西安電子科技大學
XIDIAN UNIVERSITY

IRSAM: Advancing Segment Anything Model for Infrared Small Target Detection

Mingjin Zhang, Yuchun Wang, Jie Guo,
Yunsong Li, Xinbo Gao, Jing Zhang

Xidian University, China
Chongqing University of Posts and Telecommunications, China
The University of Sydney, Australia

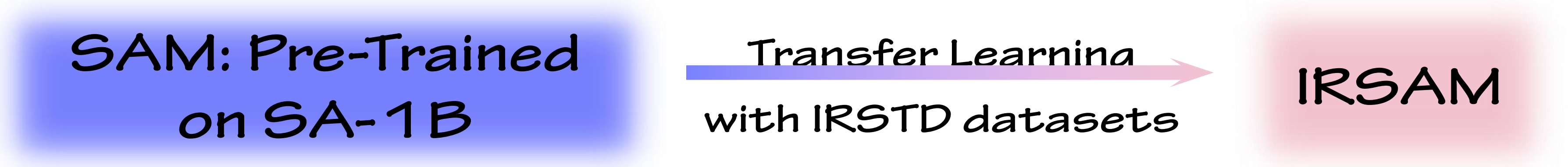
Background

- Task: The essence of infrared small target detection lies in pinpointing minuscule targets with a low signal-to-noise ratio amidst the intricacies of complex infrared imagery.

- Challenge: In comparison to natural image datasets, acquiring such data is inherently more challenging. While previous approaches have demonstrated promising results in certain straightforward scenarios, their efficacy is heavily contingent upon the particular design of the architecture and the magnitude of the training data. This dependency limits the applicability of these methods to more complex scenarios.

Motivation

With extensive research on deep models in the natural image domain and the proven effectiveness of transfer learning in mitigating generalization issues with limited training data for downstream tasks, a crucial question arises: Can a well-designed model pre-trained on a large-scale natural image dataset effectively kickstart the IRSTD task?



Method

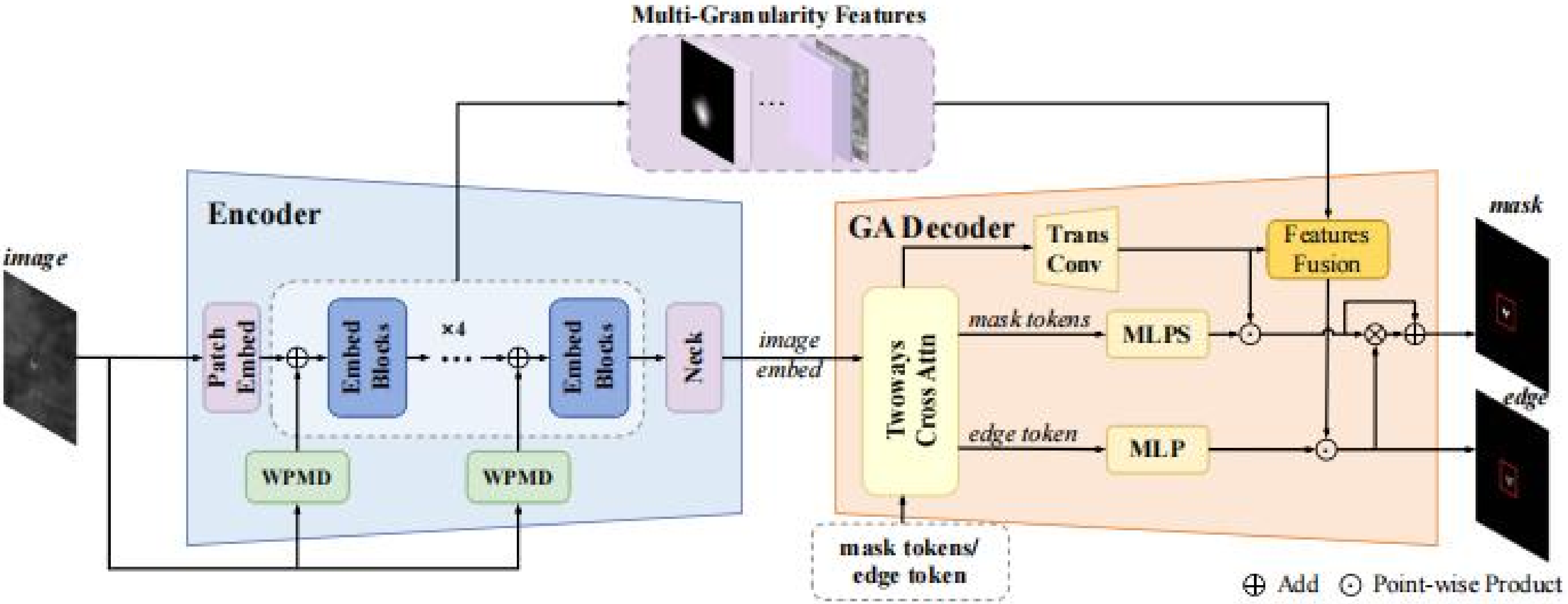


Fig. 2: Overall Architecture of IRSAM. Utilizing an encoder-decoder structure rooted in SAM, IRSAM incorporates two novel modules: WPMD and GAD, crafted specifically for the IRSTD task.

Method

Although similar to edge extraction algorithms like the Sobel operator, which can attenuate edges, Perona-Malik diffusion operates on a different principle. The anisotropic Perona-Malik function promotes diffusion (smoothing) within smoother regions while suppressing it at the edges. As a result, the output of WPMD would be a smoother version of the input, preserving essential structural information while eliminating noise.

Given a picture u, its corresponding PMD equation is given by $g\left(|\nabla u|\right) = 1/\left(1+|\nabla u|^2/k^2\right)$,where diffusion coefficient $\frac{\partial u}{\partial t} = div\left(g\left(|\nabla u|\right)\nabla u\right)$

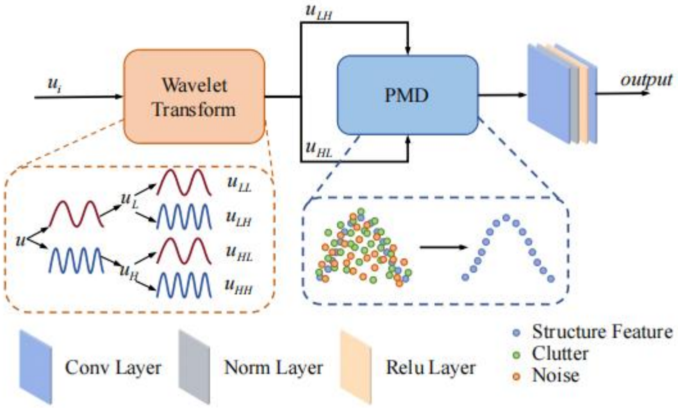


Fig. 3: Structure of the WPMD.

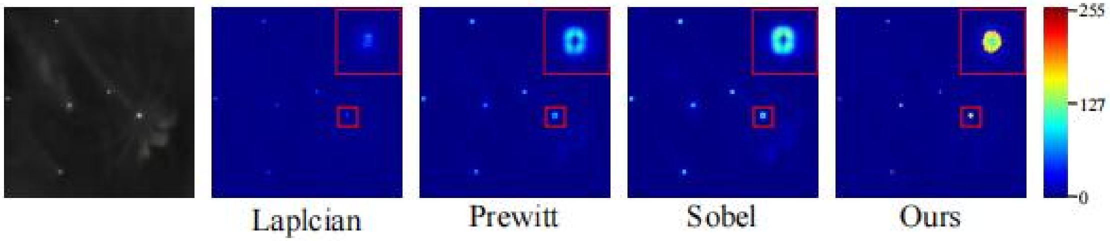


Fig. 8: Comparison of the edge maps utilizing various methods.

Method

Small infrared targets usually have limited visual features and are easily confounded with the background or similar targets. To improve the performance of infrared small target segmentation, it is necessary to consider both the global context information, which can help extract the overall semantics of the image and enhance the detection of small objects, as well as the local boundary information, which can help preserve the spatial details of small objects and improve the precision of segmentation boundaries. SAM adopts the ViT architecture, which excels at capturing long-term dependence and global information. The early layer of the ViT structure has been demonstrated to preserve more general image boundary details in previous works while the deep layer contains higherlevel semantics . To improve the performance of SAM in the IRSTD task, we devise the Granularity-Aware Decoder to fuse the multi-granularity features by feeding global semantic context and local fine-grained features to the decoder, GAD enjoys a richer multi-view knowledge, as shown in Architecture of Network.

Experiments

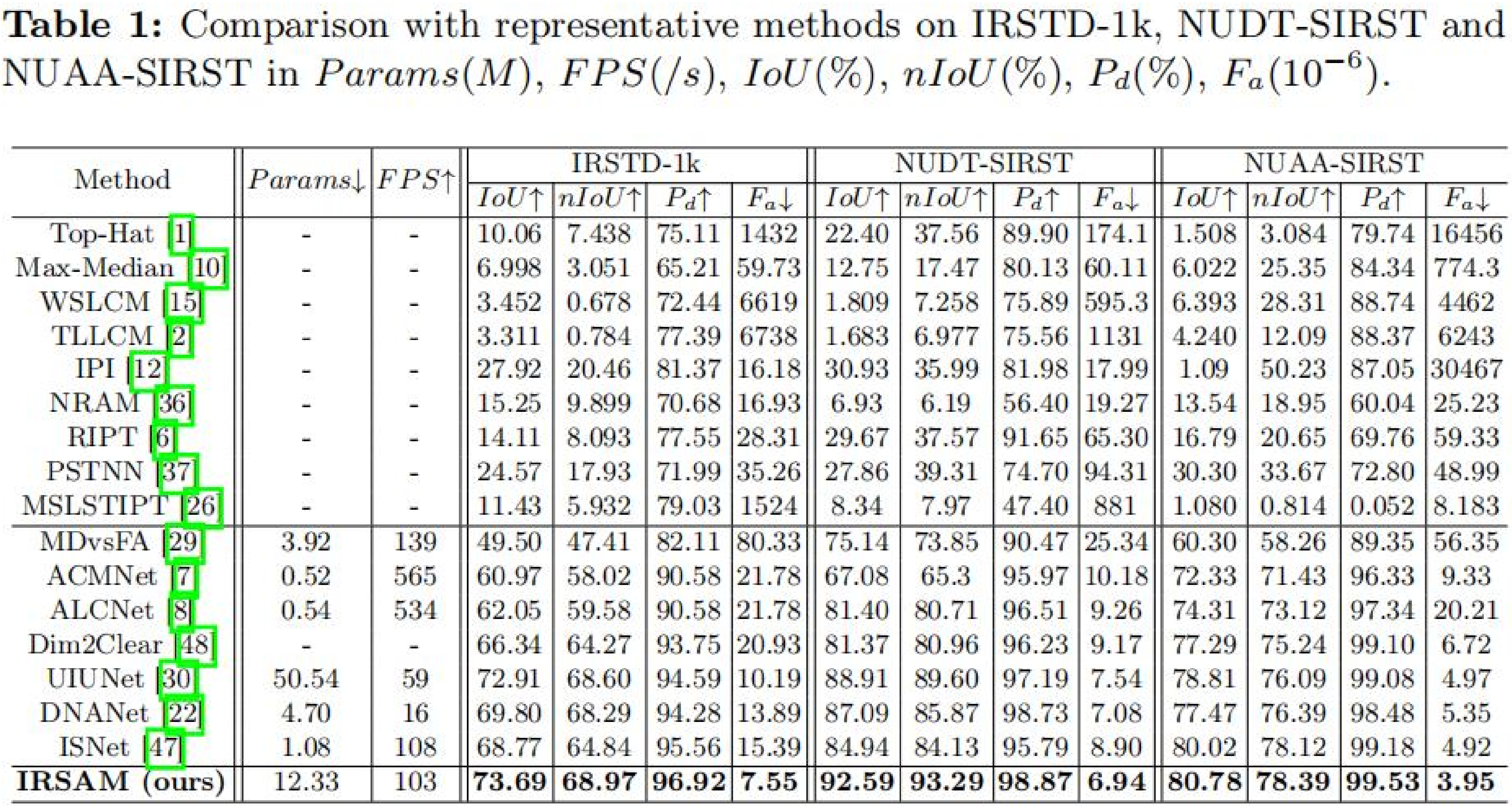


Table 1: Comparison with representative methods on IRSTD-1k, NUDT-SIRST and NUAA-SIRST in $Params(M)$, $FPS(/s)$, $IoU(\%)$, $nIoU(\%)$, $P_d(\%)$, $F_a(10^{-6})$.

Method	$Params\downarrow$	$FPS\uparrow$	IRSTD-1k				NUDT-SIRST				NUAA-SIRST			
			$IoU\uparrow$	$nIoU\uparrow$	$P_d\uparrow$	$F_a\downarrow$	$IoU\uparrow$	$nIoU\uparrow$	$P_d\uparrow$	$F_a\downarrow$	$IoU\uparrow$	$nIoU\uparrow$	$P_d\uparrow$	$F_a\downarrow$
Top-Hat [1]	-	-	10.06	7.438	75.11	1432	22.40	37.56	89.90	174.1	1.508	3.084	79.74	16456
Max-Median [10]	-	-	6.998	3.051	65.21	59.73	12.75	17.47	80.13	60.11	6.022	25.35	84.34	774.3
WSLCM [15]	-	-	3.452	0.678	72.44	6619	1.809	7.258	75.89	595.3	6.393	28.31	88.74	4462
TLLCM [2]	-	-	3.311	0.784	77.39	6738	1.683	6.977	75.56	1131	4.240	12.09	88.37	6243
IPI [12]	-	-	27.92	20.46	81.37	16.18	30.93	35.99	81.98	17.99	1.09	50.23	87.05	30467
NRAM [36]	-	-	15.25	9.899	70.68	16.93	6.93	6.19	56.40	19.27	13.54	18.95	60.04	25.23
RIPT [6]	-	-	14.11	8.093	77.55	28.31	29.67	37.57	91.65	65.30	16.79	20.65	69.76	59.33
PSTNN [37]	-	-	24.57	17.93	71.99	35.26	27.86	39.31	74.70	94.31	30.30	33.67	72.80	48.99
MSLSTIPT [26]	-	-	11.43	5.932	79.03	1524	8.34	7.97	47.40	881	1.080	0.814	0.052	8.183
MDvsFA [29]	3.92	139	49.50	47.41	82.11	80.33	75.14	73.85	90.47	25.34	60.30	58.26	89.35	56.35
ACMNet [7]	0.52	565	60.97	58.02	90.58	21.78	67.08	65.3	95.97	10.18	72.33	71.43	96.33	9.33
ALCNet [8]	0.54	534	62.05	59.58	90.58	21.78	81.40	80.71	96.51	9.26	74.31	73.12	97.34	20.21
Dim2Clear [48]	-	-	66.34	64.27	93.75	20.93	81.37	80.96	96.23	9.17	77.29	75.24	99.10	6.72
UIUNet [30]	50.54	59	72.91	68.60	94.59	10.19	88.91	89.60	97.19	7.54	78.81	76.09	99.08	4.97
DNANet [22]	4.70	16	69.80	68.29	94.28	13.89	87.09	85.87	98.73	7.08	77.47	76.39	98.48	5.35
ISNet [47]	1.08	108	68.77	64.84	95.56	15.39	84.94	84.13	95.79	8.90	80.02	78.12	99.18	4.92
IRSAM (ours)	12.33	103	**73.69**	**68.97**	**96.92**	**7.55**	**92.59**	**93.29**	**98.87**	**6.94**	**80.78**	**78.39**	**99.53**	**3.95**

Experiments

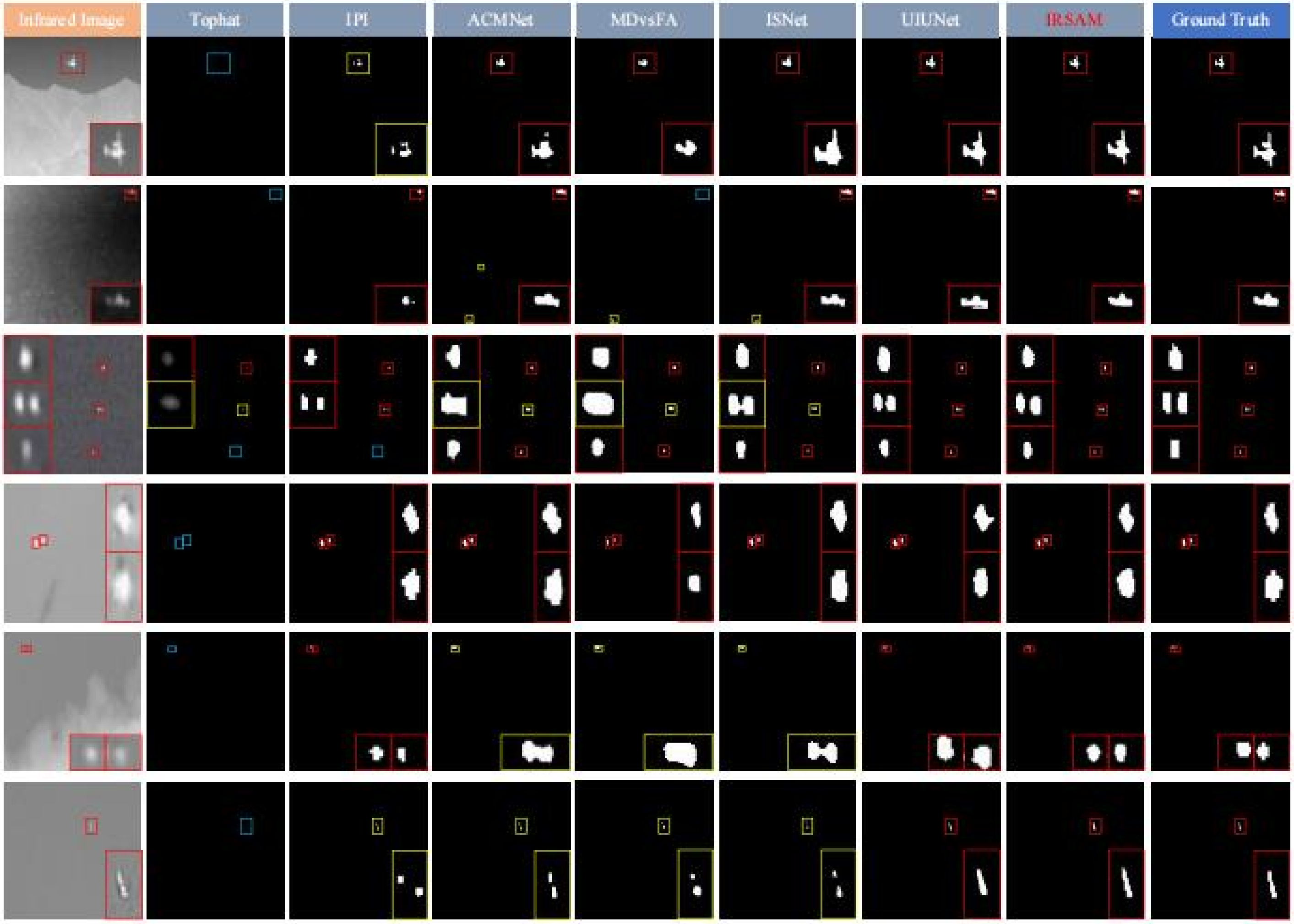


Fig. 4: Visualization results using different IRSTD methods. The closed views are shown at the border. In each prediction result, red, blue, and yellow boxes represent the correct detection, miss detection, and false detection, respectively.

Thank you