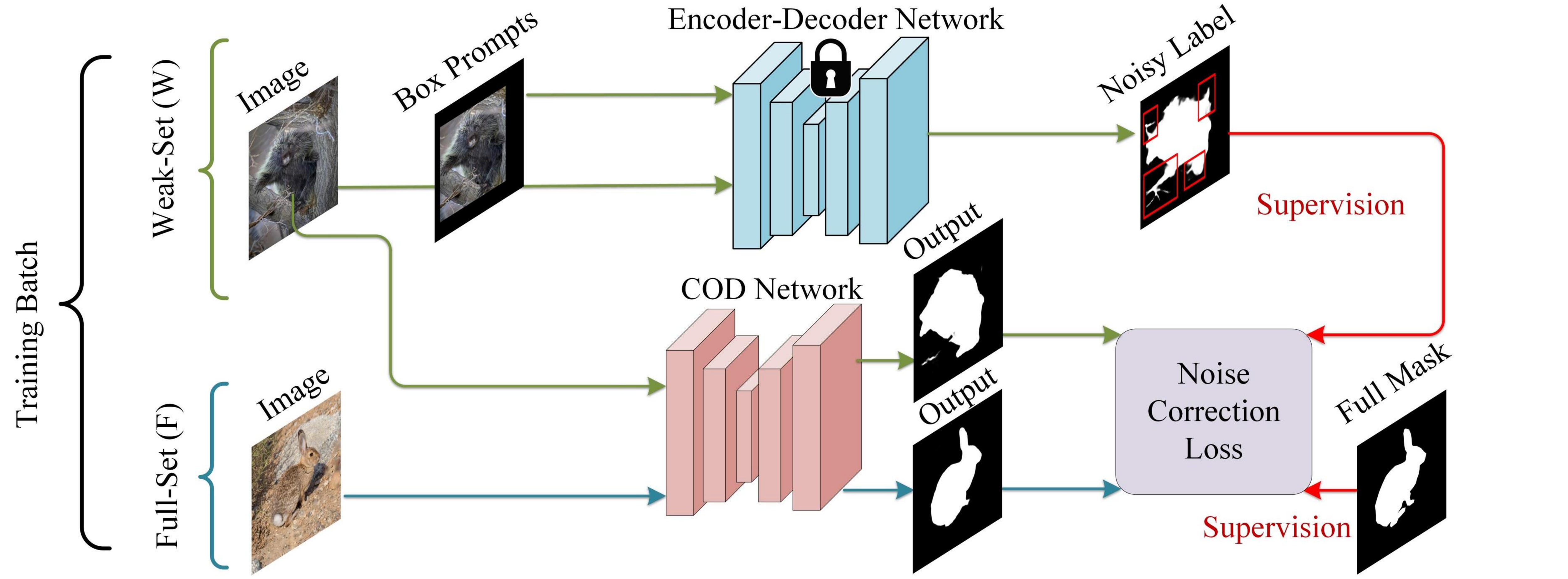
$$\nabla\mathcal{L}(m(x_t;\theta), g_t) = \underbrace{\nabla\mathcal{L}(m(x_t), g_t)}_{\text{noisy pixels}} + \underbrace{\nabla\mathcal{L}(m(x_t), \tilde{g}_t)}_{\text{clean pixels}}$$



Noise Correction Loss:

$$\mathcal{L}_{NC} = \frac{\sum_{i=1}^{H\times W} |p_i - g_i|^q}{\sum_{i=1}^{H\times W}(p_i + g_i) - \sum_{i=1}^{H\times W} p_i \cdot g_i}$$

$$\frac{\partial \mathcal{L}_{NC}}{\partial p_i} = \frac{sign(p_i - g_i)}{\sum_{i=1}^{H\times W}(p_i + g_i) - \sum_{i=1}^{H\times W} p_i \cdot g_i}$$



Effect of Noise Correction Loss

Losses	$\mathcal{M}\downarrow$	$E_\phi\uparrow$	$F_\beta\uparrow$	$S_\alpha\uparrow$
CE + IoU	0.027	0.910	0.780	0.844
CE + $\mathcal{L}_{NC}^{q=2.0}$	0.027	0.912	0.778	0.849
$\mathcal{L}_{NC}^{q=1.0}$	0.024	0.922	0.780	0.855
GCE [42]	0.029	0.900	0.759	0.835
$\mathcal{L}_{NC}$	**0.023**	**0.932**	**0.792**	**0.860**

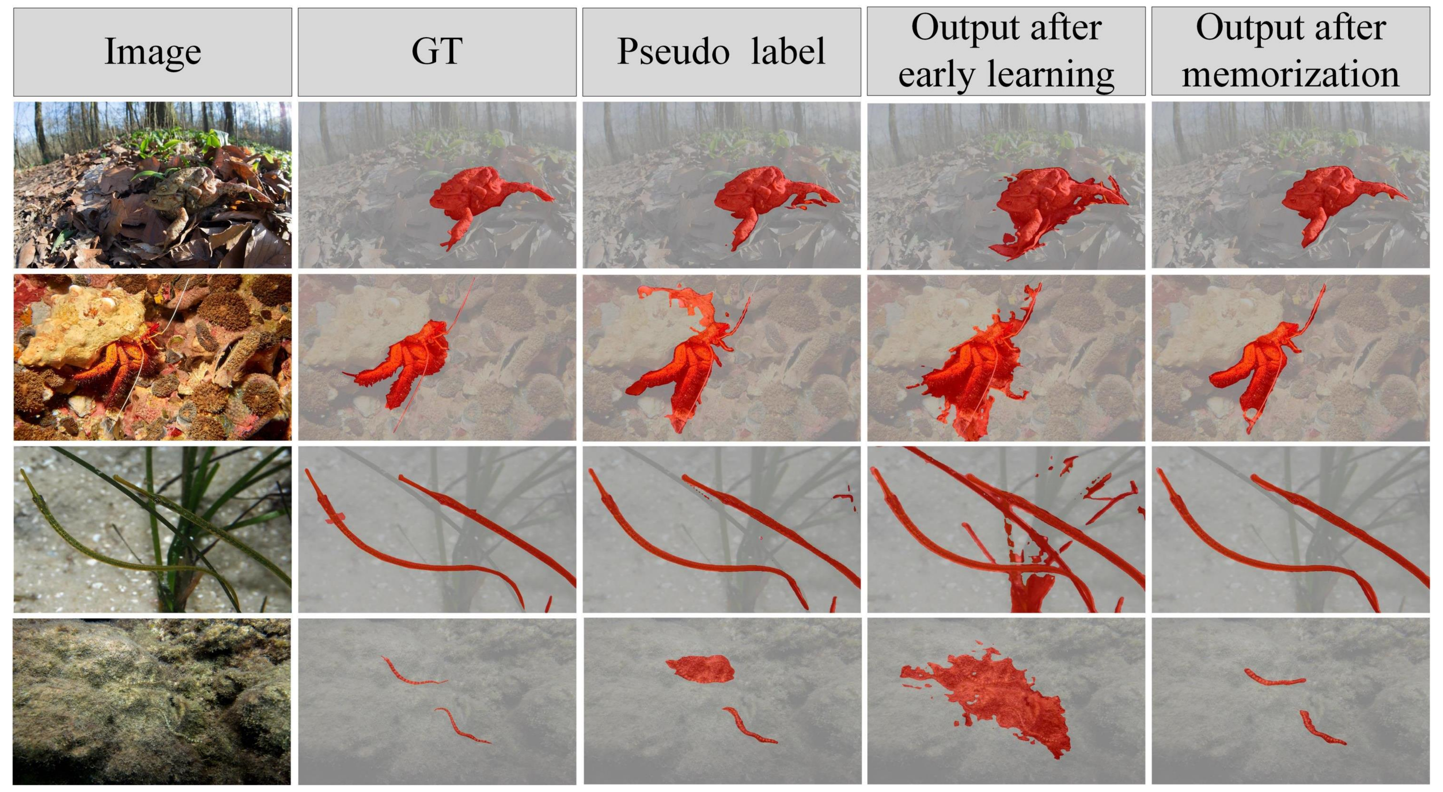
Universality of Noise Correction Loss

Model	Losses	$\mathcal{M}\downarrow$	$E_\phi\uparrow$	$F_\beta\uparrow$	$S_\alpha\uparrow$
SINetv2	w/o $\mathcal{L}_{NC}$	0.037	0.887	0.719	0.814
	w/ $\mathcal{L}_{NC}$	**0.033**	**0.896**	**0.734**	**0.814**
SCOD	w/o $\mathcal{L}_{NC}$	0.060	0.802	0.628	0.711
	w/ $\mathcal{L}_{NC}$	**0.045**	**0.853**	**0.665**	**0.759**

EXPERIMENT

Model	Encoder	Annotation	CAMO (250) $\mathcal{M}\downarrow$	$E_\phi\uparrow$	$F_\beta\uparrow$	$S_\alpha\uparrow$	CHAMELEON (76) $\mathcal{M}\downarrow$	$E_\phi\uparrow$	$F_\beta\uparrow$	$S_\alpha\uparrow$	COD10K (2026) $\mathcal{M}\downarrow$	$E_\phi\uparrow$	$F_\beta\uparrow$	$S_\alpha\uparrow$
Weakly Supervised Methods														
WSSA^{20} [37]	ResNet-50	$\mathcal{S}$ 100%	0.118	0.686	0.632	0.696	0.067	0.860	0.719	0.782	0.071	0.770	0.598	0.684
SCWS^{21} [35]	ResNet-50	$\mathcal{S}$ 100%	0.102	0.658	0.651	0.713	0.053	0.881	0.721	0.792	0.055	0.805	0.644	0.710
TEL^{22} [22]	ResNet-50	$\mathcal{B}$ 100%	0.104	0.681	0.654	0.717	0.073	0.827	0.706	0.785	0.057	0.801	0.659	0.724
SCOD^{23} [14]	ResNet-50	$\mathcal{P}$ 100%	0.129	0.688	0.592	0.663	0.092	0.746	0.692	0.725	0.060	0.802	0.628	0.711
Fully Supervised Methods														
SINet^{20} [9]	ResNet-50	$\mathcal{F}$ 100%	0.100	0.771	0.675	0.751	0.043	0.891	0.787	0.869	0.051	0.806	0.634	0.771
FEDER^{23} [13]	ResNet-50	$\mathcal{F}$ 100%	0.071	0.898	0.781	0.802	0.030	0.959	0.851	0.887	0.032	0.905	0.751	0.822
SINetv2^{21} [7]	Res2Net-50	$\mathcal{F}$ 100%	0.070	0.895	0.782	0.820	0.030	0.961	0.835	0.888	0.037	0.906	0.718	0.815
BSA-Net^{22} [44]	Res2Net-50	$\mathcal{F}$ 100%	0.079	0.851	0.763	0.794	0.026	0.946	0.856	0.896	0.034	0.891	0.738	0.818
BGNet^{22} [32]	Res2Net-50	$\mathcal{F}$ 100%	0.073	0.870	0.789	0.812	0.027	0.943	0.857	0.901	0.033	0.901	0.753	0.831
CamoFormer^{22} [34]	PVTv2-B4	$\mathcal{F}$ 100%	0.046	0.929	0.854	0.872	0.022	0.957	0.880	0.909	0.023	0.932	0.811	0.869
FSPNet^{23} [16]	ViT-B16	$\mathcal{F}$ 100%	0.050	0.899	0.830	0.856	0.022	0.942	0.865	0.908	0.026	0.895	0.769	0.851
HitNet^{23} [15]	PVTv2-B2	$\mathcal{F}$ 100%	0.055	0.906	0.831	0.849	0.019	0.966	0.898	0.921	0.023	0.935	0.823	0.871
MSCAF-Net^{23} [23]	PVTv2-B2	$\mathcal{F}$ 100%	0.046	0.929	0.852	0.873	0.022	0.958	0.874	0.911	0.024	0.927	0.798	0.865
Prompt Based Methods														
SAM^{23} [19]	ViT-H	-	0.209	0.304	0.039	0.394	0.157	0.276	0.017	0.418	0.111	0.315	0.018	0.445
SAM-P^{23} [19]	ViT-H	-	0.126	0.653	0.595	0.658	0.068	0.737	0.666	0.731	0.084	0.725	0.613	0.706
SAM-B^{23} [19]	ViT-H	-	0.139	0.495	0.346	0.535	0.121	0.467	0.276	0.524	0.073	0.433	0.218	0.534
Weakly Semi-Supervised Methods														
• **PNet**$_{F1}$	PVTv2-B4	$\mathcal{F}$ 1% + $\mathcal{B}$ 99%	0.051	0.922	0.835	0.852	0.038	0.921	0.812	0.847	0.031	0.903	0.745	0.828
• **PNet**$_{F5}$	PVTv2-B4	$\mathcal{F}$ 5% + $\mathcal{B}$ 95%	0.050	0.924	0.845	0.857	0.032	0.943	0.821	0.865	0.027	0.921	0.771	0.845
• **PNet**$_{F10}$	PVTv2-B4	$\mathcal{F}$ 10% + $\mathcal{B}$ 90%	0.048	0.925	0.841	0.861	0.028	0.949	0.830	0.878	0.024	0.927	0.782	0.855
• **PNet**$_{F20}$	PVTv2-B4	$\mathcal{F}$ 20% + $\mathcal{B}$ 80%	**0.043**	**0.934**	**0.856**	**0.872**	**0.024**	**0.954**	**0.861**	**0.892**	**0.023**	**0.932**	**0.792**	**0.860**
• **PNet**$^{\dagger}_{F20}$	PVTv2-B4	$\mathcal{F}$ 20% + $\mathcal{B}$ 240%	***0.039***	***0.942***	***0.870***	***0.882***	***0.021***	***0.964***	***0.886***	***0.908***	***0.016***	***0.960***	***0.857***	***0.901***

Compare with SOTA COD Methods



Effect of Noise Correction Loss

ACKNOWLEDGEMENT

We express our gratitude to all co-authors for their invaluable contributions to this paper. We also thank the reviewers and area chairs for their thoughtful reviews and recognition of our work.