NEURAL INFORMATION PROCESSING SYSTEMS

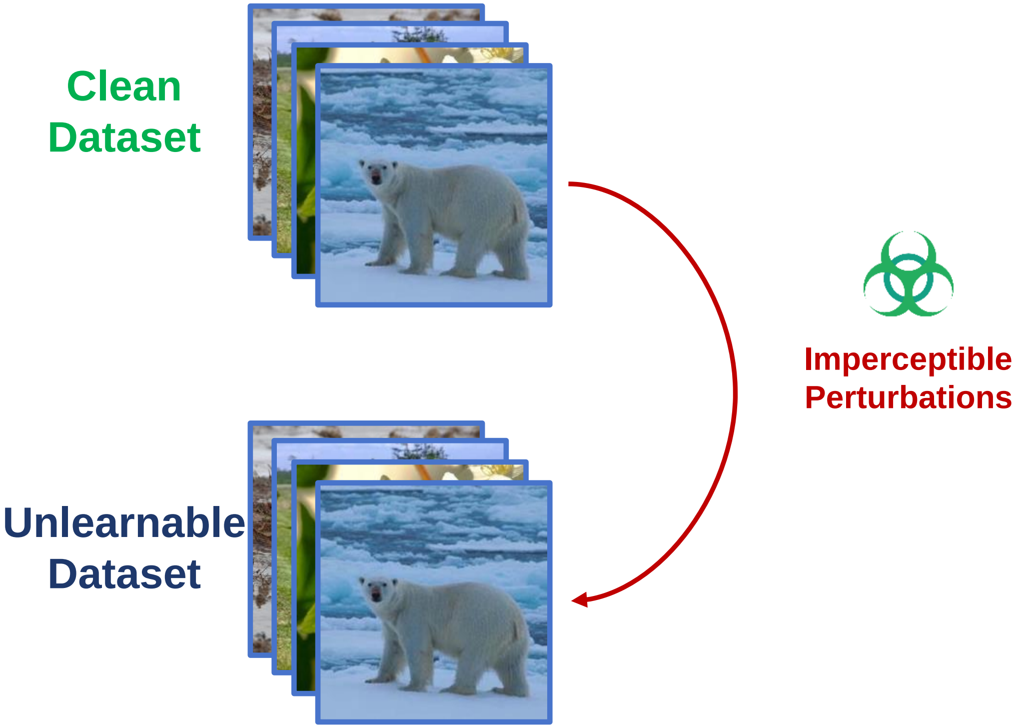
UnSeg: One Universal Unlearnable Example Generator is Enough against All Image Segmentation

Ye Sun · Hao Zhang · Tiehua Zhang · Xingjun Ma · Yu-Gang Jiang

Fudan University · HKUST · Tongji University

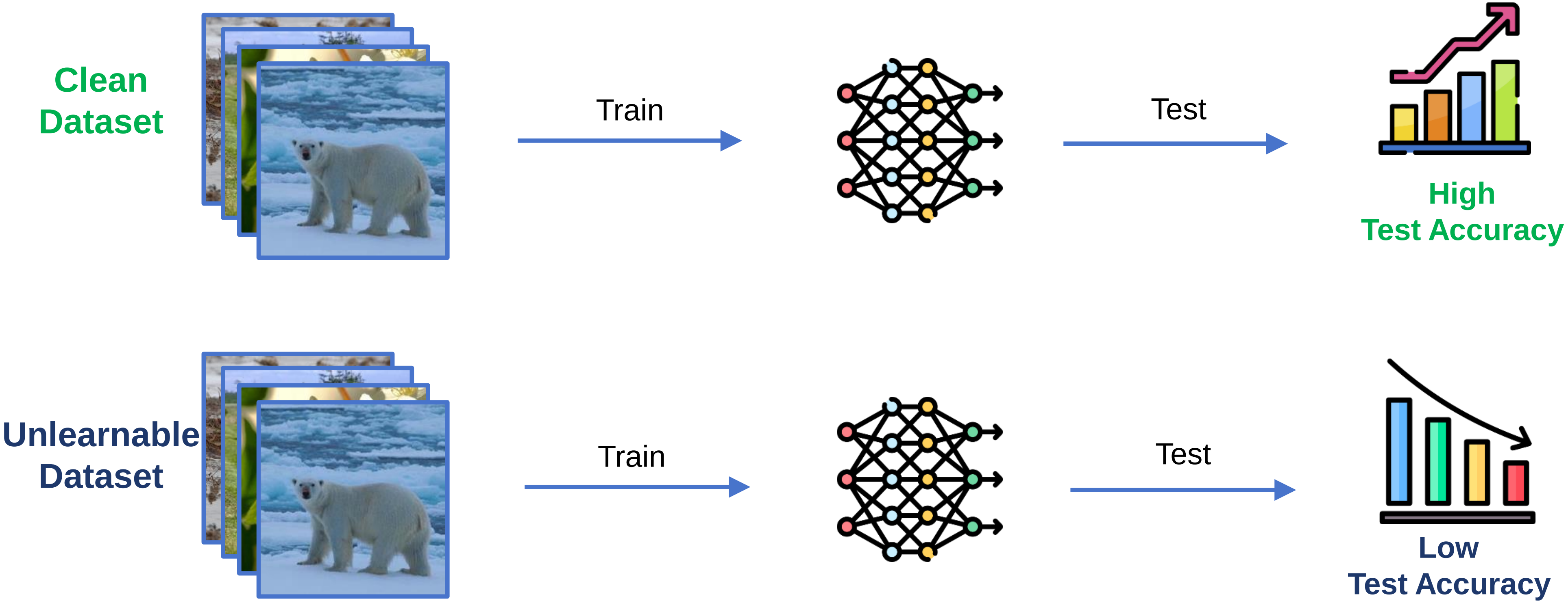
What are Unlearnable Examples?

Def. Unlearnable Examples (UEs, or “availability attacks”): manipulate the training data to prevent machine learning models from illegally learning useful representations.



What are Unlearnable Examples?

Def. Unlearnable Examples (UEs, or "availability attacks"): manipulate the training data to prevent machine learning models from illegally learning useful representations.



Challenges of UEs in Image Segmentation

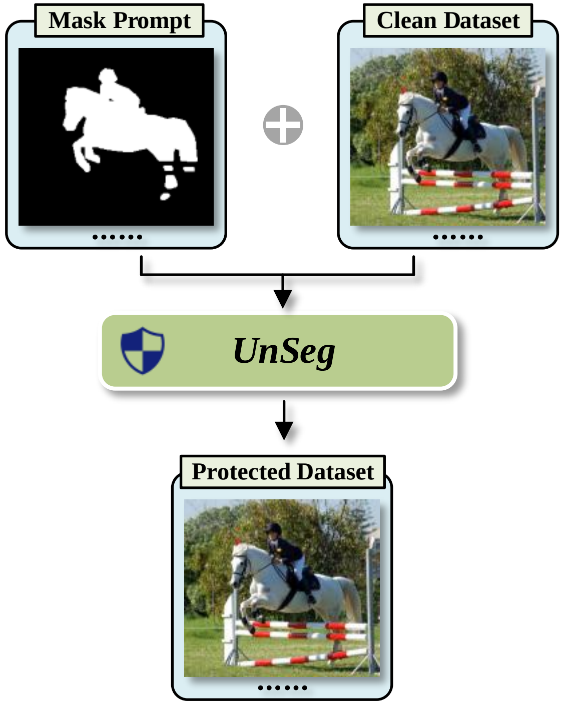
I. ☆ Data Efficiency Challenge: Effective UEs should be crafted based on a small number of images rather than existing large-scale image segmentation datasets.

II. ☆ Generation Efficiency Challenge : Effective method should be able to craft UEs directly without the need to optimize for each image .

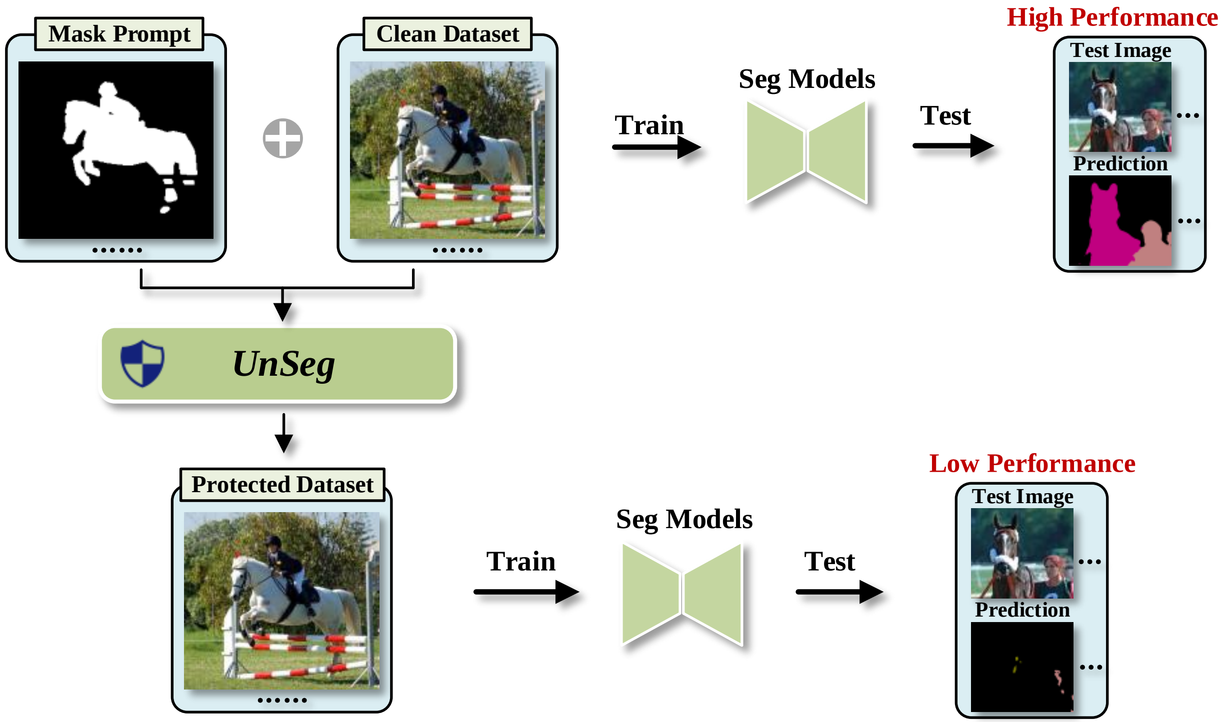
III. ☆ Transferability Challenge: The UE generation method should stay effective when transferred to protect different downstream tasks and datasets.

Method	***Data Efficiency***	***Generation Efficiency***	***Transferability***
UEs (Huang et al., ICLR 2021)	No	No	No
Robust UEs (Fu et al., ICLR 2022)	No	No	No
Stable UEs (Liu et al., AAAI 2024)	No	No	No
Transferable UEs (Ren et al., ICLR 2023)	No	No	Yes
Synthetic Perturbations (Yu et al, KDD 2022)	Yes	Yes	No
UnSeg **(Ours)**	**Yes**	**Yes**	**Yes**

Proposed Unlearnable Segmentation Pipeline



Proposed Unlearnable Segmentation Pipeline

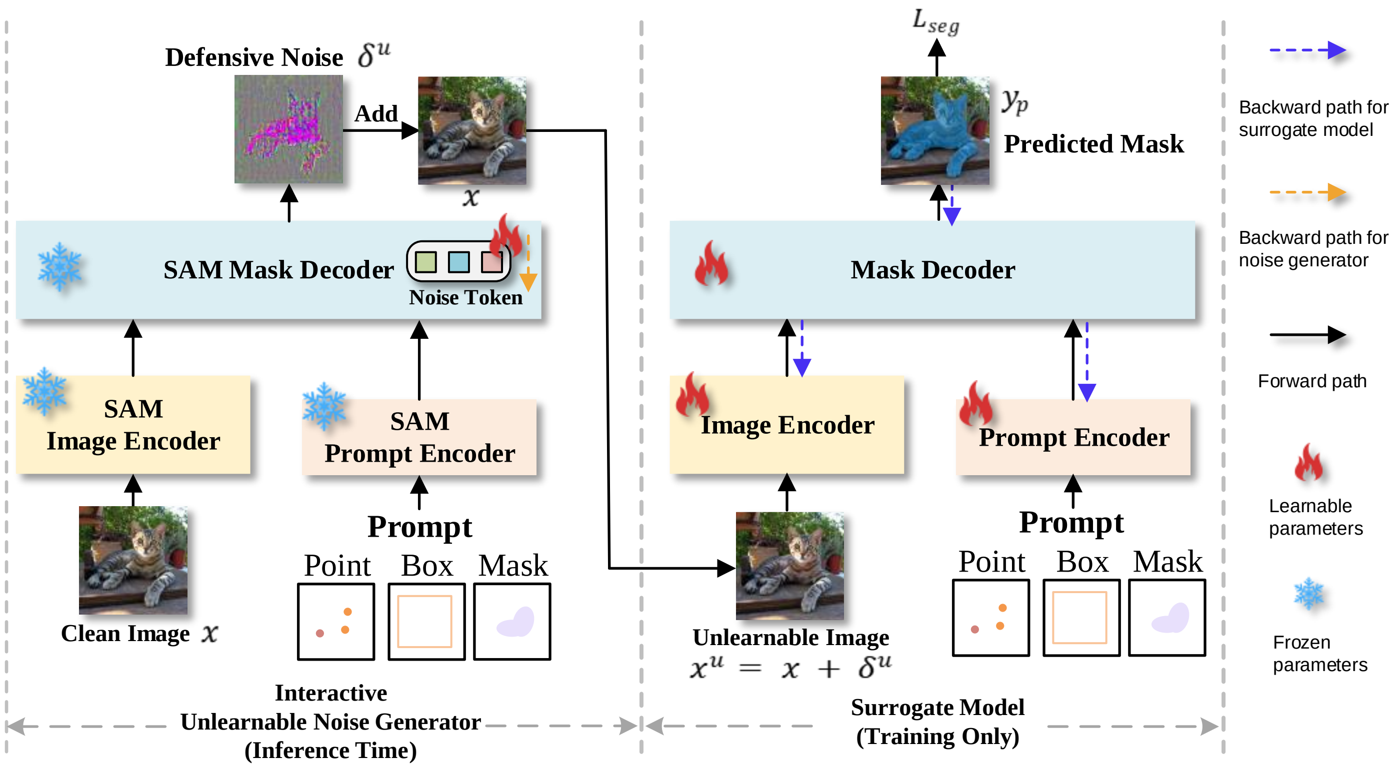


☆ Generative and interactive

☆ Instead of label information, requires only the mask prompt to protect the object.

☆ Can be finetuned on a small-scale dataset to achieve reasonable protection performance.

The UnSeg Framework



☆ Unlearnable Noise Generator:

$$\delta^u = \tanh(F \otimes T_{\text{noise}}^\top) \times \boxed{\epsilon} \quad \text{s.t.} \quad \|\delta^u\|_\infty \le \epsilon$$

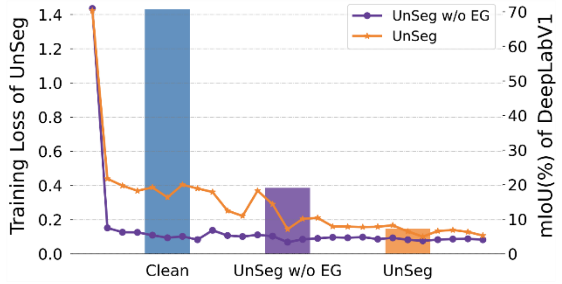
$$\boxed{\epsilon / v}$$

☆ Training the Unlearnable Noise Generator:

$$\arg\min_{\theta} \mathbb{E}_{(\boldsymbol{x},p,y)\sim\mathcal{D}_c} \left[\mathcal{L}_{seg}(\mathcal{F}(\boldsymbol{x},p;\theta),y)\right],$$

$$\arg\min_{\theta} \mathbb{E}_{(\boldsymbol{x},p,y)\sim\mathcal{D}_c} \left[\min_{\delta^u} \mathcal{L}_{seg}(\mathcal{F}'(\boldsymbol{x}+\delta^u,p;\theta),y)\right] \quad \text{s.t.} \quad \|\delta^u\|_\infty \le \epsilon,$$

☆ Training Stability



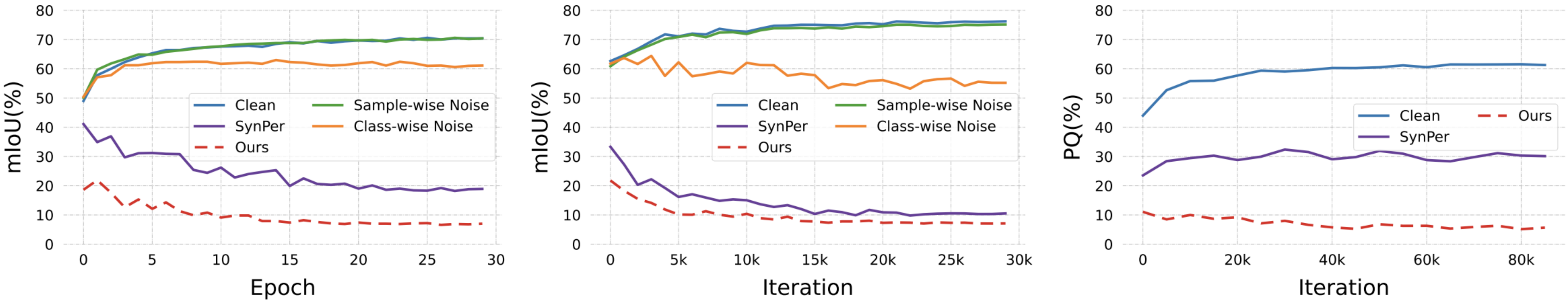
Evaluation Summary

Table 1: A summary of our considered evaluation tasks, datasets, models, and performance metrics.

Task	Model	Dataset	Metric
Semantic segmentation [41]	DeepLabV1 [8]/DeepLabV3 [10]/Mask2Former [11]	Pascal VOC2012 [14]/ADE20K [66]/Cityscapes [13]	mIoU [15]
Instance segmentation [21]	Mask2Former [11]	ADE20K [66]/COCO [37]/Cityscaptes [13]	AP [37]
Panoptic segmentation [30]	Mask2Former [11]	ADE20K [66]/COCO [37]/Cityscaptes [13]	PQ [30]
Interactive segmentation [31]	SAM-HQ [29]	HQSeg-44K [29]/DIS [45]/COIFT [36]/HRSOD [59]/ThinObject [36]	mIoU [15]
Remote sensing instance segmentation [7]	Rsprompter [7]	WHU [28]/NWPU [12]/SSDD [64]	mAP [7]
Medical image segmentation [49]	UNet++ [67]	Lung segmentation [2]/Kvasir-seg [27]	IoU [67]
Object detection [3]	DINO [60]	COCO [37]	AP [37]

UnSeg is…

☆ More effective than random noise and synthetic noise.



(a) Pascal VOC (DeepLabV1) (b) Pascal VOC (DeepLabV3) (c) Cityscapes (Mask2Former)

UnSeg is…

☆ **More effective than random noise and synthetic noise.**

☆ **More effective on mainstream image segmentation tasks.**

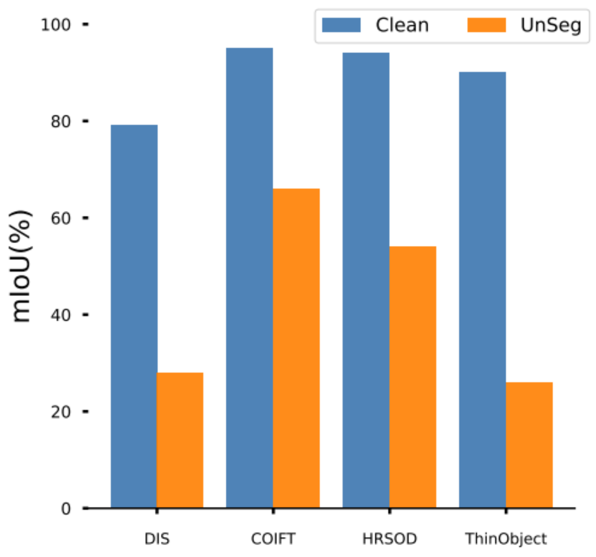
Dataset	**Method**	**Backbone**	**Panoptic** PQ	AP^{Th}_{pan}	$mIoU_{pan}$	**Instance** AP	AP^{S}	AP^{M}	AP^{L}	**Semantic** mIoU
ADE20k	Clean	R50	39.7	26.5	46.1	26.4	10.4	28.9	43.1	47.2
		Swin-T	41.6	27.7	49.3	27.9	10.8	29.8	46.2	47.7
	SynPer [58]	R50	18.6	13.6	28.7	9.3	7.1	13.4	9.7	25.4
	AR [51]	R50	37.8	24.9	43.1	25.4	9.4	27.7	43.3	43.9
	CUDA [50]	R50	**10.7**	8.4	19.6	12.0	**3.9**	14.6	22.5	19.6
	UnSeg(Ours)	R50	11.7(28.0↓)	**7.5**(19.0↓)	**17.7**(28.4↓)	**6.2**(20.2↓)	5.0(5.4↓)	**8.6**(20.3↓)	**7.3**(35.8↓)	**16.7**(30.5↓)
		Swin-T	**4.1**(37.5↓)	**3.4**(24.3↓)	**10.6**(38.7↓)	**4.1**(23.8↓)	**4.0**(6.8↓)	**5.8**(24.0↓)	**3.4**(42.8↓)	**7.8**(39.9↓)
COCO	Clean	R50	51.9	41.7	61.7	43.7	23.4	47.2	64.8	-
		Swin-T	53.2	43.3	63.2	45	24.5	48.3	67.4	-
	SynPer [58]	R50	11.3	9.5	11	10.8	13.4	15.2	5	-
	CUDA [50]	R50	6.7	4.7	11.2	9.7	**3.7**	10.9	18.8	-
	UnSeg(Ours)	R50	**4.2**(47.7↓)	**3.2**(38.5↓)	**5.2**(57.5↓)	**4.0**(39.7↓)	5.8(17.6↓)	**3.7**(43.5↓)	**1.7**(63.1↓)	-
		Swin-T	**4.1**(49.1↓)	**2.8**(40.5↓)	**6.0**(57.2↓)	**2.7**(42.3↓)	**4.4**(20.1↓)	**1.9**(46.4↓)	**0.7**(66.7↓)	-
Cityscapes	Clean	R50	62.1	37.3	77.5	37.4	-	-	-	79.4
		Swin-T	63.9	39.1	80.5	39.7	-	-	-	82.1
	SynPer [58]	R50	30.1	23.0	37.1	20.5	-	-	-	25.5
	AR [51]	R50	51.6	36.0	68.3	35.5	-	-	-	68.9
	CUDA [50]	R50	51.6	31.4	69.1	29.9	-	-	-	65.8
	UnSeg(Ours)	R50	**5.7**(56.4↓)	**1.1**(36.2↓)	**7.8**(69.7↓)	**2.3**(35.1↓)	-	-	-	**10.9**(68.5↓)
		Swin-T	**7.2**(56.7↓)	**1.7**(37.4↓)	**12.6**(67.9↓)	**1.5**(38.2↓)	-	-	-	**17.8**(61.6↓)

UnSeg is…

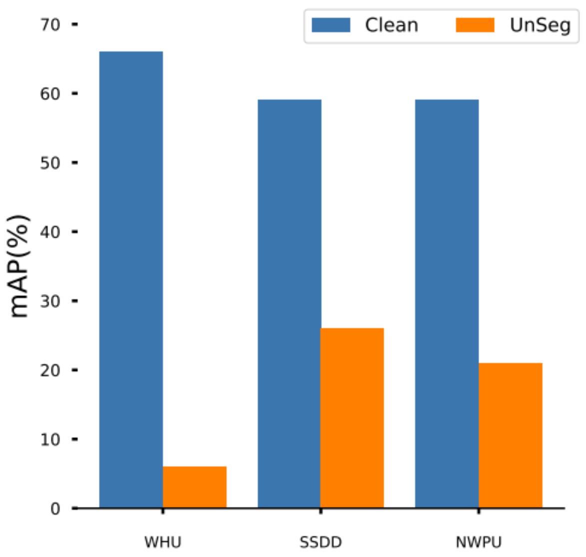
☆ **More effective than random noise and synthetic noise.**

☆ **More effective on mainstream image segmentation tasks.**

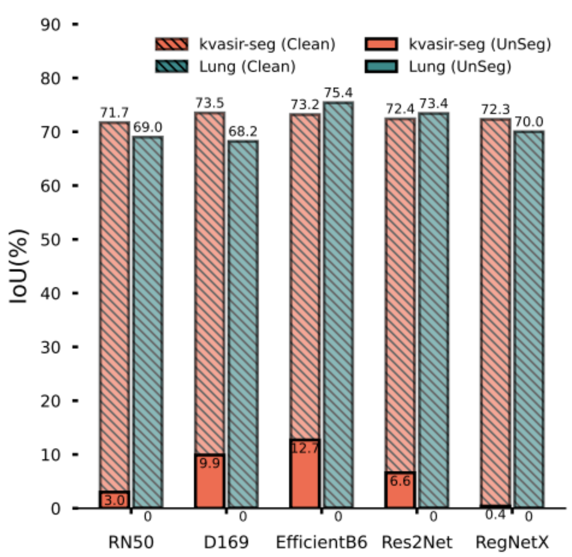
☆ **Effective on downstream related vision tasks.**



(a) Interactive Segmentation



(b) Remote Sensing Segmentation



(c) Medical Image Segmentation

UnSeg is…

☆ More effective than random noise and synthetic noise.

☆ More effective on mainstream image segmentation tasks.

☆ Effective on downstream related vision tasks.

☆ Resistant to Potential Defenses.

Clean	No Defense	Gaussian	JPEG [40]	AT [43]	DDC-AT [56]
75.1	5.8	7.3	44.8	23.1	28.5

UnSeg is…

☆ **More effective than random noise and synthetic noise.**

☆ **More effective on mainstream image segmentation tasks.**

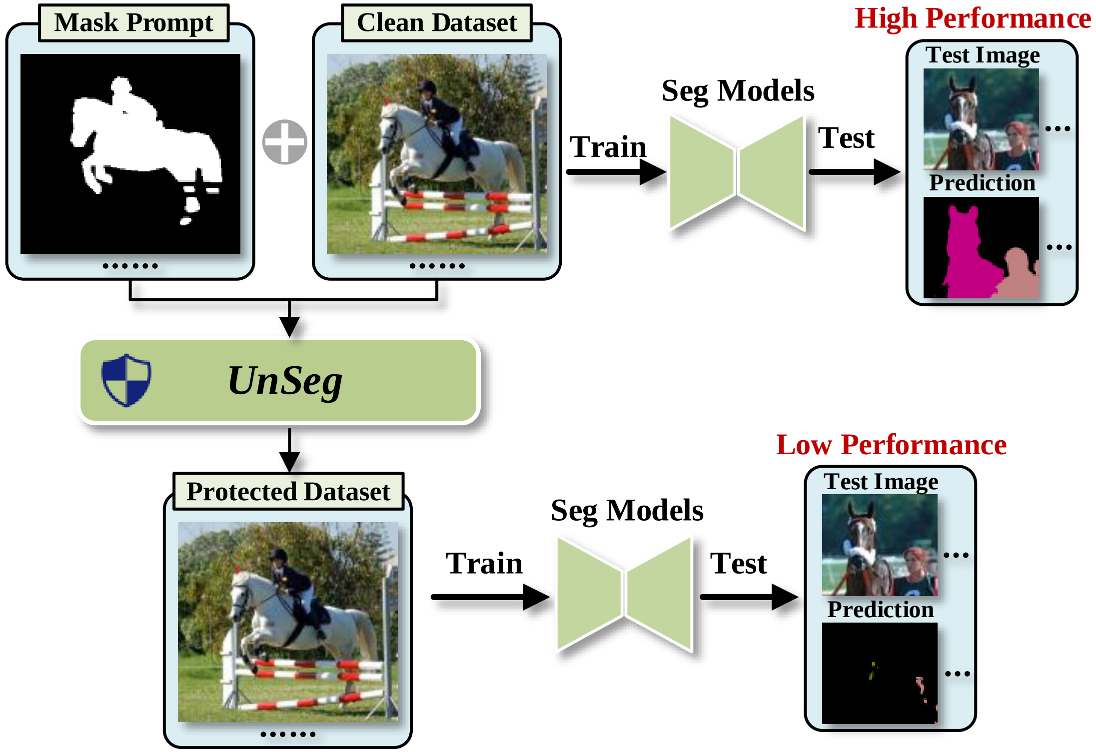
☆ **Effective on downstream related vision tasks.**

☆ **Resistant to Potential Defenses.**

☆ **Effective when mixed with clean data.**

Method	**Backbone**	**Clean Proportion**					
		0%	**20%**	**40%**	**60%**	**80%**	**100%**
Clean Only	ResNet50	-	67.3	70.1	71.0	71.6	72.3
	DenseNet169	-	69.3	70.7	72.1	72.2	73.6
	EfficientNetB6	-	69.7	71.2	73.5	72.7	74.0
	Res2Net	-	67.6	70.8	71.2	71.7	73.6
	RegNetX	-	68.1	69.2	71.1	71.3	72.6
Mixed Data	ResNet50	2.5	67.2	68.7	70.3	71.8	-
	DenseNet169	6.0	69.0	69.5	71.2	72.4	-
	EfficientNetB6	7.4	70.6	71.9	73.3	73.2	-
	Res2Net	6.7	68.7	70.5	71.7	72.6	-
	RegNetX	2.1	69.8	69.7	71.1	71.4	-

Thank you!



See the paper for more details!

Fudan University · HKUST · Tongji University