

# Lightweight Deep Unfolding Network With Enhanced Robustness For Infrared Small Target Detection

Xianchao Xiu

Department of Automation

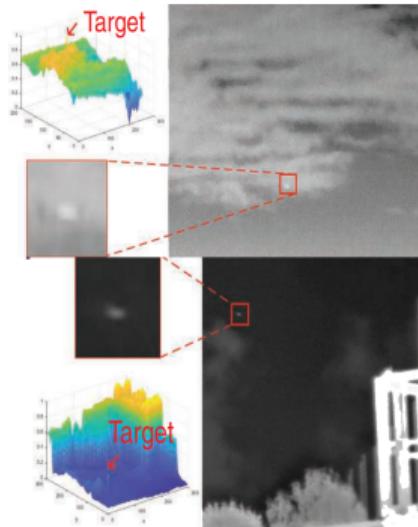
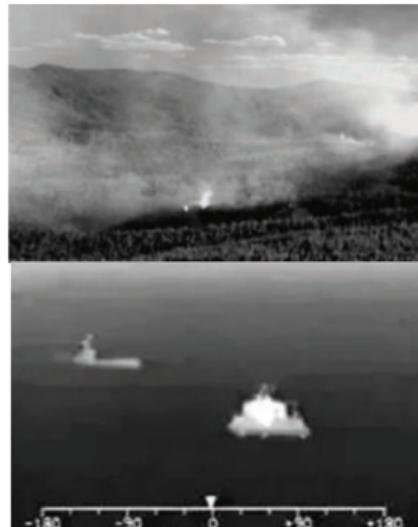


Mathematical Optimization Society, May 16-19, 2025

Joint work with [Yinчao Han](#) (SHU), [Jingjing Liu](#) (SHU) and [Wanquan Liu](#) (SYSU)

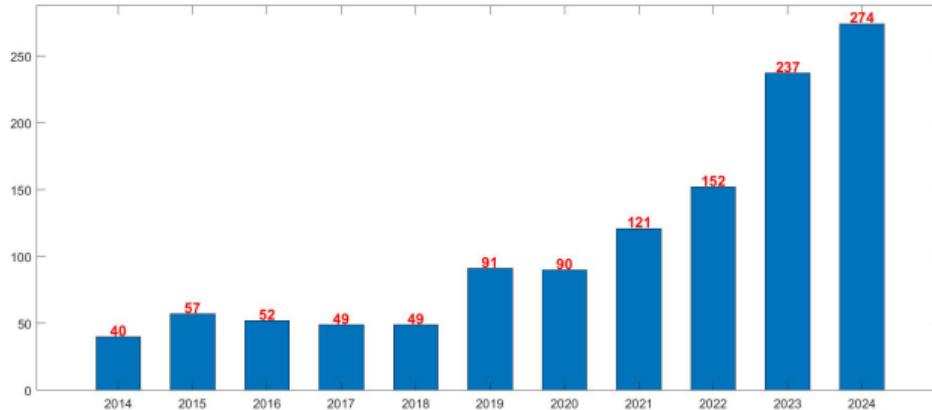
# ISTD

- ▶ Infrared small target detection (ISTD)



- ▶ Difficulties: **small size, low SNR, weak contrast**
- ▶ Advantages: **strong concealment, good anti-interference**

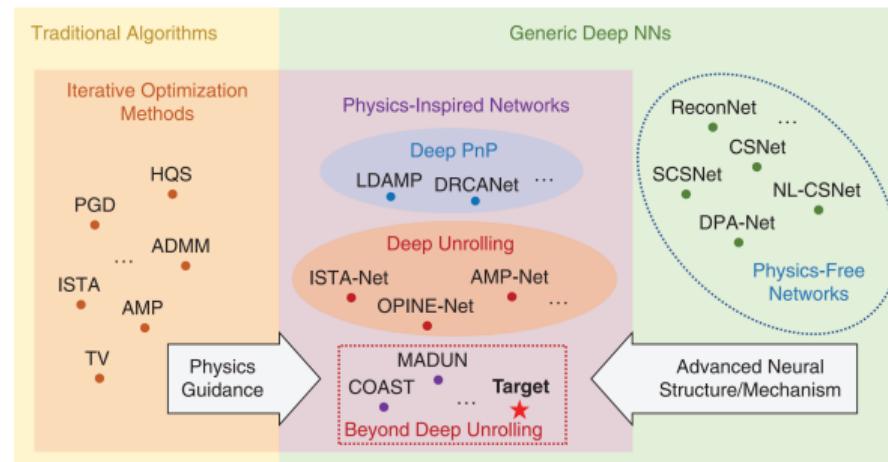
- ▶ Trends of “infrared small target detection” on Google Scholar



- ▶ Existing methods
  - ▶ Filter-based: spatial domain, transformed domain
  - ▶ Local information-based: local contrast, local entropy
  - ▶ Data structure-based: subspace, dictionary, tensor representation
  - ▶ Deep learning-based

# DUN

- ▶ From iterative optimization to deep unfolding networks
  - ▶ Gregor-LeCun, ICML, 2010
  - ▶ Yang-Sun-Li-Xu, IEEE TPAMI, 2020
  - ▶ Chen-Liu-Yin, Science China Mathematics, 2024



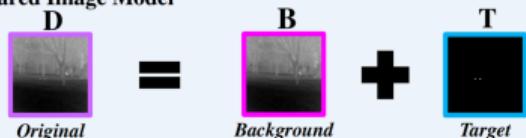
- ▶ <https://github.com/xianchaoxiu/AI4OPT>

# RPCANet

## ► From RPCA to RPCANet

$$\min_{B, T} \|B\|_* + \lambda \|T\|_1 \quad \text{s.t.} \quad D = B + T$$

(a) Infrared Image Model



(c) Model Iterative Solving

Updating of  $B^k$        $B^k = \text{prox}_{\mu}(\mathbf{D}^{k-1} - \mathbf{T}^{k-1})$

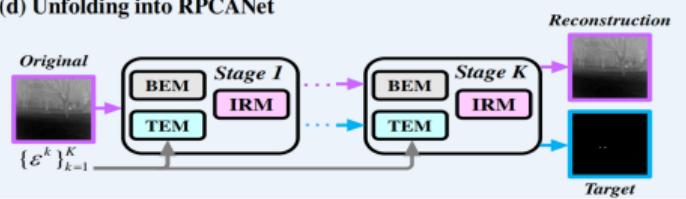
Updating of  $\mathbf{T}^k$        $\mathbf{T}^k = \mathbf{T}^{k-1} + \mathbf{D}^{k-1} - \mathbf{B}^k - \varepsilon \nabla S(\mathbf{T}^{k-1})$

Updating of  $\mathbf{D}^k$        $\mathbf{D}^k = \mathbf{B}^k + \mathbf{T}^k$

(b) RPCA Modeling

$$\begin{aligned} & \boxed{\text{Optimization Problem}} \quad \min_{\mathbf{B}, \mathbf{T}} \bar{\mathcal{R}}(\mathbf{B}) + \bar{\lambda} S(\mathbf{T}), \quad \text{s.t.} \quad \mathbf{D} = \mathbf{B} + \mathbf{T} \\ & \downarrow \\ & \boxed{\text{Unconstrained Problem}} \quad \mathcal{L}(\mathbf{B}, \mathbf{T}) = \mathcal{R}(\mathbf{B}) + \lambda S(\mathbf{T}) + \frac{\mu}{2} \|\mathbf{D} - \mathbf{B} - \mathbf{T}\|_F^2 \end{aligned}$$

(d) Unfolding into RPCANet



How to enhance robustness? How to realize lightweight?

# Methodology

- ▶ From RPCANet to RPCANet+

$$\min_{B, T} \|B\|_* + \lambda \|T\|_1$$

$$\text{s.t. } D = B + T$$

↓

$$\min_{B, T, N} \|B\|_* + \lambda \|T\|_1 + \mu \|N\|_F^2$$

$$\text{s.t. } D = B + T + N$$

↓

$$\min_{B, T, N} \mathcal{R}(B) + \lambda \mathcal{S}(T) + \mu \mathcal{G}(N)$$

$$\text{s.t. } D = B + T + N$$

- ▶ Unconstrained version

$$\mathcal{L}(B, T, N) = \mathcal{R}(B) + \lambda \mathcal{S}(T) + \mu \mathcal{G}(N) + \frac{\alpha}{2} \|D - B - T - N\|_F^2$$

## Update $B$

- ▶ Background estimation module + squeeze-and-excitation network (SEBEM)

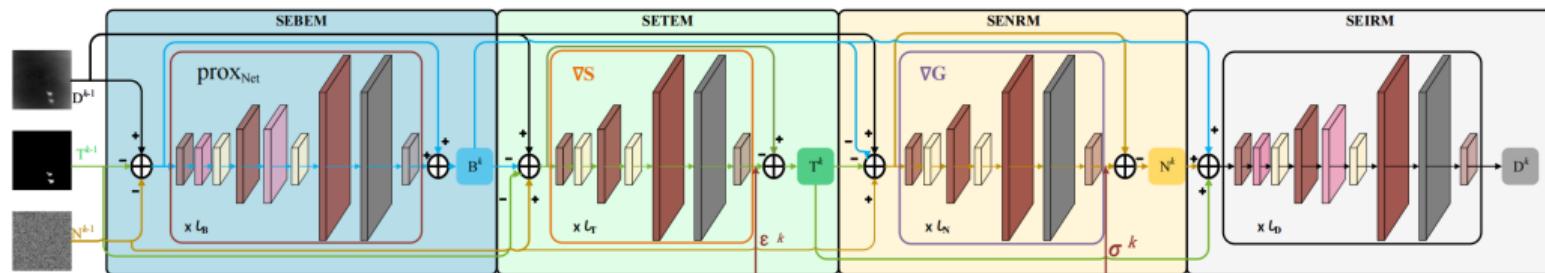
$$B^k = \arg \min_B \mathcal{R}(B) + \frac{\alpha}{2} \|D^{k-1} - B - T^{k-1} - N^{k-1}\|_F^2$$

↓

$$B^k = \text{prox}_{\alpha}(D^{k-1} - T^{k-1} - N^{k-1})$$

↓

$$B^k = D^{k-1} - T^{k-1} - N^{k-1} + \mathcal{W}^k(D^{k-1} - T^{k-1} - N^{k-1})$$



## Update $T$

- Target estimation module + squeeze-and-excitation network (SETEM)

$$T^k = \arg \min_T \lambda \mathcal{S}(T) + \frac{\alpha}{2} \|D^{k-1} - B^k - T - N^{k-1}\|_F^2$$

$\Downarrow$

$$T^k = \arg \min_T \frac{\lambda L_S}{2} \|T - T^{k-1} + \frac{1}{L_S} \nabla \mathcal{S}(T^{k-1})\|_F^2 + \frac{\alpha}{2} \|D^{k-1} - B^k - T - N^{k-1}\|_F^2$$

$\Downarrow$

$$T^k = \frac{\lambda L_S}{\lambda L_S + \alpha} T^{k-1} + \frac{\alpha}{\lambda L_S + \alpha} (D^{k-1} - B^k - N^{k-1}) - \frac{\lambda}{\lambda L_S + \alpha} \nabla \mathcal{S}(T^{k-1})$$

$\Downarrow$

$$T^k = \gamma T^{k-1} + (1 - \gamma)(D^{k-1} - B^k - N^{k-1}) - \varepsilon \nabla \mathcal{S}(T^{k-1})$$

## Update $T$

- Set  $\gamma = 0.5$

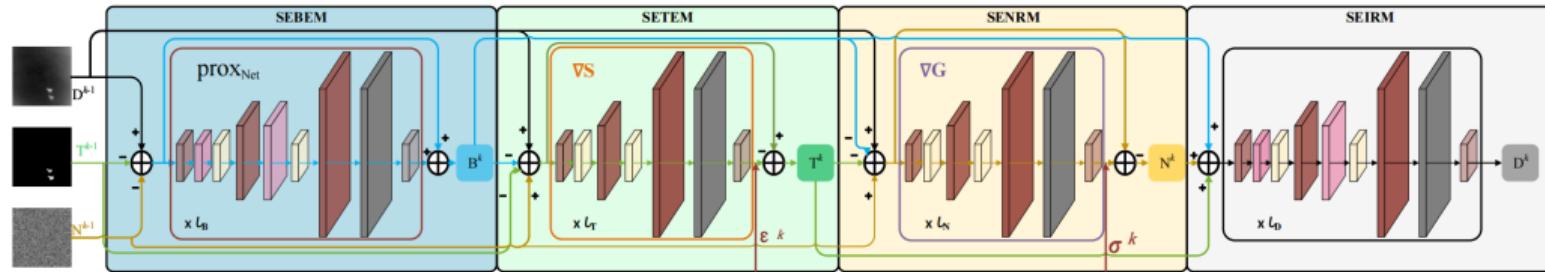
$$T^k = \gamma T^{k-1} + (1 - \gamma)(D^{k-1} - B^k - N^{k-1}) - \varepsilon \nabla S(T^{k-1})$$

↓

$$T^k = T^{k-1} + D^{k-1} - B^k - N^{k-1} - \varepsilon \nabla S(T^{k-1})$$

↓

$$T^k = T^{k-1} + D^{k-1} - B^k - N^{k-1} - \varepsilon^k \mathcal{H}^k (T^{k-1} + D^{k-1} - B^k - N^{k-1})$$



## Update $N$

- ▶ Noise reduction module + squeeze-and-excitation network (SENRM)

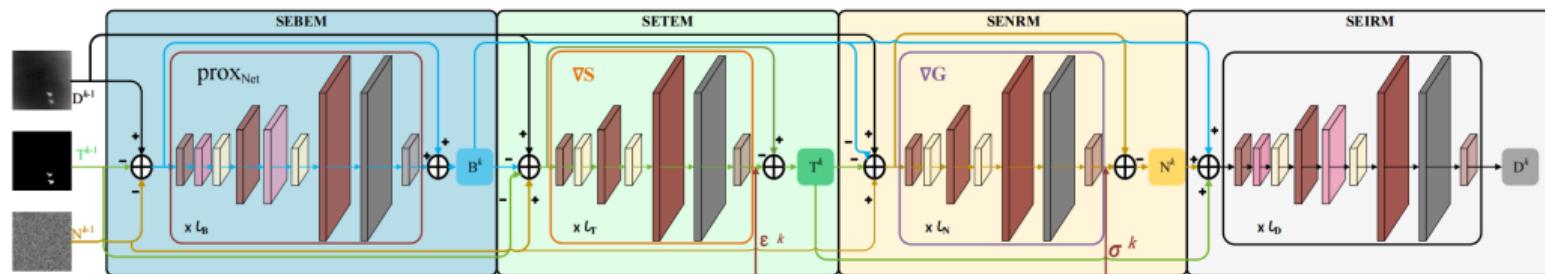
$$N^k = \arg \min_N \mu \mathcal{G}(N) + \frac{\alpha}{2} \|D^{k-1} - B^k - T^k - N\|_F^2$$

↓

$$N^k = \arg \min_N \frac{\mu L_N}{2} \|N - N^{k-1} + \frac{1}{L_N} \nabla \mathcal{G}(N^{k-1})\|_F^2 + \frac{\alpha}{2} \|D^{k-1} - B^k - T^k - N\|_F^2$$

↓

$$N^k = N^{k-1} + D^{k-1} - B^k - T^k - \sigma^k \mathcal{F}^k(N^{k-1} + D^{k-1} - B^k - T^k)$$



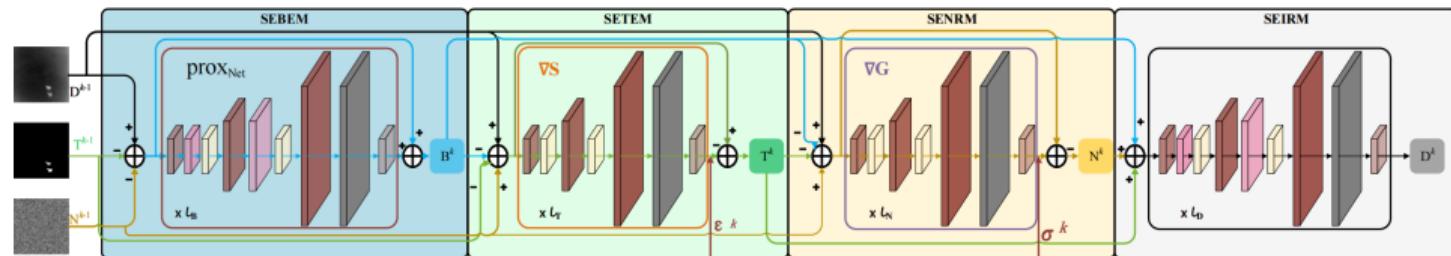
# Update $D$

- ▶ Image reconstruction module + squeeze-and-excitation network (SEIRM)

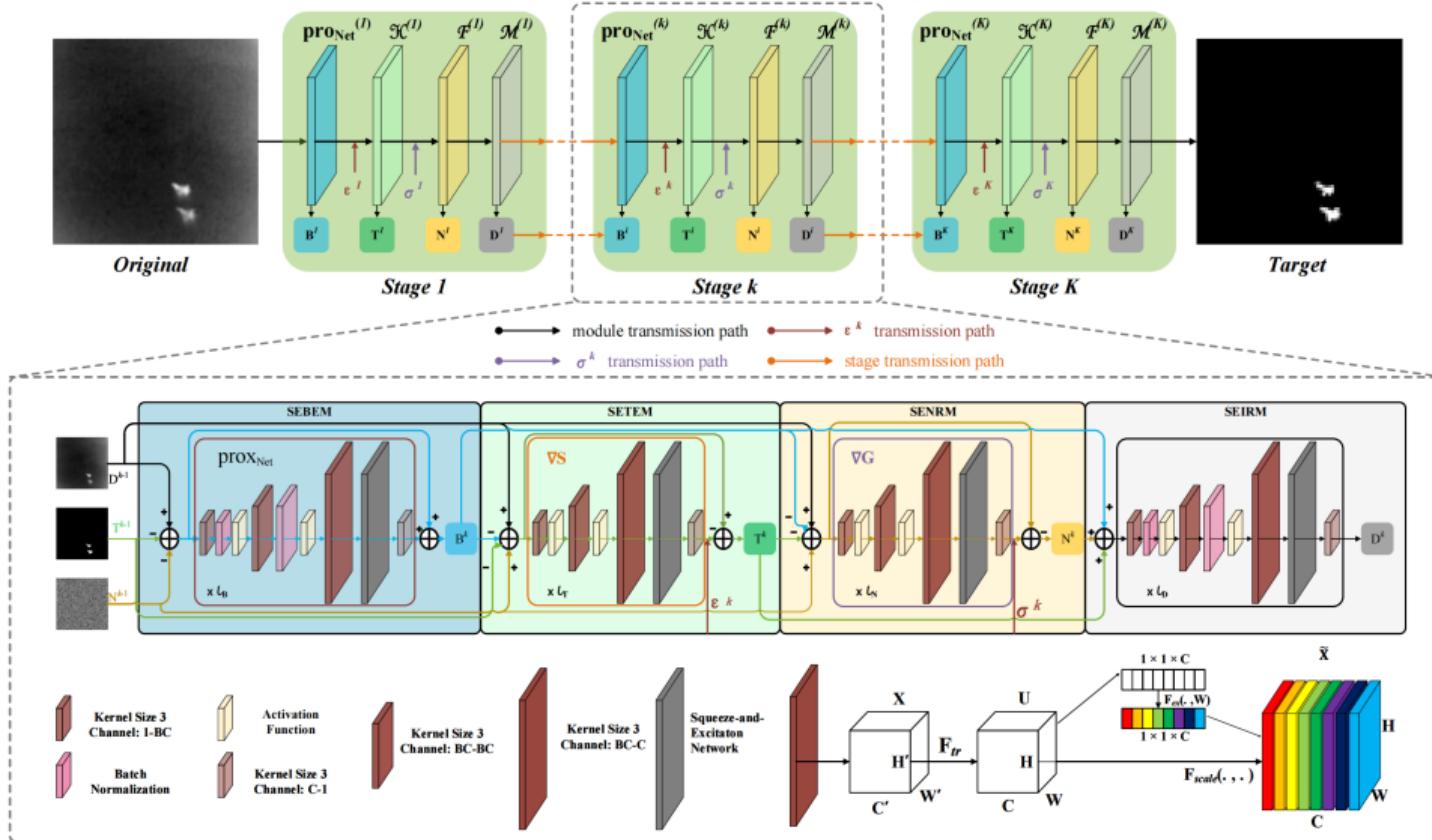
$$D^k = B^k + T^k + N^k$$

↓

$$D^k = \mathcal{M}^k(B^k + T^k + N^k)$$



# Architecture



# Experiment

- ▶ Compared methods
  - ▶ IPI: Gao-Meng-Yang-Wang-Zhou-Hauptmann, IEEE TIP, 2013
  - ▶ MPCM: Wei-You-Li, PR, 2016
  - ▶ PSTNN: Zhang-Peng, RS, 2019
  - ▶ AGPCNet: Zhang-Li-Cao-Pu-Peng, IEEE TAES, 2023
  - ▶ UIUNet: Wu-Hong-Chanussot, IEEE TIP, 2023
  - ▶ MSHNet: Liu-Liu-Zheng-Wang-Fu, CVPR, 2024
  - ▶ RPCANet: Wu-Zhang-Li-Huang-Peng, WACV, 2024
- ▶ Evaluation metrics
  - ▶ Mean intersection over union ( $mIoU \uparrow$ ),  $F_1$ -score ( $F_1 \uparrow$ ), Probability of detection ( $P_d \uparrow$ )
  - ▶ False alarm rate ( $F_a \downarrow$ )
- ▶ Loss function

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{segmentation}} + \eta \mathcal{L}_{\text{fidelity}} = \left( 1 - \frac{1}{M_t} \sum_{i=1}^{M_t} \frac{\text{TP}}{\text{FP} + \text{TP} + \text{FN}} \right) + \frac{\eta}{M_t M} \sum_{i=1}^{M_t} \|D^K - D\|_F^2$$

# Performance

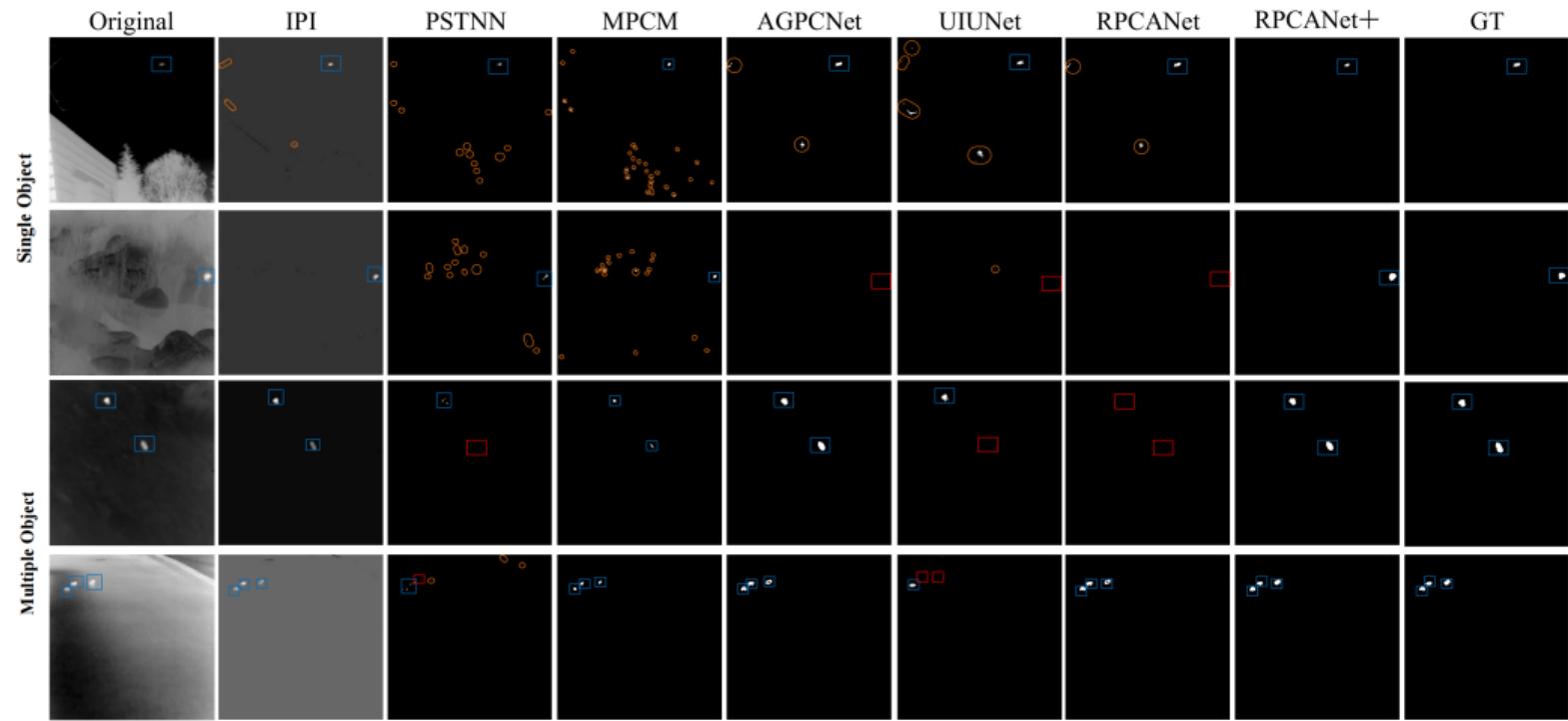
## ► Quantitative results

Methods	#Params	NUDT-SIRST				IRSTD-1k				SIRST-Aug				Time(s) CPU/GPU
		mIoU ↑	F <sub>1</sub> ↑	P <sub>d</sub> ↑	F <sub>a</sub> ↓	mIoU ↑	F <sub>1</sub> ↑	P <sub>d</sub> ↑	F <sub>a</sub> ↓	mIoU ↑	F <sub>1</sub> ↑	P <sub>d</sub> ↑	F <sub>a</sub> ↓	
IPI	-	34.83	51.49	92.58	7.14	18.67	31.48	78.54	11.11	21.90	35.97	80.36	<b>2.20</b>	3.0972/-
PSTNN	-	25.46	40.58	78.52	7.95	14.87	25.89	68.73	6.51	19.76	33.00	93.40	3.14	0.2249/-
MPCM	-	25.96	40.78	78.59	7.91	14.81	25.93	69.03	6.51	19.49	33.00	93.58	3.04	0.0624/-
AGPCNet	12.360M	85.41	92.15	98.10	4.72	61.00	75.75	89.35	5.34	72.36	83.83	99.03	35.56	-/0.0205
UIUNet	50.540M	88.71	94.01	91.43	1.89	63.06	77.35	<b>93.60</b>	6.57	71.80	83.59	98.35	28.29	-/0.0317
MSHNet	4.065M	75.99	86.57	96.07	2.63	64.50	77.55	91.68	4.46	71.64	84.16	90.78	23.09	-/0.0245
RPCANet	0.680M	89.31	94.35	97.14	2.87	63.21	77.45	88.31	<b>4.39</b>	72.54	84.08	98.21	34.14	-/0.0096
<b>RPCANet+</b>	<b>0.216M</b>	<b>92.37</b>	<b>96.04</b>	<b>98.41</b>	<b>1.79</b>	<b>64.68</b>	<b>78.55</b>	89.39	4.66	<b>74.56</b>	<b>85.43</b>	<b>99.17</b>	29.78	<b>-/0.0072</b>

	NUDT-SIRST	IRSTD-1k	SIRST-Aug
#Size	256 × 256	512 × 512	256 × 256
#Training	663	800	8525
#Testing	662	201	545

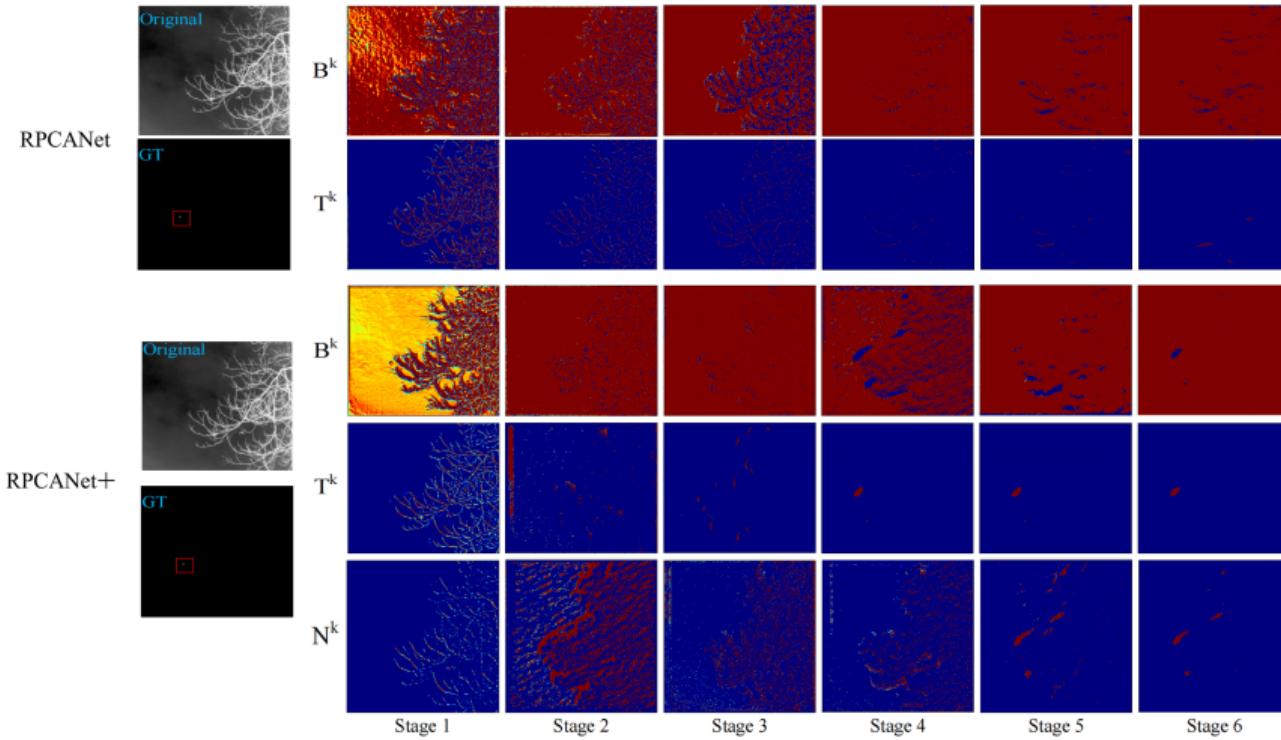
# Visualization

## ► Comparisons on IRSTD-1k



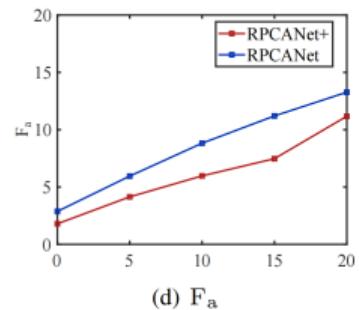
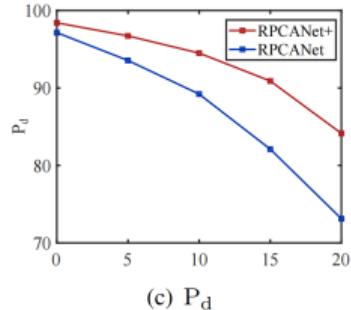
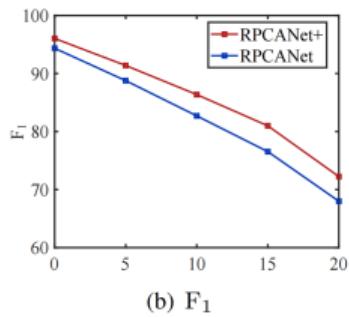
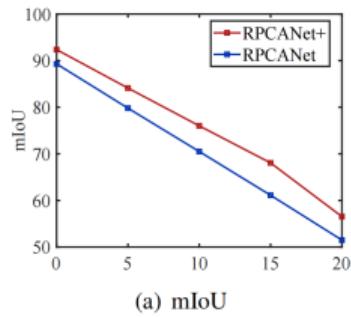
# Visualization

## ► A close look at stages

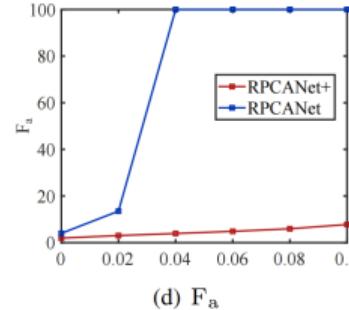
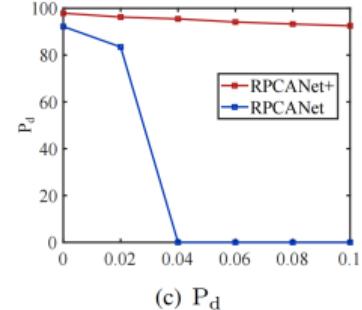
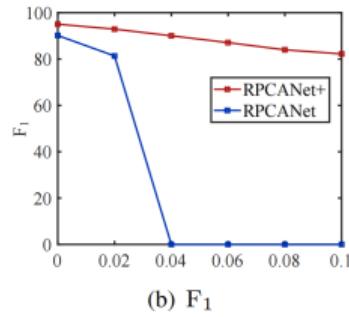
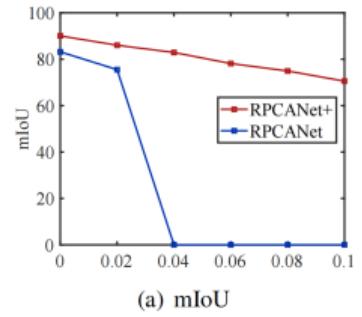


# Robustness

## ► Gaussian noise



## ► Salt-and-pepper noise



# Discussion

## ► Effects of stage $K$

K	#Params	NUDT-SIRST		IRSTD-1k		SIRST-Aug	
		mIoU ↑	F <sub>1</sub> ↑	mIoU ↑	F <sub>1</sub> ↑	mIoU ↑	F <sub>1</sub> ↑
1	0.0360M	72.83	84.28	91.39	95.51	61.26	75.98
2	0.0720M	72.96	84.37	91.61	95.62	61.59	76.24
3	0.1080M	73.58	84.78	91.53	95.58	62.60	77.00
4	0.1439M	73.81	84.93	90.06	95.06	61.98	76.53
5	0.1799M	74.28	85.24	91.64	95.64	63.45	77.63
<b>6</b>	<b>0.2159M</b>	<b>74.56</b>	<b>85.43</b>	<b>92.37</b>	<b>96.04</b>	<b>64.68</b>	<b>78.55</b>
7	0.2519M	72.29	83.92	88.53	93.58	62.29	76.76

## ► Effects of bottleneck channel (BC) and total channel (TC)

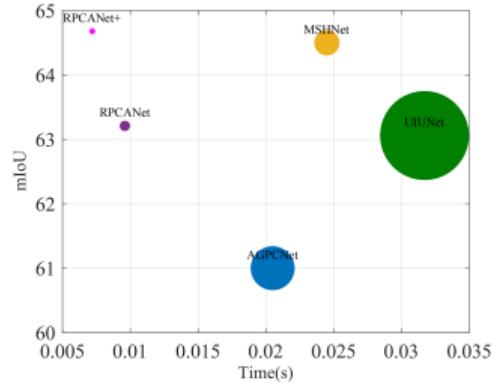
#BC	#TC	mIoU ↑	F <sub>1</sub> ↑
<b>4</b>	<b>32</b>	<b>74.56</b>	<b>85.43</b>
4	40	72.54	83.08
4	48	72.10	82.59
4	56	70.34	80.08
4	64	68.54	78.35
8	32	73.96	85.03
16	32	71.19	83.17

# Conclusion

## ► Conclusion

- How to enhance robustness?  $\Rightarrow$  RPCA + noise reduction + attention mechanism
- How to realize lightweight?  $\Rightarrow$  intermediate bottleneck + multi-layer mapping

Methods	mIoU $\uparrow$	F <sub>1</sub> $\uparrow$	P <sub>d</sub> $\uparrow$	F <sub>a</sub> $\downarrow$
IPI	25.13	39.65	83.83	6.81
PSTNN	20.03	33.16	80.22	5.87
MPCM	20.09	33.24	80.40	<b>5.82</b>
<b>RPCANet+</b>	<b>77.20</b>	<b>86.67</b>	<b>95.66</b>	<b>12.08</b>



## ► What about the convergence?

- Ryu-Liu-Wang-Chen-Wang-Yin, ICML, 2019
- Mukherjee-Hauptmann-Öktem-Pereyra-Schönlieb, IEEE SPM, 2023

## References

- ▶ Gao-Meng-Yang-Wang-Zhou-Hauptmann, Infrared Patch-Image Model for Small Target Detection in A Single Image, IEEE TIP, 2013
- ▶ Zhang-Ghanem, ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing, CVPR, 2018
- ▶ Zhao-Li-Li-Hu-Tao, Single-Frame Infrared Small-Target Detection: A Survey, IEEE GRSM, 2022
- ▶ Liu-Liu-Zheng-Wang-Fu, Infrared Small Target Detection with Scale and Location Sensitivity, CVPR, 2024
- ▶ Wu-Zhang-Li-Huang-Peng, RPCANet: Deep Unfolding RPCA Based Infrared Small Target Detection, WACV, 2024

Thank you for your attention!

[xcxiu@shu.edu.cn](mailto:xcxiu@shu.edu.cn)