SCTransNet: Spatial-Channel Cross Transformer Network for Infrared Small Target Detection

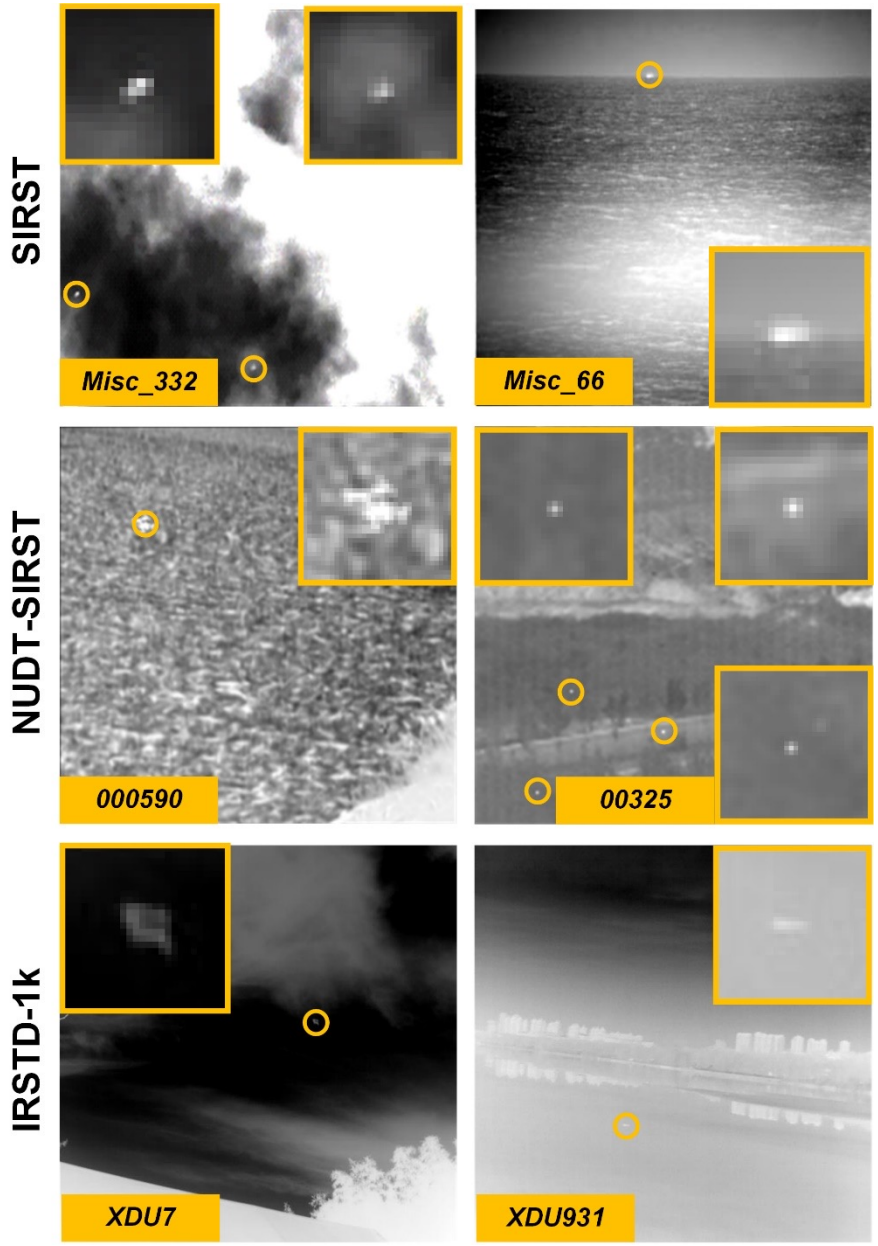
Shuai Yuan, Hanlin Qin*, Xiang Yan*, Naveed Akhtar, Ajmal Mian

Xidian University; The University of Melbourne; The University of Western Australia

IEEE Transactions on Geoscience and Remote Sensing, 2024

西安电子科技大学 墨尔本大学 袁帅

Infrared Small Target Detection (IRSTD)



1) Dim

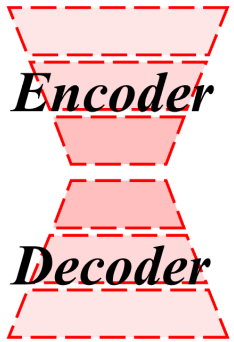
2) Small

3) Characterless

4) Uncertain shapes

Segmentation

Input:



1) Encoders

2) Decoders

3) Skip connection

4) Feature fusion

Label:

Challenges

① *Semantic interaction across full levels*

② *Information gap between encoders and decoders*

③ *Long-range contextual perception of targets and backgrounds in deeper layers*

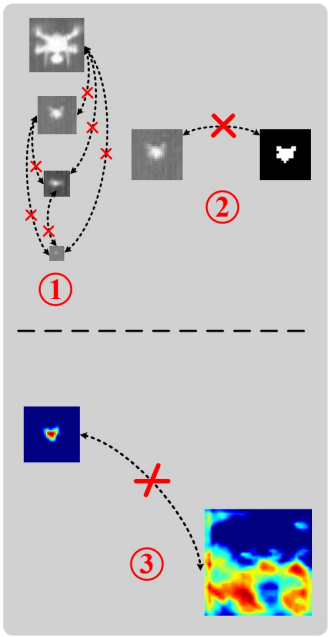
U-Net:

1) Encoders

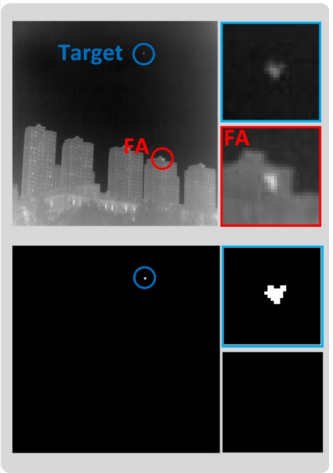
2) Decoders

3) Skip connection

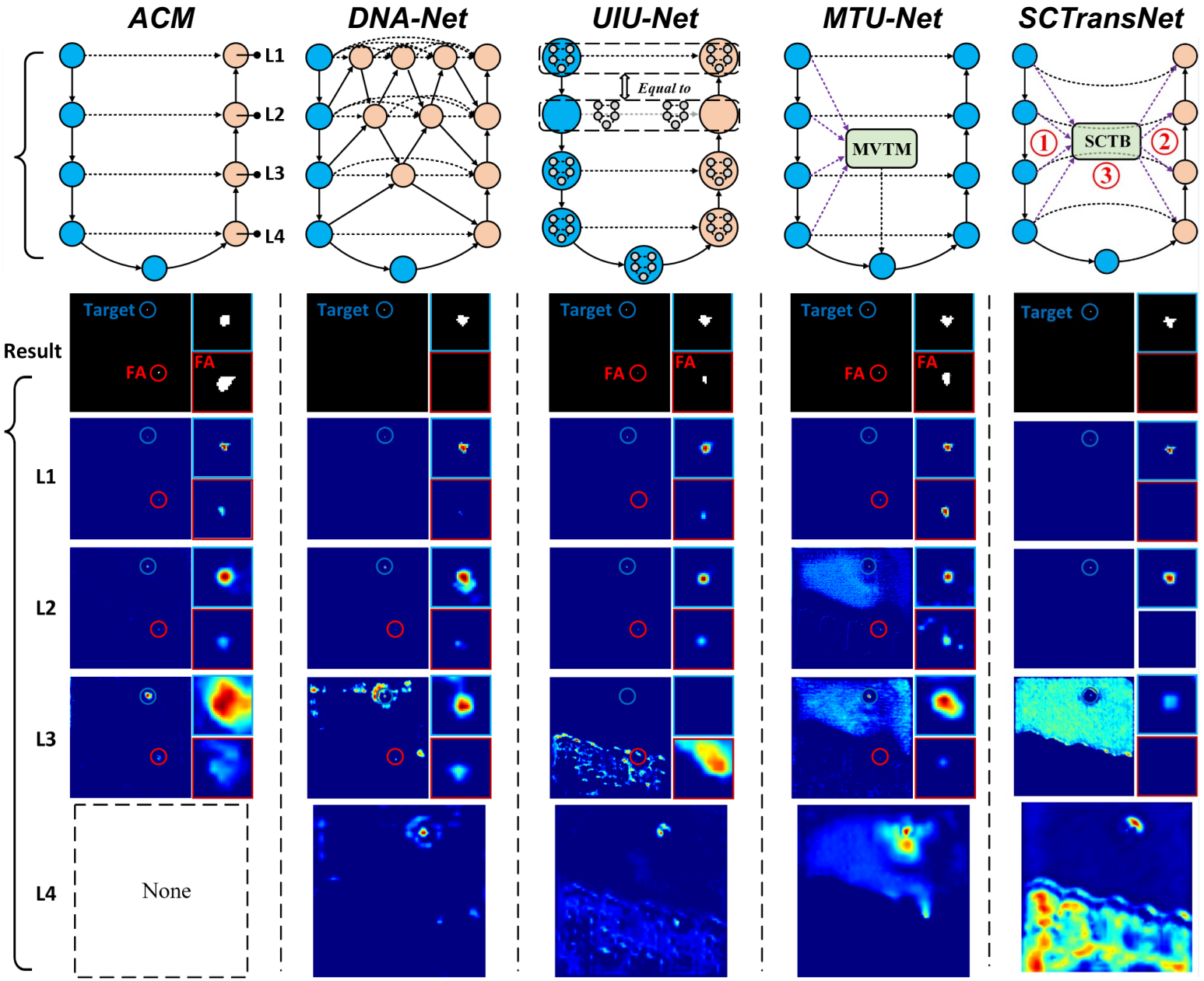
4) Feature fusion



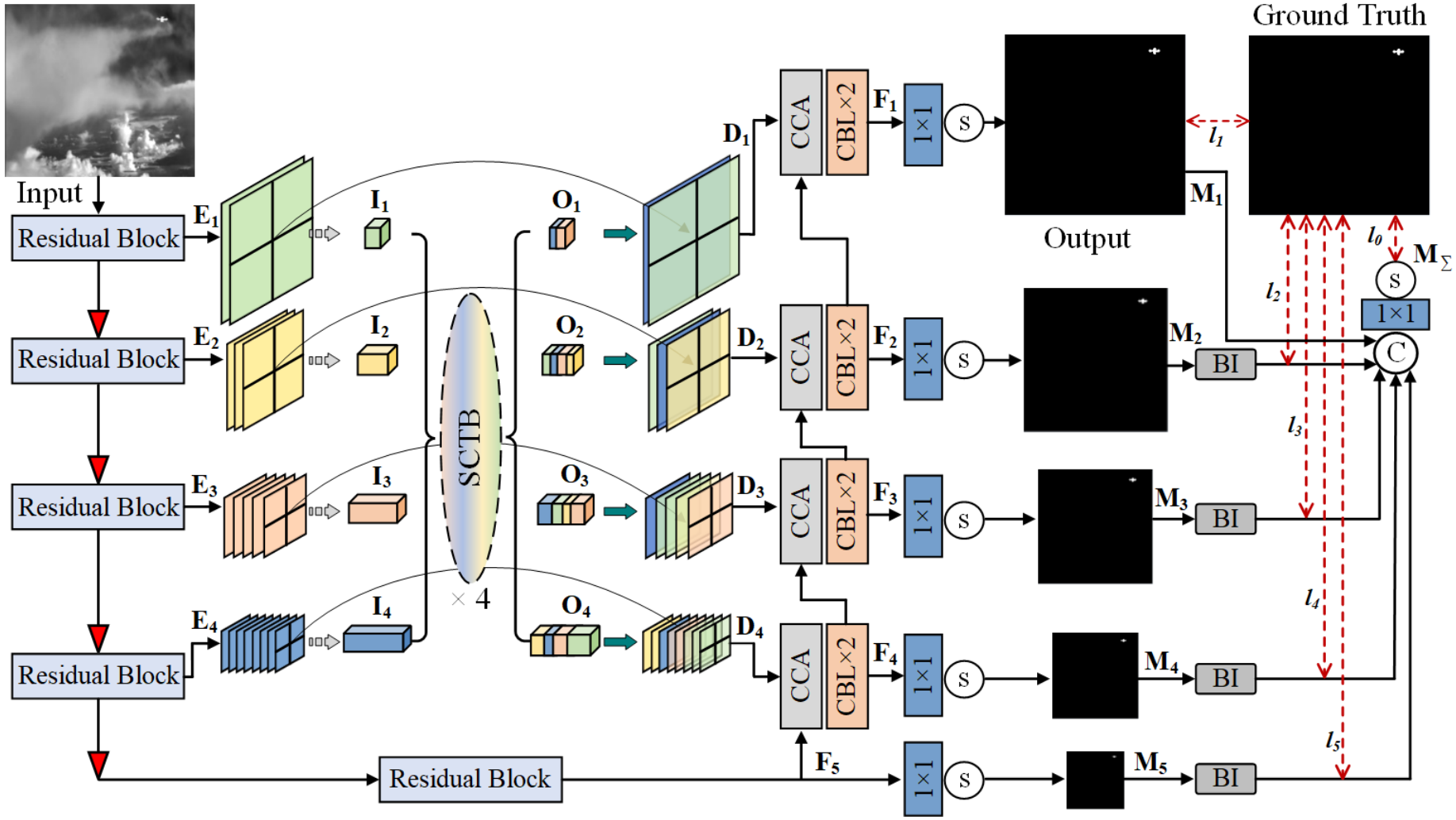
Three challenges



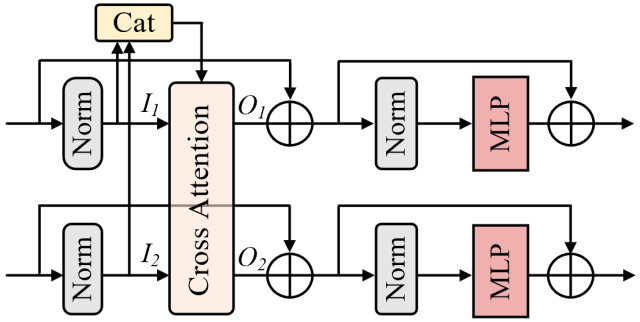
(b) Input image and GT



Framework


Input
Residual Block
Residual Block
Residual Block
Residual Block
Residual Block
SCTB
× 4
CCA
CBL×2
1×1
S
BI
C
Ground Truth
Output
Sigmoid Activation
Bilinear Interpolation
Channel Concatenate
Convolution 1×1
Maxpooling
Patch Embedding
Feature Mapping
Channel-wise Cross Attention
Convolution + Batch Normalization + LeakyReLU
Spatial-embedded Channel-cross Transformer Block

Spatial-channel Cross Transformer Block (SCTB)



Inspirations:

SSCA: 1) ***Local spatial embedding***

2) ***Multi-head Cross Attention***

CFN: 3) ***Mixed modelling of spatial and channels***

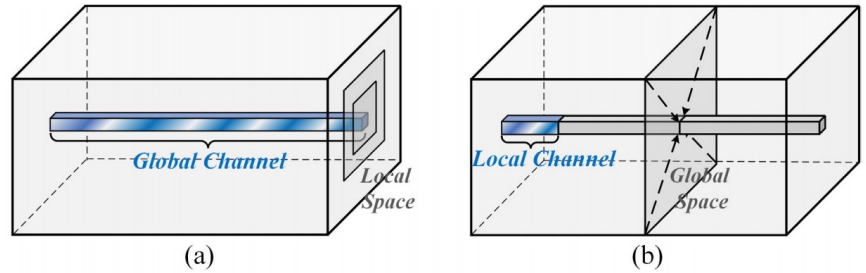
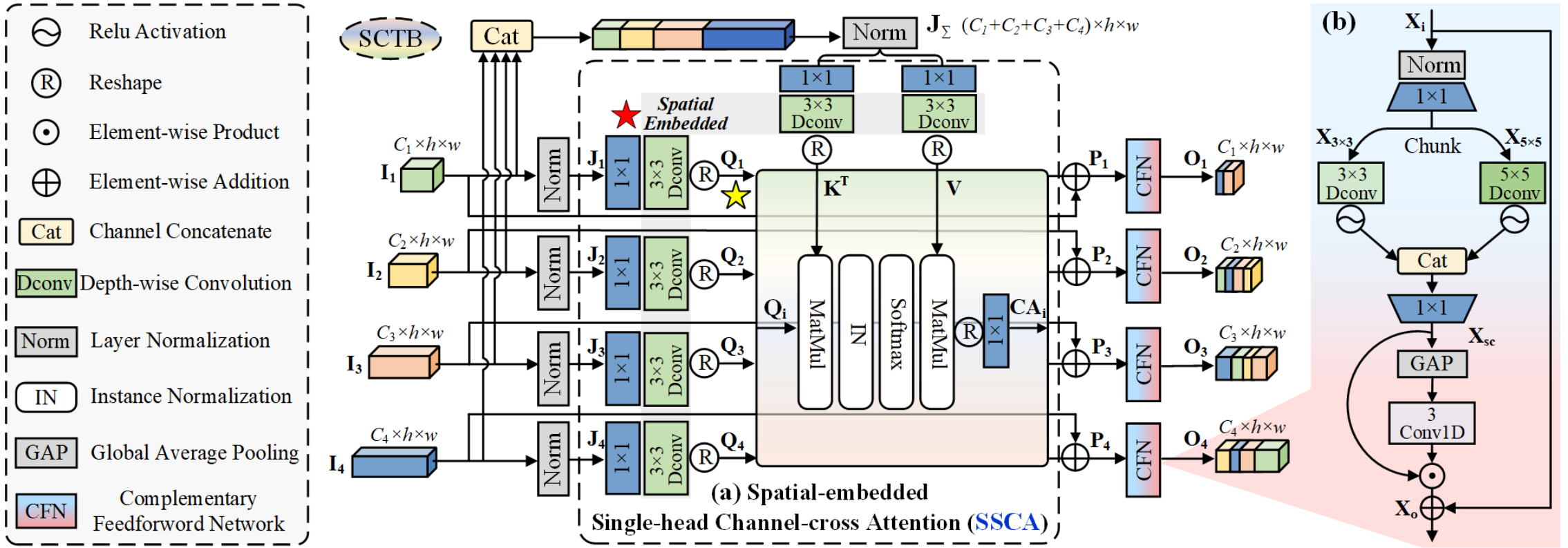


Fig. 4. Information enhancement from different perspectives. (a) LSGC paradigms. (b) GSLC paradigms. Our CFN integrates both of these information enhancement methods internally.



Quantitative Results

TABLE I

COMPARISONS WITH SOTA METHODS ON NUAA-SIRST, NUDT-SIRST AND IRSTD-1K IN $IoU(\%)$, $nIoU(\%)$, $F\text{-}Measure(\%)$, $Pd(\%)$, $Fa(10^{-6})$

Method	NUAA-SIRST [14]					NUDT-SIRST [15]					IRSTD-1K [40]				
	mIoU	nIoU	F-measure	Pd	Fa	mIoU	nIoU	F-measure	Pd	Fa	mIoU	nIoU	F-measure	Pd	Fa
Top-Hat [5]	7.143	18.27	14.63	79.84	1012	20.72	28.98	33.52	78.41	166.7	10.06	7.438	16.02	75.11	1432
Max-Median [54]	4.172	12.31	10.67	69.20	55.33	4.197	3.674	7.635	58.41	36.89	6.998	3.051	8.152	65.21	59.73
WSLCM [55]	1.158	6.835	4.812	77.95	5446	2.283	3.865	5.987	56.82	1309	3.452	0.678	2.125	72.44	6619
TLLCM [56]	1.029	4.099	4.995	79.09	5899	2.176	4.315	7.225	62.01	1608	3.311	0.784	2.186	77.39	6738
IPI [9]	25.67	50.17	43.65	84.63	16.67	17.76	15.42	26.94	74.49	41.23	27.92	20.46	35.68	81.37	16.18
PSTNN [57]	30.30	33.67	39.16	72.80	48.99	14.85	23.57	35.63	66.13	44.17	24.57	17.93	37.18	71.99	35.26
MSLSTIPT [1]	10.30	15.93	18.83	82.13	1131	8.342	10.06	18.26	47.40	888.1	11.43	5.932	12.23	79.03	1524
ACM [14]	68.93	69.18	80.87	91.63	15.23	61.12	64.40	75.87	93.12	55.22	59.23	57.03	74.38	93.27	65.28
ALCNet [18]	70.83	71.05	82.92	94.30	36.15	64.74	67.20	78.59	94.18	34.61	60.60	57.14	75.47	92.98	58.80
RDIAN [39]	68.72	75.39	81.46	93.54	43.29	76.28	79.14	86.54	95.77	34.56	56.45	59.72	72.14	88.55	26.63
ISTDU [22]	75.52	79.73	86.06	96.58	14.54	89.55	90.48	94.49	97.67	13.44	66.36	63.86	79.58	93.60	53.10
MTU-Net [17]	74.78	78.27	85.37	93.54	22.36	74.85	77.54	84.47	93.97	46.95	66.11	63.24	79.26	93.27	36.80
IAANet [58]	74.22	75.58	85.02	93.53	22.70	90.22	92.04	94.88	97.26	8.32	66.25	65.77	78.34	93.15	14.20
AGPCNet [19]	75.69	76.60	85.26	96.48	14.99	88.87	90.64	93.88	97.20	10.02	66.29	65.23	79.58	92.83	13.12
DNA-Net [15]	75.80	79.20	86.24	95.82	**8.78**	88.19	88.58	93.73	**98.83**	9.00	65.90	66.38	79.44	90.91	12.24
UIU-Net [16]	76.91	79.99	86.95	95.82	14.13	93.48	93.89	96.63	98.31	7.79	66.15	66.66	79.63	**93.98**	22.07
SCTransNet	**77.50**	**81.08**	**87.32**	**96.95**	13.92	**94.09**	**94.38**	**96.95**	98.62	**4.29**	**68.03**	**68.15**	**80.96**	93.27	**10.74**

ROC

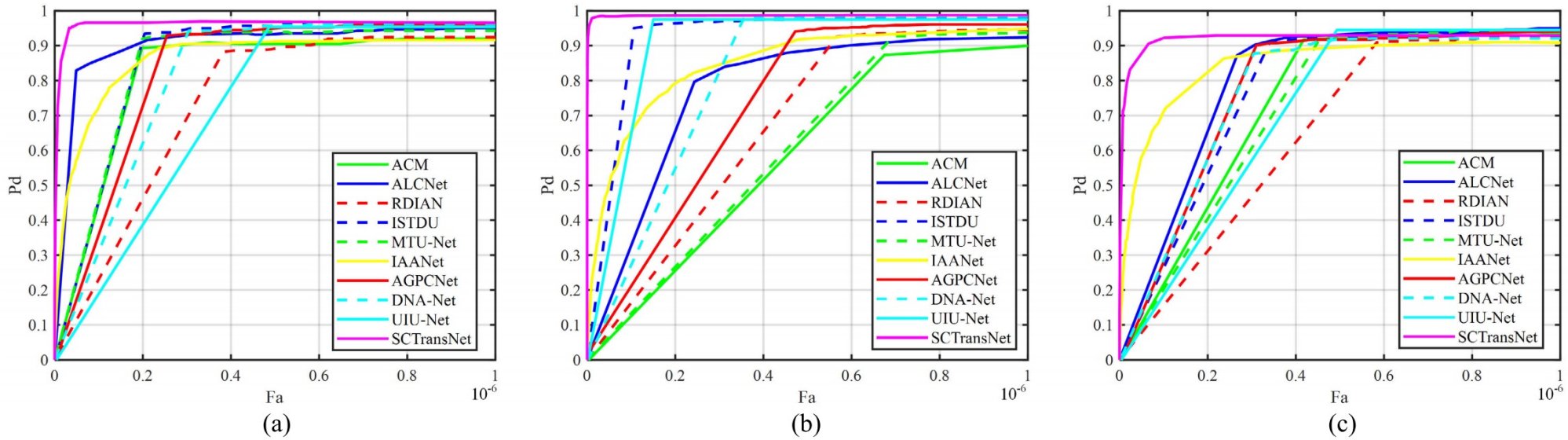


Fig. 5. ROC curves of different methods on the NUAA-SIRST, NUDT-SIRST, and IRSTD-1K dataset. Our SCTransNet can achieve the highest P_d at very low F_a. (a) NUAA-SIRST. (b) NUST-SIRST. (c) IRSTD-1K.

TABLE III

AUC WITH DIFFERENT THRESHOLDS OF THE SOTA METHODS ON THE NUAA-SIRST, NUDT-SIRST, AND IRSTD-1K DATASETS

Dataset	Index	ACM	ALCNet	RDIAN	ISTDU	MTU-Net	IAANet	AGPCNet	DNA-Net	UIU-Net	SCTransNet
NUAA-SIRST [14]	$AUC_{F_a=0.5}$	0.7223	<u>0.8618</u>	0.5461	0.7515	0.7457	0.8081	0.6953	0.6582	0.4854	**0.9539**
	$AUC_{F_a=1}$	0.8180	<u>0.9025</u>	0.7321	0.8579	0.8437	0.8614	0.8262	0.8098	0.7197	**0.9589**
NUDT-SIRST [15]	$AUC_{F_a=0.5}$	0.4392	0.6321	0.4630	<u>0.8635</u>	0.4640	0.7569	0.5038	0.6300	0.8275	**0.9853**
	$AUC_{F_a=1}$	0.5865	0.7716	0.6695	<u>0.9211</u>	0.6064	0.8463	0.7306	0.8072	0.9013	**0.9863**
IRSTD-1K [40]	$AUC_{F_a=0.5}$	0.5374	0.6606	0.4545	0.6014	0.5018	<u>0.7862</u>	0.6211	0.6162	0.4749	**0.9107**
	$AUC_{F_a=1}$	0.7366	0.8006	0.6480	0.7687	0.7198	<u>0.8456</u>	0.7752	0.7684	0.7099	**0.9200**

Ablation

Baselines Analysis

TABLE IV

BASED ON U-NET, ABLATION STUDY OF THE RBS, DEEP SUPERVISION (DS), SSCA, CFN, AND CCA MODULE IN AVERAGE *IoU*(%), *nIoU*(%), AND *F-Measure*(%) ON NUAA-SIRST, NUDT-SIRST, AND IRSTD-1K

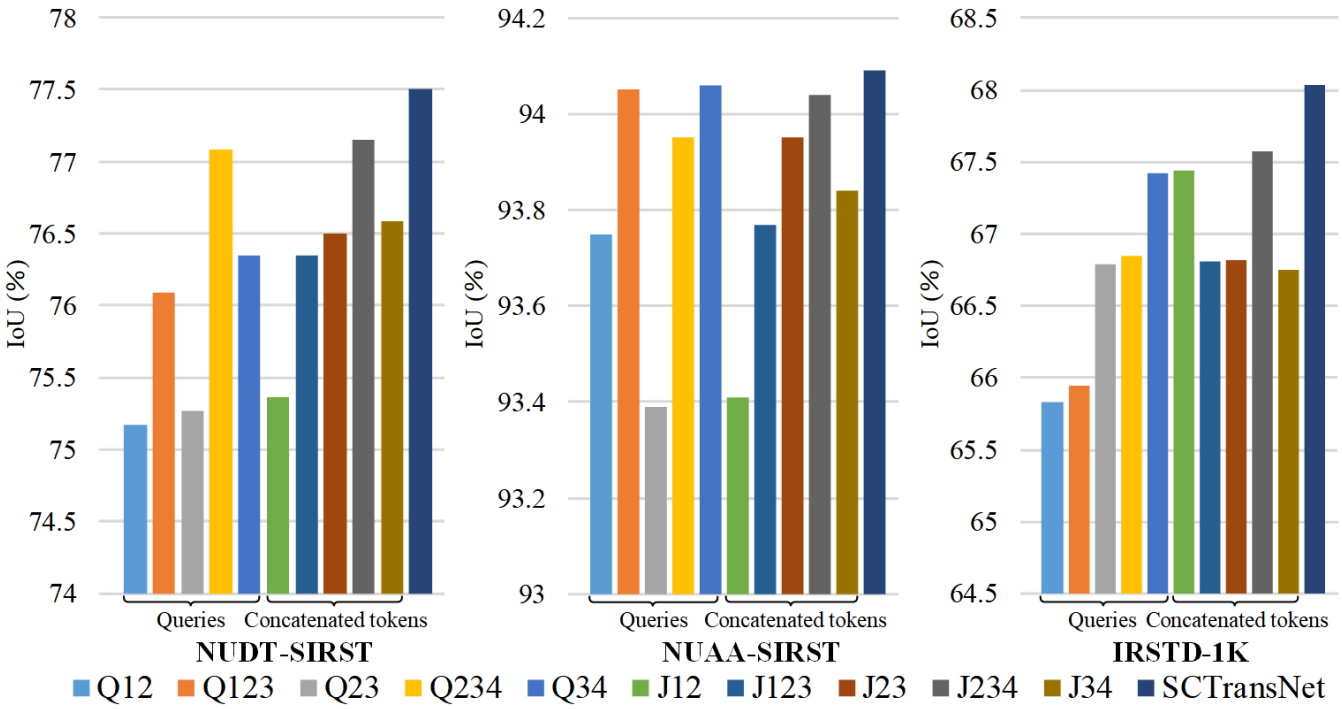
U-Net	+RBs	+DS	+SSCA	+CFN	+CCA	IoU	nIoU	F-measure
✓	✗	✗	✗	✗	✗	75.29	78.60	86.36
✓	✓	✗	✗	✗	✗	77.07	80.13	87.05
✓	✓	✓	✗	✗	✗	77.73	80.78	87.47
✓	✓	✓	✓	✗	✗	82.39	85.71	90.34
✓	✓	✓	✓	✓	✗	82.89	86.28	90.66
✓	✓	✓	✓	✓	✓	83.43	86.86	90.96

TABLE V

BASED ON UCTRANSNET, ABLATION STUDY OF THE RBS, DS, SKS AND SCTB, REPORTING AVERAGE *IoU*(%), *nIoU*(%), AND *F-Measure*(%) ON NUAA-SIRST, NUDT-SIRST, AND IRSTD-1K. NOTE THAT, WE REPLACE THE CCT IN UCTRANSNET USING THE PROPOSED SCTB

UCTransNet	+RBs	+DS	+SKs	SCTB r/ CCT	IoU	nIoU	F-measure
✓	✗	✗	✗	✗	78.78	81.56	87.80
✓	✓	✗	✗	✗	79.95	82.97	88.45
✓	✓	✓	✗	✗	81.47	83.89	88.92
✓	✓	✓	✓	✗	82.03	84.98	89.54
✓	✓	✓	✓	✓	83.43	86.86	90.66

Full Level Feature Interaction



Ablation

TABLE VI

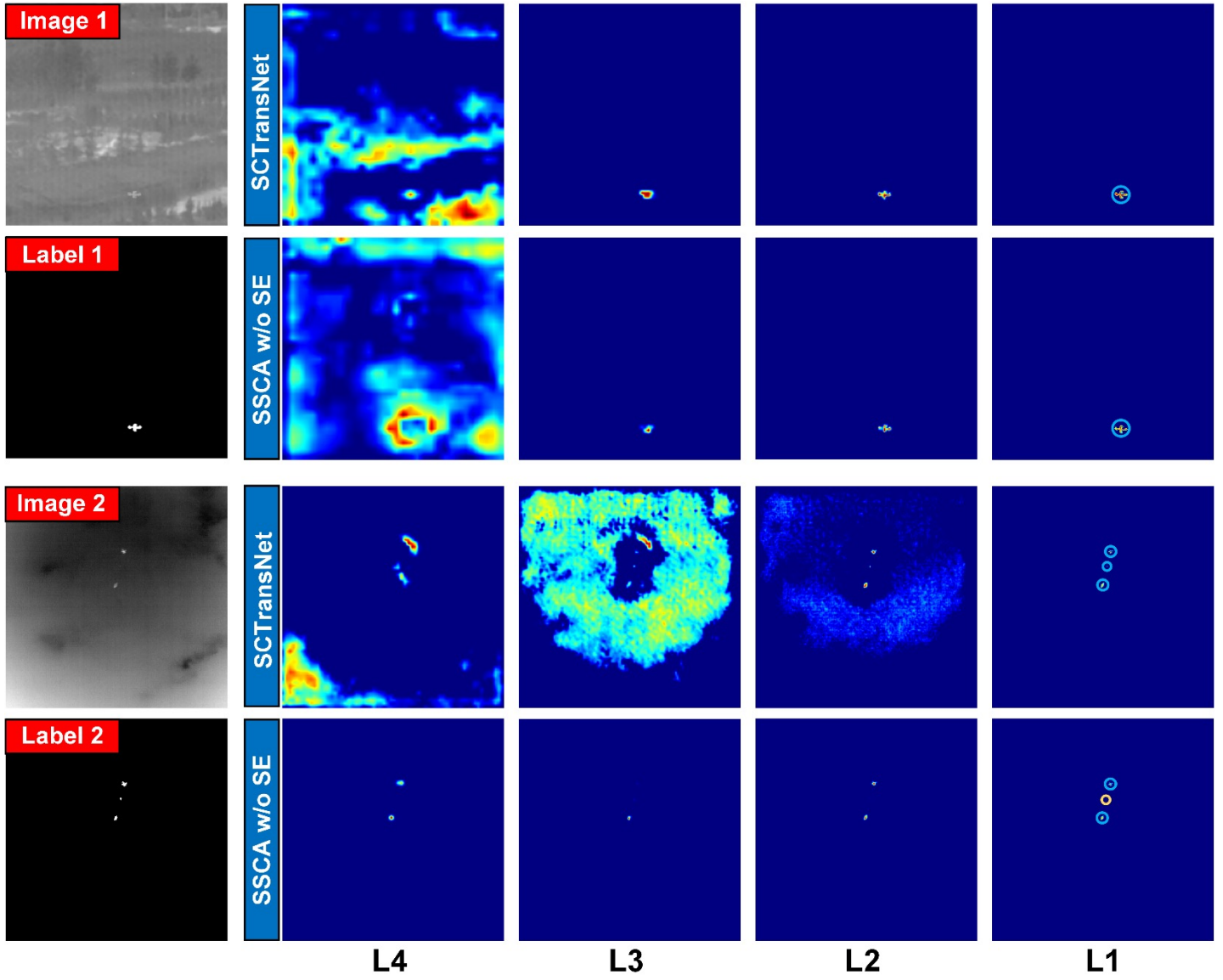
IoU(%)/nIoU(%)/F-Measure(%) Values Achieved by Variants of SSCA and MCA on NUAA-SIRST, NUDT-SIRST, and IRSTD-1K

Model	Dataset		
	NUAA-SIRST	NUDT-SIRST	IRSTD-1K
MCA [27]	74.72/78.35/85.53	93.07/93.61/96.41	65.60/66.57/79.22
SSCA w PE	77.10/79.88/87.07	94.03/94.25/96.93	66.01/65.29/79.52
SSCA w MH	76.35/79.56/86.59	93.72/94.13/96.76	67.08/67.55/80.30
SSCA w/o SE	76.40/79.19/86.62	93.23/93.49/96.50	66.10/65.48/79.59
SSCA	**77.50/81.08/87.32**	**94.09/94.38/96.95**	**68.03/68.15/80.96**

SSCA w PE: Incorporate positional encoding during the patch embedding stage.

SSCA w MH: Use a typical multi-head cross-attention mechanism to replace the single-head cross attention mechanism in SSCA.

SSCA w/o SE: To validate the effectiveness of local spatial information coding, we eliminate the depth-wise convolution in the QKV matrix generation process in SCTB.



Ablation

TABLE VII

IoU(%)/nIoU(%) Values Achieved by the Representative FFNs and the Variants of CFN on NUAA-SIRST and NUDT-SIRST

Model	Params(M)	Flops(G)	Dataset	
			NUAA-SIRST	NUDT-SIRST
FFN [41]	**11.0292**	**20.1474**	76.87/80.08	93.58/93.85
LeFF [32]	11.1312	20.1944	76.49/80.21	93.92/94.07
GDFN [30]	10.1841	19.7210	75.48/79.32	93.40/93.64
MSFN [31]	11.7107	20.5026	77.35/79.89	93.88/94.24
CFN w/o GSLC	11.1905	20.2362	76.54/80.56	93.95/94.18
CFN	11.1905	20.2372	**77.50/81.08**	**94.09/94.38**

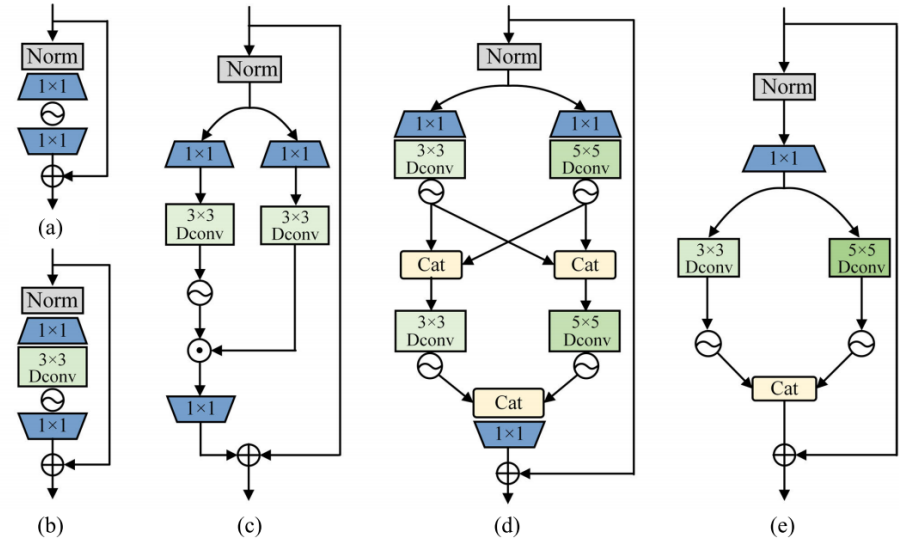
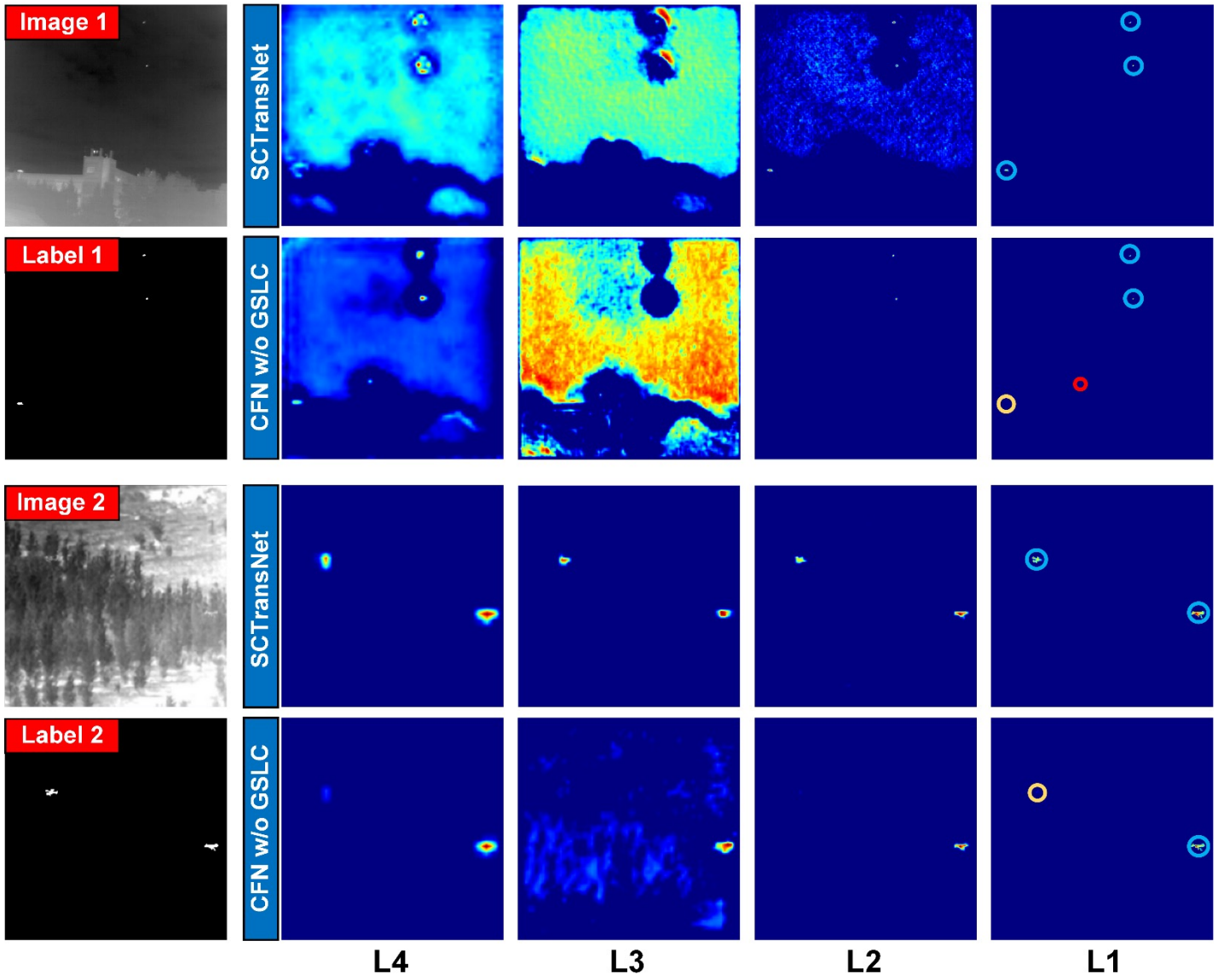


Fig. 10. Structure of representative FFNs and CFN w/o GSLC. (a) FFN. (b) LeFF. (c) GDFN. (d) MSFN. (e) CFN w/o GSLC.

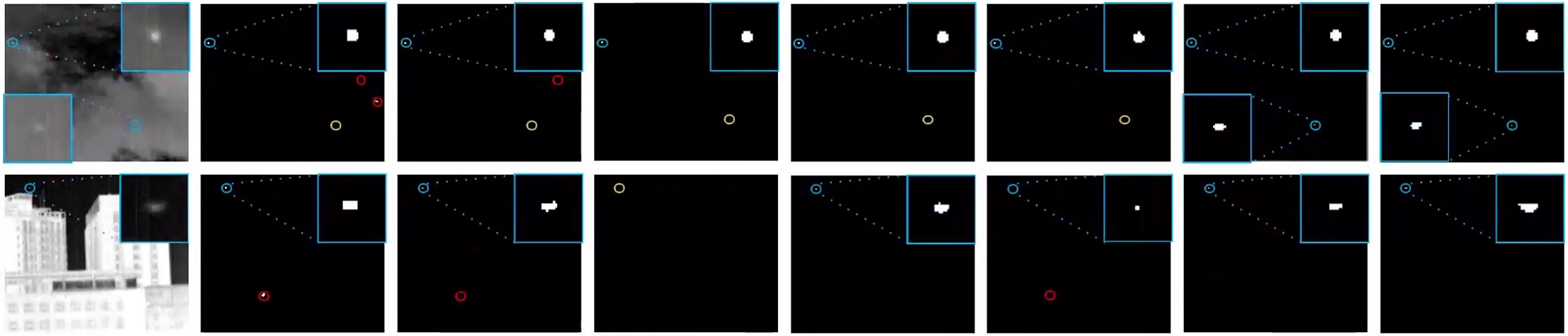


Ablation

TABLE VIII

IoU(%)/nIoU(%) Values Achieved by the Different Cross-Layer Feature Fusing Modules on NUAA-SIRST and NUDT-SIRST

Dataset								
✗	✗	✗	✗	82.29	85.77	90.26	20.0212	11.1462
✓	✗	✗	✗	82.33	85.89	90.31	20.0967	11.1484
✓	✓	✗	✗	82.49	86.11	90.40	20.1680	11.1569
✓	✓	✓	✗	82.95	86.27	90.68	20.2372	11.1905
✓	✓	✓	✓	83.43	86.86	90.96	20.2372	11.1905



(a) Stripy (b) ACM (c) RDIAN (d) DNA-Net (e) UIU-Net (f) MTU-Net (g) SCTransNet (h) GT

TABLE X

Hyperparameter Study of the Number of SCTBs, the Channel Expansion Factor of CFN, and the Basic Width of the Model in Average IoU(%), nIoU(%), *F*-Measure(%) on NUAA-SIRST, NUDT-SIRST, and IRSTD-1K

Hyper-param	IoU	nIoU	F-measure	Params(M)	Flops(G)
η = 3.00	83.24	86.69	90.84	20.4938	11.6917
η = 3.99	83.10	86.60	90.77	21.2306	13.1307
The basic width of the model					
W = 8	77.52	80.55	87.33	1.3321	0.7468
W = 16	81.02	84.50	89.51	5.1488	2.8609
W = 32	83.43	86.86	90.96	20.2372	11.1905
W = 48	82.95	86.48	90.60	45.2687	24.994

Summary

SCTransNet: Spatial-Channel Cross Transformer Network for Infrared Small Target Detection

Thanks

Project code is available at:
https://github.com/xdFai/SCTransNet

WeChat

yuansy@stu.xidian.edu.cn

yuansy2@student.unimelb.edu.au