Text-IRSTD: Leveraging Semantic Text to Promote Infrared Small Target Detection in Complex Scenes

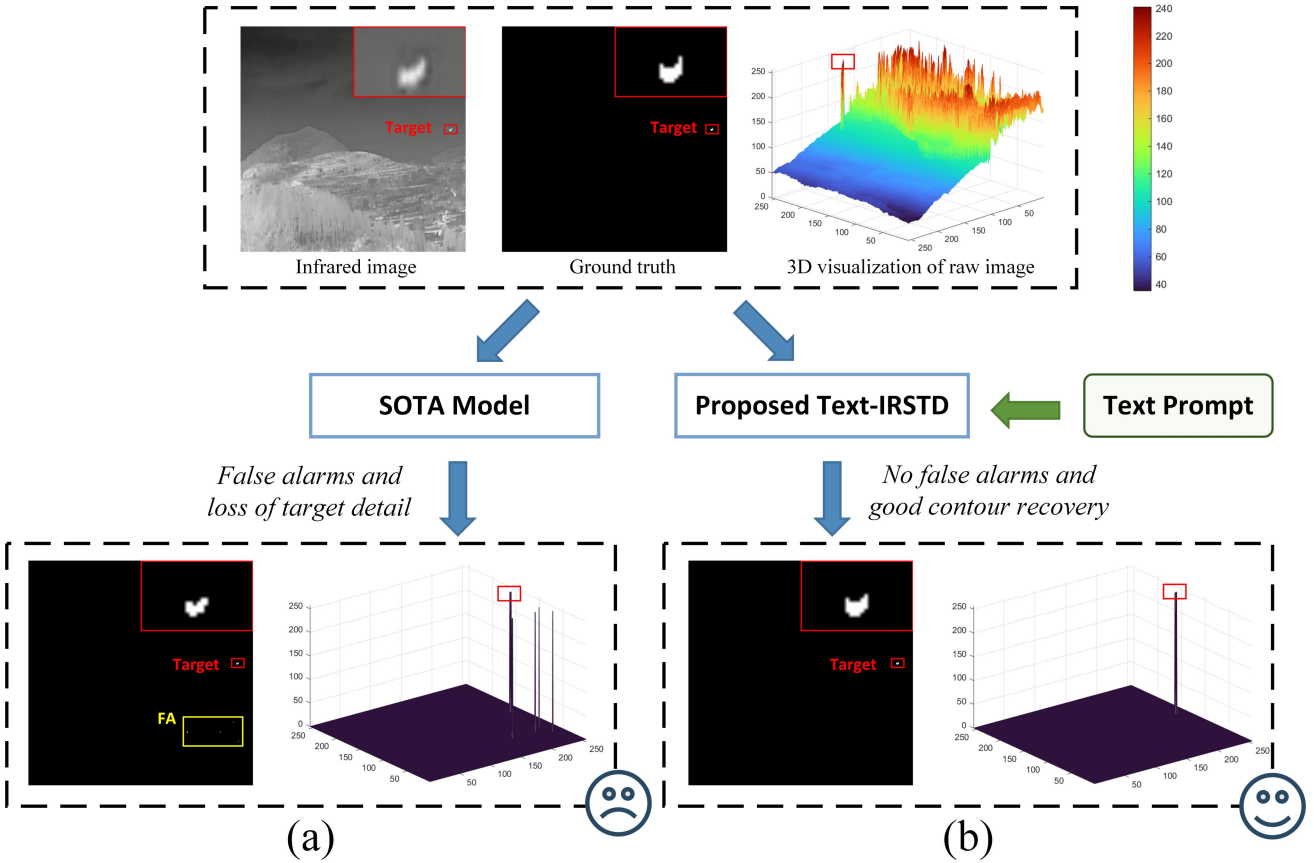
Feng Huang*, Shuyuan Zheng*, Zhaobing Qiu*,†, Huanxian Liu, Huanxin Bai, Liqiong Chen†
Fuzhou University, Fuzhou, China

Background

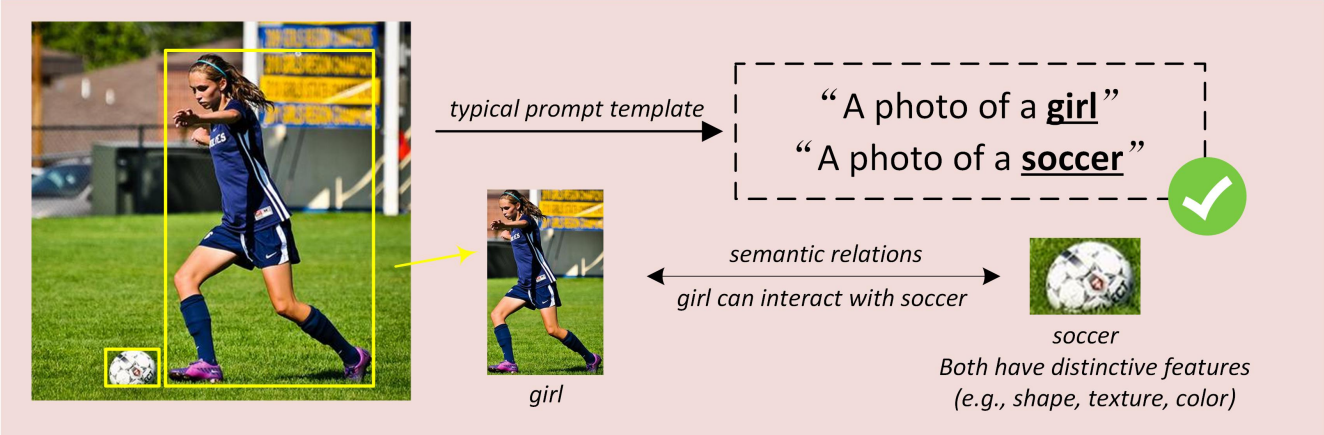
Challenge：

(1) Existing methods usually focus on mining visual features of targets, which struggles to cope with complex and diverse detection scenarios.

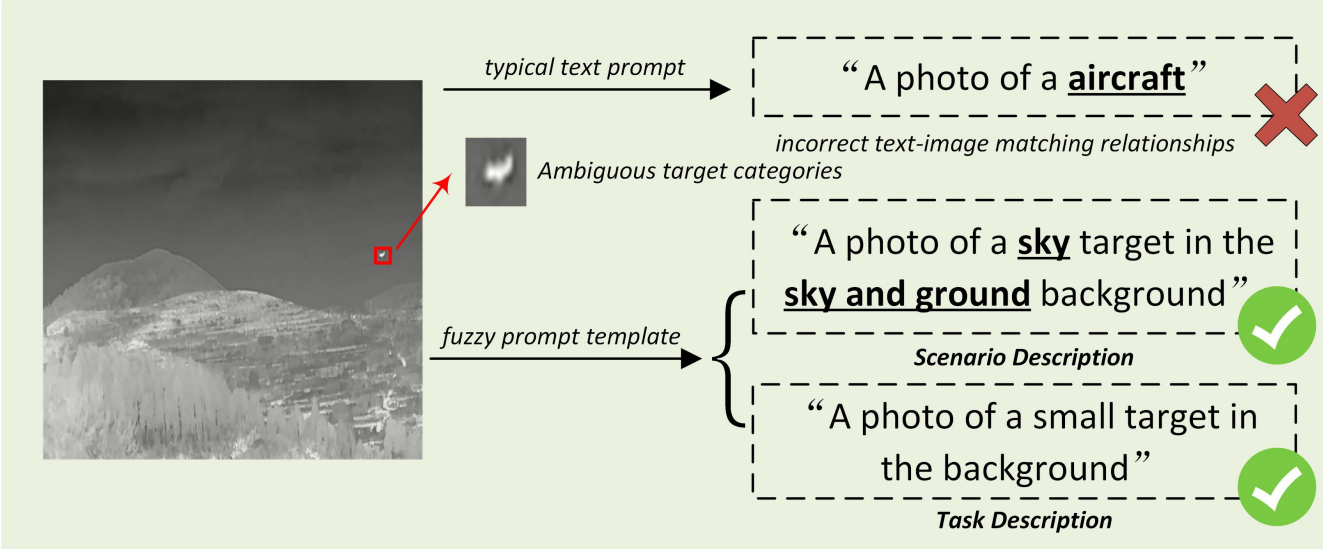
(2) Semantic text–image models excel in generic detection but perform poorly on infrared small targets, mainly due to ambiguous categories invalidate existing data annotation and feature fusion methods.



Fuzzy Semantic Text Prompt For IRSTD



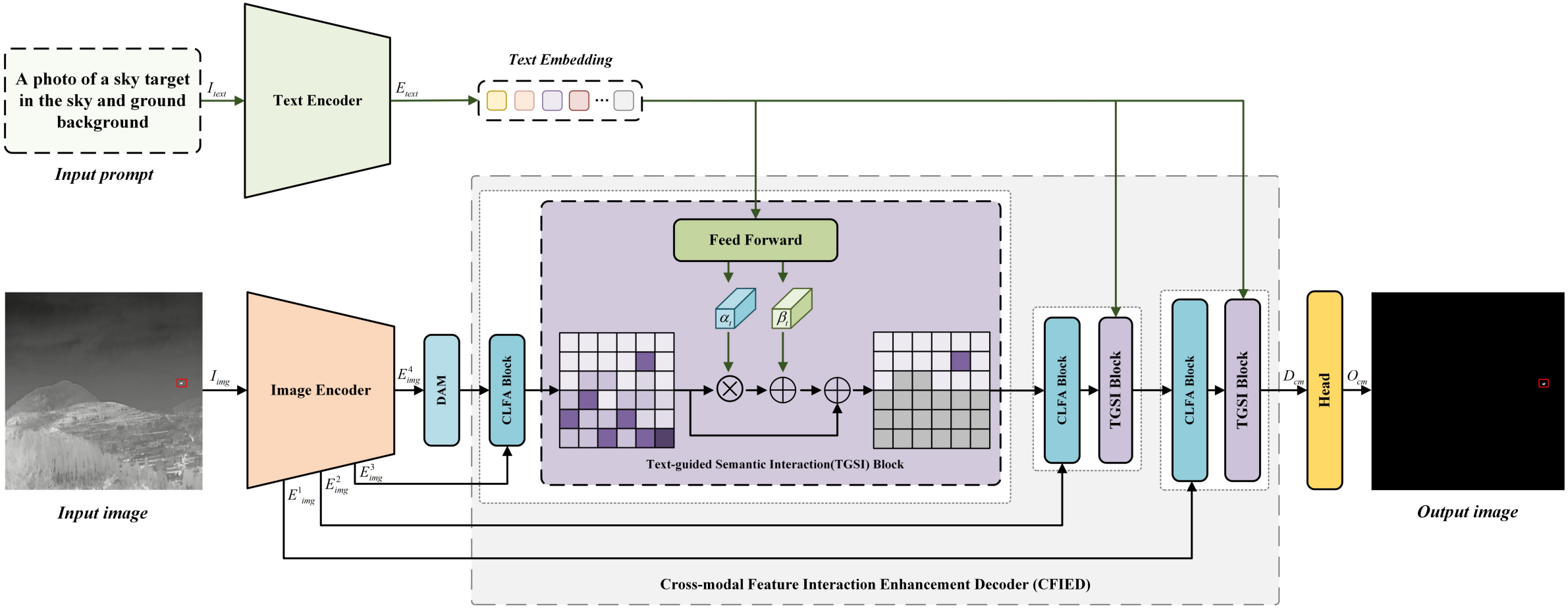
(a) Typical prompt template for object detection



(b) Proposed fuzzy semantic prompt for IRSTD

- In generic target detection, researchers usually directly describe specific categories of targets.
- IR small targets are usually presented as bright spots without specific categories, using existing text prompts causes incorrect text-image matching relationships.
- We propose a novel fuzzy semantic text prompt for IRSTD, which includes two templates: scenario description and task description, which not only establishes correct and robust matching relationships, but also provides more semantic associations.

Method


A photo of a sky target in the sky and ground background
Input prompt
I_{text}
Text Encoder
E_{text}
Text Embedding
Feed Forward
α_t
β_t
Text-guided Semantic Interaction(TGSI) Block
Input image
I_{img}
Image Encoder
E^4_{img}
E^3_{img}
E^2_{img}
E^1_{img}
DAM
CLFA Block
CLFA Block
TGSI Block
CLFA Block
TGSI Block
D_{cm}
Head
O_{cm}
Output image
Cross-modal Feature Interaction Enhancement Decoder (CFIED)

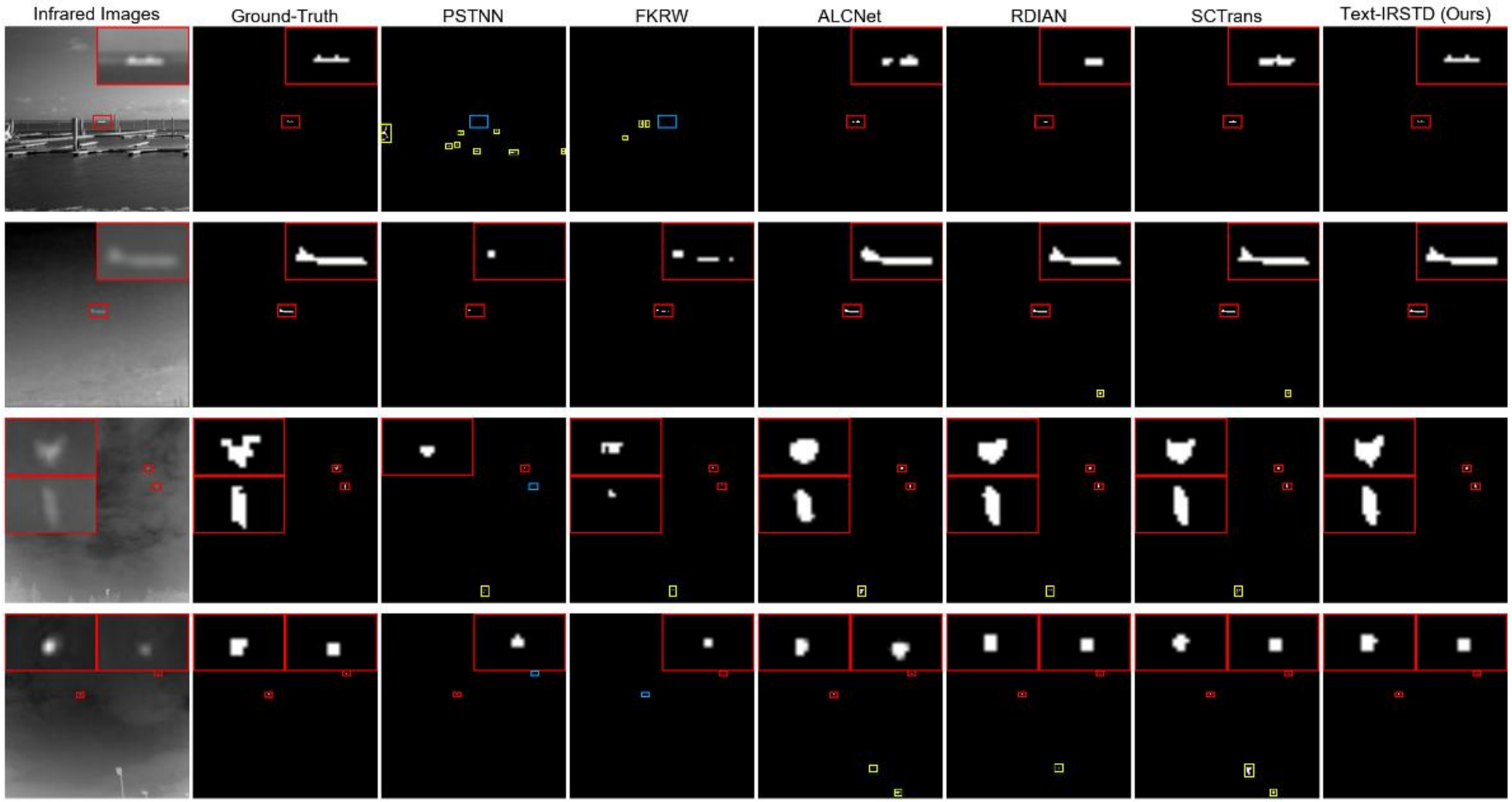
Experiments

Quantitative results of different methods. The best values are highlighted with **bold**, the second best values are underlined.

Method	NUDT-SIRST				IRSTD-1k			
	IoU	P_d	F_a	F_1	IoU	P_d	F_a	F_1
IPI [4]	38.15	90.78	369.9	61.52	30.95	78.91	180.4	55.15
PSTNN [31]	17.21	54.47	81.62	29.45	25.58	62.96	78.91	41.35
FKRW [15]	25.38	60.26	107.5	38.54	15.89	54.54	26.75	27.08
GSWLCM [16]	4.909	65.00	15.71	9.301	3.068	59.25	26.94	6.045
RUMFR [1]	26.08	67.89	96.42	39.25	10.26	53.87	25.29	18.59
ACMNet [2]	68.14	96.57	17.38	83.41	61.14	87.88	41.79	74.17
ALCNet [3]	72.56	96.05	7.572	86.78	58.27	89.90	44.69	72.92
AGPCNet [33]	85.97	97.63	7.228	93.51	62.51	91.58	20.97	77.68
DNANet [9]	94.70	98.68	3.212	97.31	64.51	90.23	18.27	78.77
UIUNet [21]	89.15	97.11	5.506	94.23	64.32	90.24	28.33	79.02
RDIAN [18]	86.08	98.15	6.711	92.36	61.86	88.56	41.44	76.27
DMFNet [6]	87.67	98.42	2.294	93.30	64.45	90.14	18.24	78.52
SCTrans [29]	94.83	98.42	2.065	97.19	62.95	92.24	23.40	78.04
Text-IRSTD (w/o text)	95.25	98.94	1.663	97.37	65.50	91.56	16.13	78.73
Text-IRSTD (Ours)	**95.84**	**99.73**	**1.032**	**97.95**	**69.57**	**92.59**	**14.97**	**79.24**

Experiments

Visual results of different IRSTD methods. The red, blue, and yellow boxes represent correctly detected targets, missed targets, and false detections, respectively



Thank you