

# Rethinking Sparse Optimization Through Deep Learning

Xianchao Xiu

Department of Automation



Mathematical Optimization Society, May 16-19, 2025

Joint work with [Long Chen](#) (SHU), [Jingjing Liu](#) (SHU), [Wanquan Liu](#) (SYSU) and others

# Outline

Introduction

Feature Selection

Infrared Small Target Detection

Conclusions and Future Work

# Sparse Optimization

- Sparse optimization considers

$$\min_{x \in \mathbb{R}^n} f(x) + \lambda \|x\|_0$$

$\Updownarrow$

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad \|x\|_0 \leq s$$

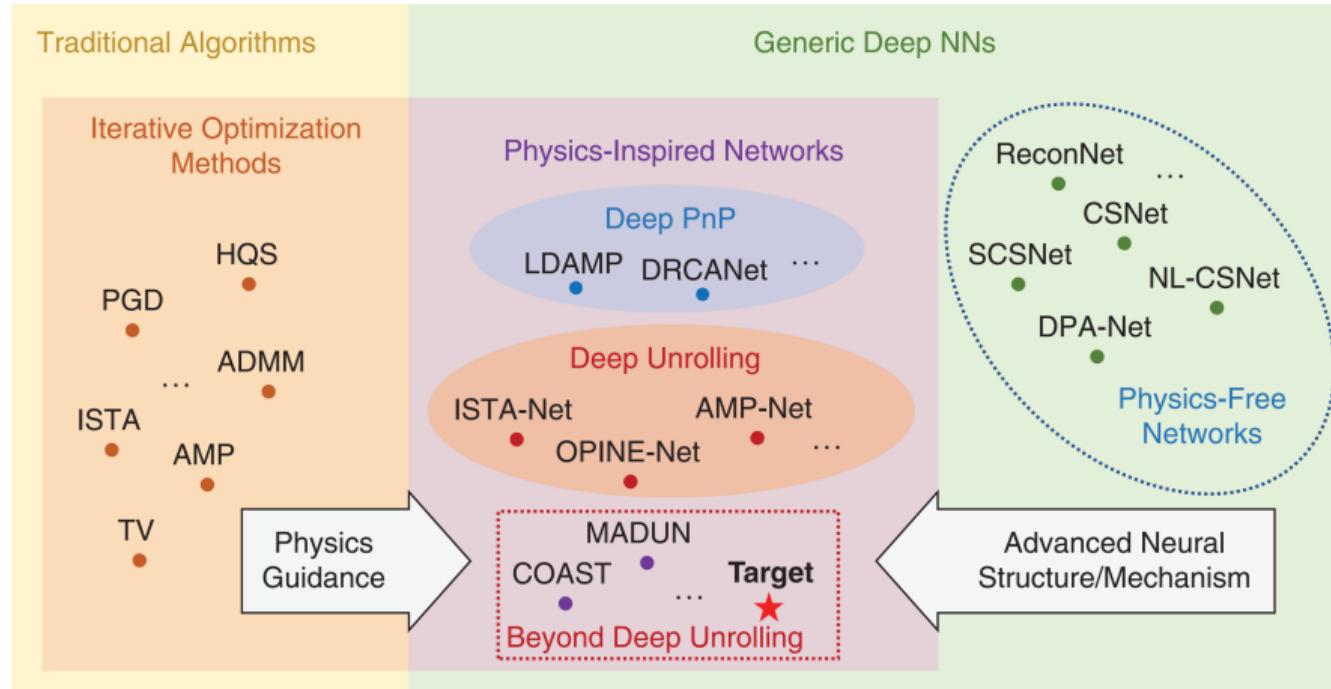
- $x$  can be extended to matrices and tensors
- $f(x)$  may be nonsmooth even nonconvex
- $\|x\|_0$  counts the number of nonzeros
- $\lambda$  and  $s$  are parameters
- Also compressed sensing, variable selection
- Broad applications: Machine learning, pattern recognition, engineering
- Efficient algorithms: Convex, nonconvex, direct

# Algorithms

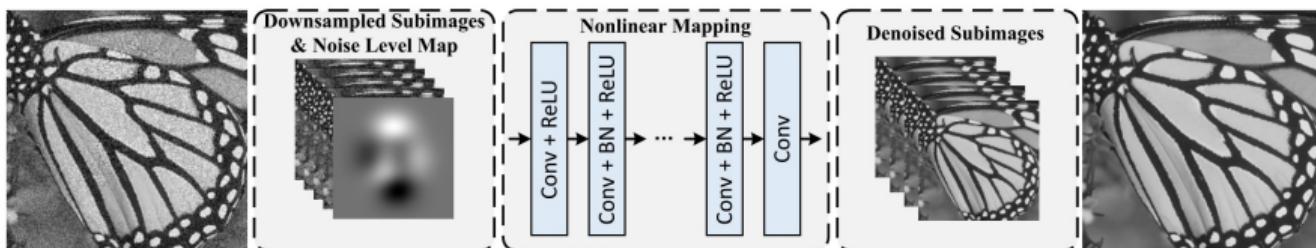
- ▶ Bach-Jenatton-Mairal-Obozinski, Optimization with Sparsity-Inducing Penalties, [Foundations and Trends in Machine Learning](#), 2012
- ▶ Jain-Kar, Non-Convex Optimization for Machine Learning, [Foundations and Trends in Machine Learning](#), 2017
- ▶ Hastie-Tibshirani-Wainwright, Statistical Learning with Sparsity: The Lasso and Generalizations, [CRC Press](#), 2015
- ▶ Zhao, Sparse Optimization Theory and Methods, [CRC Press](#), 2018
- ▶ Fan-Li-Zhang-Zou, Statistical Foundations of Data Science, [CRC Press](#), 2020
- ▶ Wright-Ma, High-Dimensional Data Analysis with Low-Dimensional Models: Principles, Computation, and Applications, [Cambridge University Press](#), 2022
- ▶ Parhi-Nowak, Deep Learning Meets Sparse Regularization: A Signal Processing Perspective, [IEEE SPM](#), 2023
- ▶ Tillmann-Bienstock-Lodi-Schwartz, Cardinality Minimization, Constraints, and Regularization: A Survey, [SIAM Review](#), 2024
- ▶ <https://sparseopt.github.io>

# Deep Learning

- From iterative optimization to deep learning

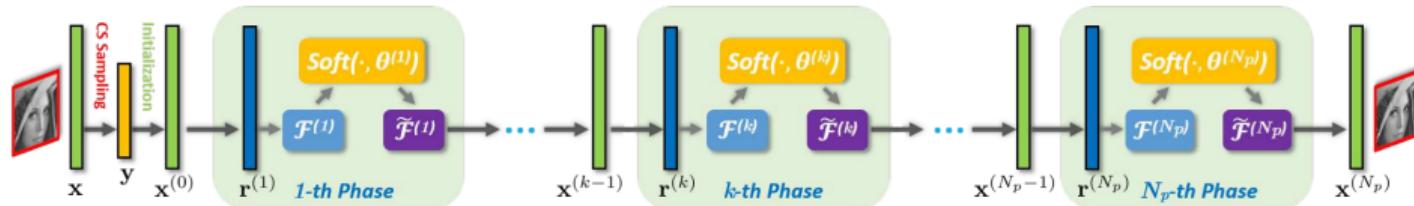


- ▶ Venkatakrishnan-Bouman-Wohlberg, Plug-and-Play Priors for Model Based Reconstruction, [IEEE GlobalSIP](#), 2013
- ▶ Zhang-Zuo-Chen et al, Beyond A Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising, [IEEE TIP](#), 2017
- ▶ Ryu-Liu-Wang et al, Plug-and-Play Methods Provably Converge with Properly Trained Denoisers, [ICML](#), 2019
- ▶ Mai-Lam-Lee, Attention-Guided Low-Rank Tensor Completion, [IEEE TPAMI](#), 2024
- ▶ Tan-Mukherjee-Tang et al, Provably Convergent Plug-and-Play Quasi-Newton Methods, [SIAM Journal on Imaging Sciences](#), 2025
- ▶ Liu-Feng-Xiu et al, Tensor Low-Rank Approximation via Plug-and-Play Priors for Anomaly Detection in Remote Sensing Images, [IEEE TIM](#), 2025



# DUN

- ▶ Gregor-LeCun, Learning Fast Approximations of Sparse Coding, [ICML](#), 2010
- ▶ Zhang-Ghanem, ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing, [CVPR](#), 2018
- ▶ Yang-Sun-Li et al, ADMM-CSNet: A Deep Learning Approach for Image Compressive Sensing, [IEEE TPAMI](#), 2020
- ▶ Chen-Chen-Chen et al, Learning to Optimize: A Primer and A Benchmark, [Journal of Machine Learning Research](#), 2022
- ▶ Zheng-Tang-Wagner et al, PhaseNet: A Deep Learning Based Phase Reconstruction Method for Ground-Based Astronomy, [SIAM Journal on Imaging Sciences](#), 2024
- ▶ Liu-Jin-Xiu et al, STAR-Net: An Interpretable Model-Aided Network for Remote Sensing Image Denoising, [Pattern Recognition](#), 2026



# Outline

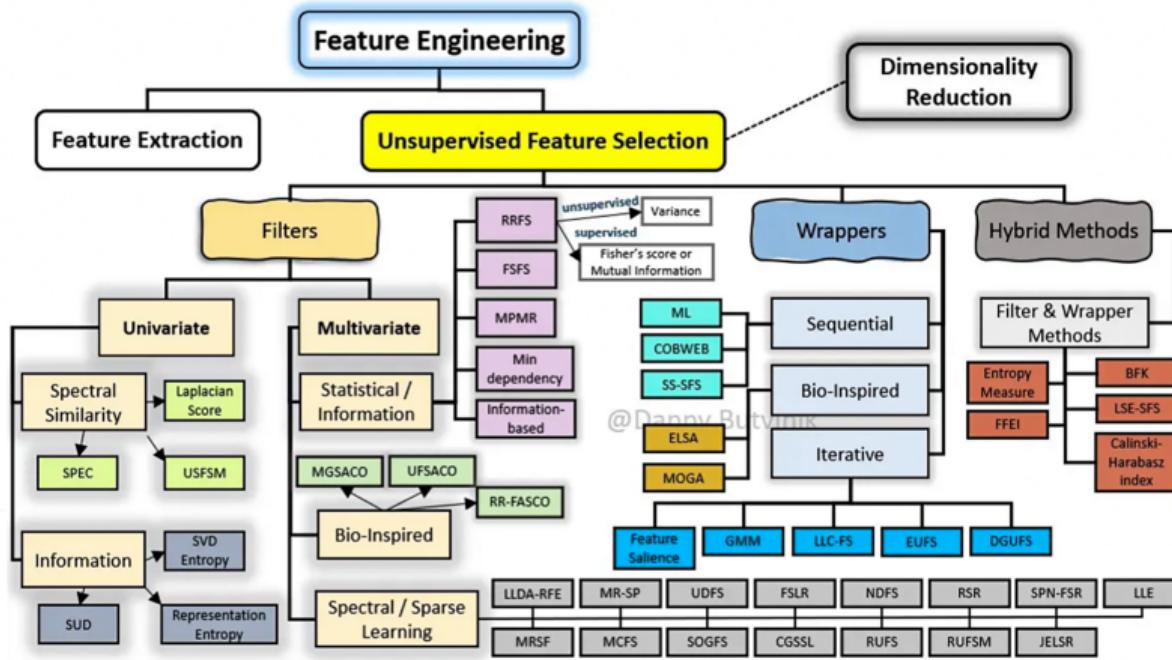
Introduction

Feature Selection

Infrared Small Target Detection

Conclusions and Future Work

- ▶ Unsupervised feature selection *vs.* Feature extraction
- ▶ Select a subset of input features without labels



# PCA

- Given  $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \in \mathbb{R}^{d \times n}$ , principal component analysis (PCA) is

$$\min_{W \in \mathbb{R}^{d \times p}} \frac{1}{2} \|X - WW^\top X\|_F^2$$

$$\text{s.t. } W^\top W = I_p$$

$\Updownarrow$

$$\min_{W \in \mathbb{R}^{d \times p}} -\text{Tr}(W^\top X X^\top W)$$

$$\text{s.t. } W^\top W = I_p$$

- Unsupervised feature selection by sparse PCA

$$\min_{W \in \mathbb{R}^{d \times p}} -\text{Tr}(W^\top X X^\top W)$$

$$\text{s.t. } W^\top W = I_p, \|W\|_{2,0} \leq s$$

- The  $i$ -th feature can be measured by  $\|\mathbf{w}^i\|$  since  $\mathbf{z}_i = (\mathbf{w}^{1\top}, \mathbf{w}^{2\top}, \dots, \mathbf{w}^{d\top})\mathbf{x}_i$
- The dimension number is often omitted when it does not cause ambiguity

## Motivation

- ▶ Li-Nie-Bian et al, Sparse PCA via  $\ell_{2,p}$ -Norm Regularization for Unsupervised Feature Selection, [IEEE TPAMI](#), 2023

$$\begin{aligned} \min_W \quad & -\text{Tr}(W^\top X X^\top W) + \lambda \|W\|_{2,p}^p \quad (0 < p < 1) \\ \text{s.t.} \quad & W^\top W = I \end{aligned}$$

- ▶ Li-Sun-Zhang, Unsupervised Feature Selection via Nonnegative Orthogonal Constrained Regularized Minimization, [arXiv](#), 2024

$$\begin{aligned} \min_{W,Y} \quad & \text{Tr}(Y^\top LY) + \alpha \|Y - X^\top W\|_{2,1} + \beta \|W\|_{2,1} + \gamma \|W\|_F^2 \\ \text{s.t.} \quad & Y^\top Y = I, \quad Y \geq 0 \end{aligned}$$

- ▶ Xiu-Huang-Shang et al, Bi-Sparse Unsupervised Feature Selection, [IEEE TIP](#), 2025

$$\begin{aligned} \min_W \quad & -\text{Tr}(W^\top X X^\top W) + \lambda \|W\|_{2,p}^p + \mu \|W\|_q^q \quad (0 \leq p < 1) \\ \text{s.t.} \quad & W^\top W = I \end{aligned}$$

# Model

- ▶ Consider structured sparse PCA

$$\begin{aligned} \min_W \quad & \frac{1}{2} \|X - WW^\top X\|_F^2 + \lambda \|W\|_{2,1} + \mu \|W\|_1 \\ \text{s.t.} \quad & W^\top W = I \end{aligned}$$

- ▶ Alternating direction method of multipliers (ADMM)

$$\begin{aligned} \min_W \quad & \frac{1}{2} \|X - WW^\top X\|_F^2 + \lambda \|Y\|_{2,1} + \mu \|Z\|_1 \\ \text{s.t.} \quad & W^\top W = I, \quad W = Y, \quad W = Z \end{aligned}$$

↓

$$\begin{aligned} \mathcal{L}(W, Y, Z, \Lambda, \Pi) = & \frac{1}{2} \|X - WW^\top X\|_F^2 + \lambda \|Y\|_{2,1} + \mu \|Z\|_1 \\ & + \langle \Lambda, W - Y \rangle + \frac{\alpha}{2} \|W - Y\|_F^2 + \langle \Pi, W - Z \rangle + \frac{\beta}{2} \|W - Z\|_F^2 \end{aligned}$$

# SPCA-Net

- ▶ Update  $W$ -block

$$\min_W \quad f(W) := \frac{1}{2} \|X - WW^\top X\|_F^2 + \frac{\alpha}{2} \|W - Y^k + \Lambda^k/\alpha\|_F^2 + \frac{\beta}{2} \|W - Z^k + \Pi^k/\beta\|_F^2$$

$$\text{s.t. } W^\top W = I$$

↓

$$\min_W \quad f(W^k) + \langle \nabla f(W^k), W - W^k \rangle + \frac{1}{2\eta} \|W - W^k\|_F^2$$

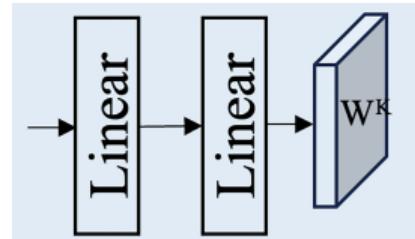
$$\text{s.t. } W^\top W = I$$

↓

$$W^{k+1} = UV^\top$$

↓

$$W^{k+1} = \text{LargNet}(U, V^\top)$$



# SPCA-Net

- ▶ Update  $Y$ -block

$$\min_Y \lambda \|Y\|_{2,1} + \frac{\alpha}{2} \|X^{k+1} - Y + \Lambda^k/\alpha\|_F^2$$

$\Downarrow$

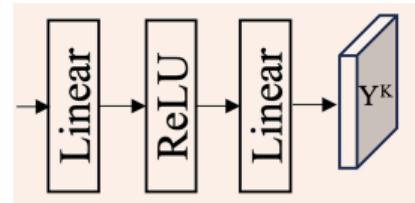
$$Y^{k+1} = \text{sign}(\|X^{k+1} + \Lambda^k/\alpha\|_2) \circ \max(\|X^{k+1} + \Lambda^k/\alpha\|_2 - \lambda/\alpha, 0)$$

$\Downarrow$

$$Y^{k+1} = \frac{X^{k+1} + \Lambda^k/\alpha}{\|X^{k+1} + \Lambda^k/\alpha\|_2} \text{ReLU}(\|X^{k+1} + \Lambda^k/\alpha\|_2 - \lambda/\alpha)$$

$\Downarrow$

$$Y^{k+1} = \text{GSoftNet}(X^{k+1} + \Lambda^k/\alpha, \lambda/\alpha)$$



# SPCA-Net

- ▶ Update  $Z$ -block

$$\min_Z \mu \|Z\|_1 + \frac{\beta}{2} \|X^{k+1} - Z + \Pi^k / \beta\|_F^2$$

↓

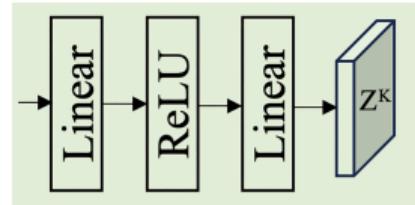
$$Z^{k+1} = \text{sign}(X^{k+1} + \Pi^k / \beta) \circ \max(|X^{k+1} + \Pi^k / \beta| - \mu / \beta, 0)$$

↓

$$Z^{k+1} = \frac{X^{k+1} + \Pi^k / \beta}{|X^{k+1} + \Pi^k / \beta|} \text{ReLU}(|X^{k+1} + \Pi^k / \beta| - \mu / \beta)$$

↓

$$Z^{k+1} = \text{SoftNet}(X^{k+1} + \Pi^k / \beta, \mu / \beta)$$



# SPCA-Net

- ▶ **Input:**  $X, \lambda, \mu, \alpha, \beta$
- ▶ **Initialize:**  $(W^0, Y^0, Z^0, \Lambda^0, \Pi^0)$
- ▶ **While**  $k = 1, \dots, K$  **do**
  - ▶ Update  $W^{k+1}$  by

$$W^{k+1} = \text{LargNet}(U, V^\top)$$

- ▶ Update  $Y^{k+1}$  by

$$Y^{k+1} = \text{GSoftNet}(X^{k+1} + \Lambda^k / \alpha, \lambda / \alpha)$$

- ▶ Update  $Z^{k+1}$  by

$$Z^{k+1} = \text{SoftNet}(X^{k+1} + \Pi^k / \beta, \mu / \beta)$$

- ▶ Update  $\Lambda^{k+1}, \Pi^{k+1}$  by

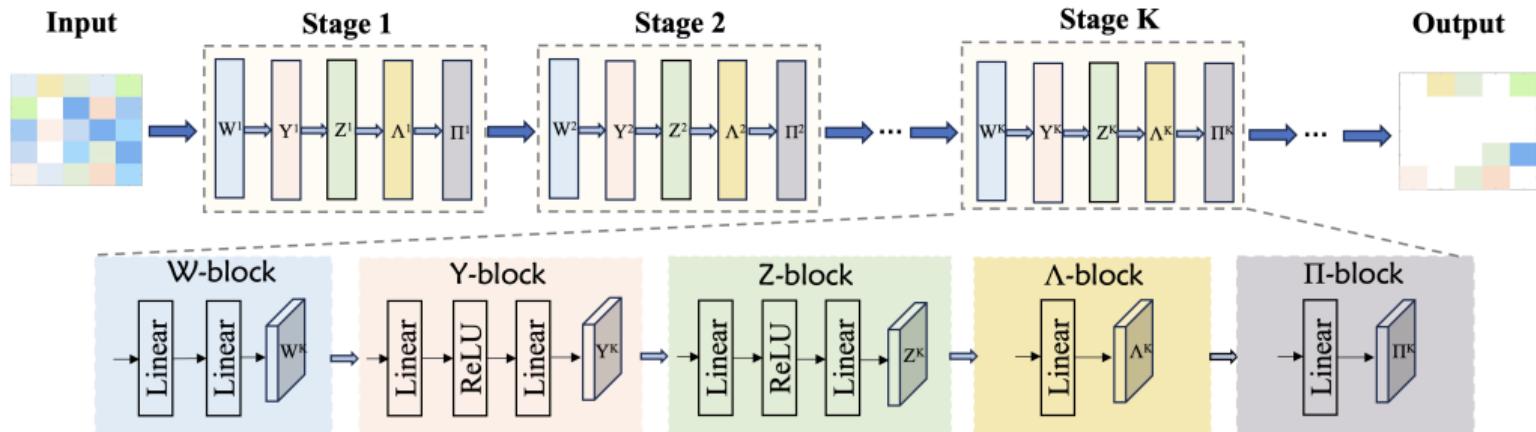
$$\Lambda^{k+1} = \text{Linear}(W^{k+1}, Y^{k+1}, \Lambda^k, \alpha), \quad \Pi^{k+1} = \text{Linear}(W^{k+1}, Z^{k+1}, \Pi^k, \beta)$$

- ▶ **Output:** Trained  $W$

# Architecture

- ▶ All parameters  $(\lambda, \mu, \alpha, \beta)$  are trained in an end-to-end manner
- ▶ The loss is defined as

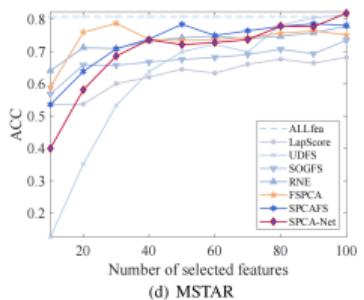
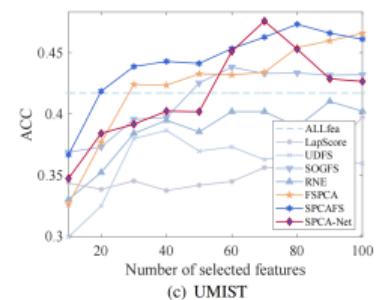
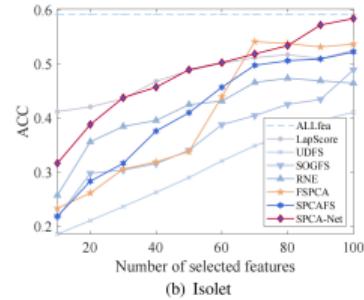
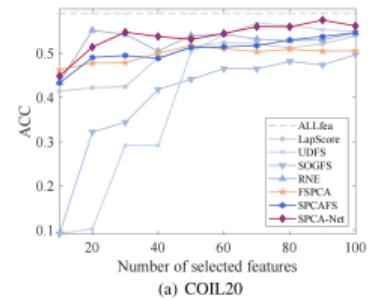
$$\text{Loss} = \frac{1}{2} \|X - \bar{W}\bar{W}^T X\|_F^2 + \lambda\|\bar{W}\|_{2,1} + \mu\|\bar{W}\|_1$$



# Experiments

► Real datasets: Accuracy (ACC) ↑

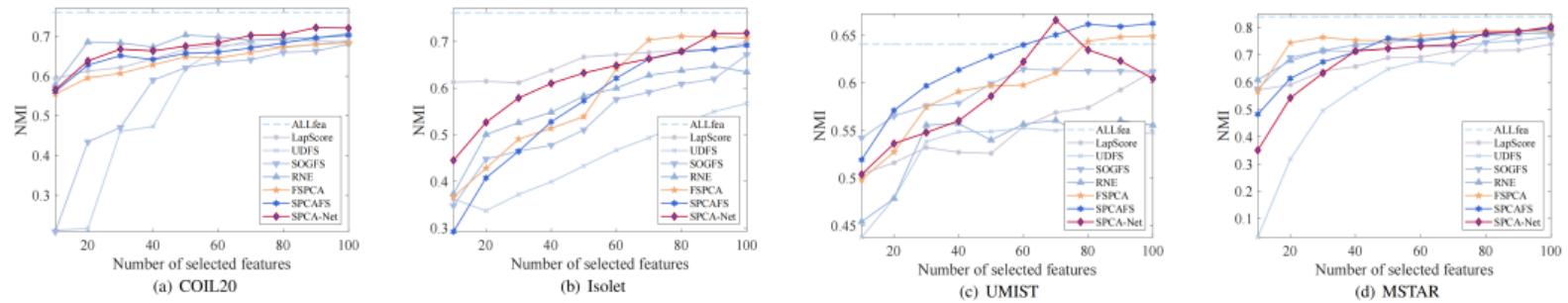
Datasets	ALLfea	LapScore	UDFS	SOGFS	RNE	FSPCA	SPCAFS	SPCA-Net
COIL20	58.97±4.99 (10)	53.91±3.61 (100)	<b>56.70±3.09 (70)</b>	49.66±3.63 (100)	55.16±3.35 (20)	51.71±3.05 (50)	54.63±3.64 (100)	<b>57.46±2.76 (90)</b>
Isolet	59.18±3.19 (10)	52.55±2.83 (100)	41.11±1.71 (100)	48.93±2.69 (100)	47.39±2.91 (80)	<b>54.15±2.69 (70)</b>	52.26±2.81 (100)	<b>58.43±4.31 (100)</b>
UMIST	41.68±2.46 (10)	39.71±3.28 (100)	38.64±1.61 (40)	43.81±2.98 (80)	41.01±2.25 (90)	46.58±2.34 (100)	<b>47.32±3.48 (80)</b>	<b>47.58±4.97 (70)</b>
MSTAR	80.81±8.76 (10)	68.21±4.57 (100)	<b>81.25±7.48 (100)</b>	73.46±5.61 (100)	77.82±6.16 (100)	78.74±5.20 (30)	78.63±8.68 (90)	<b>81.90±6.87 (100)</b>



# Experiments

- Real datasets: Normalized mutual information (NMI) ↑

Datasets	ALLfea	LapScore	UDFS	SOGFS	RNE	FSPCA	SPCAF	SPCA-Net
COIL20	76.04±1.69 (10)	69.01±1.53 (100)	69.12±1.17 (80)	68.03±1.59 (100)	<b>70.76±2.07</b> <b>(100)</b>	68.41±1.60 (100)	70.29±1.31 (100)	<b>72.21±2.68</b> <b>(90)</b>
Isolet	76.09±1.77 (10)	69.86±1.26 (100)	56.73±1.05 (100)	67.15±1.45 (100)	64.74±1.28 (90)	<b>71.12±1.11</b> <b>(80)</b>	69.18±1.33 (100)	<b>71.80±1.59</b> <b>(100)</b>
UMIST	64.07±1.76 (10)	61.23±2.15 (100)	55.43±1.50 (80)	61.46±2.03 (70)	56.08±1.80 (60)	64.94±1.65 (100)	<b>66.26±1.74</b> <b>(100)</b>	<b>66.62±7.52</b> <b>(70)</b>
MSTAR	83.96±3.14 (10)	73.90±1.62 (100)	78.18±3.64 (90)	76.56±1.54 (100)	78.26±2.51 (100)	78.87±2.52 (90)	<b>79.62±2.30</b> <b>(100)</b>	<b>80.67±3.47</b> <b>(90)</b>



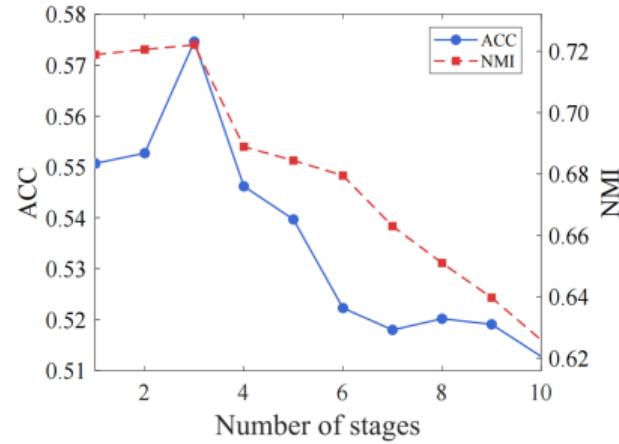
# Experiments

## ► Ablation studies

Datasets	Network	ACC ↑	NMI ↑
COIL20	✗	<b>55.12±2.67</b>	<b>70.44±1.37</b>
	✓	<b>57.46±2.76</b>	<b>72.21±2.68</b>
Isolet	✗	<b>51.84±2.82</b>	<b>67.02±1.43</b>
	✓	<b>58.43±4.31</b>	<b>71.80±1.59</b>
UMIST	✗	<b>40.65±2.29</b>	<b>55.88±1.62</b>
	✓	<b>47.58±4.97</b>	<b>66.62±7.52</b>
MSTAR	✗	<b>80.65±6.47</b>	<b>80.53±2.41</b>
	✓	<b>81.90±6.87</b>	<b>80.67±3.47</b>

Datasets	Dynamic	ACC ↑	NMI ↑
COIL20	✗	<b>56.71±3.83</b>	<b>71.49±3.67</b>
	✓	<b>57.46±2.76</b>	<b>72.21±2.68</b>
Isolet	✗	<b>52.06±3.71</b>	<b>68.91±2.36</b>
	✓	<b>58.43±4.31</b>	<b>71.80±1.59</b>
UMIST	✗	<b>42.63±2.78</b>	<b>60.12±1.69</b>
	✓	<b>47.58±4.97</b>	<b>66.62±7.52</b>
MSTAR	✗	<b>80.74±5.28</b>	<b>80.59±3.67</b>
	✓	<b>81.90±6.87</b>	<b>80.67±3.47</b>

## ► Effect of deep unfolding stages



# Outline

Introduction

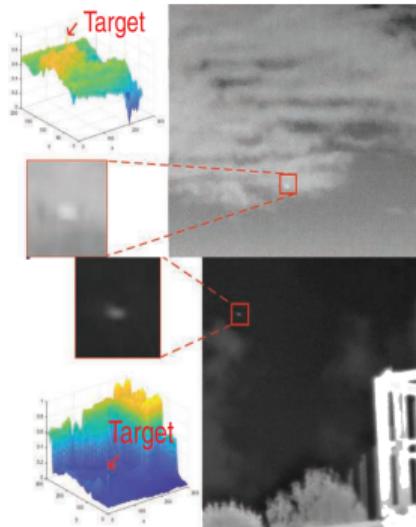
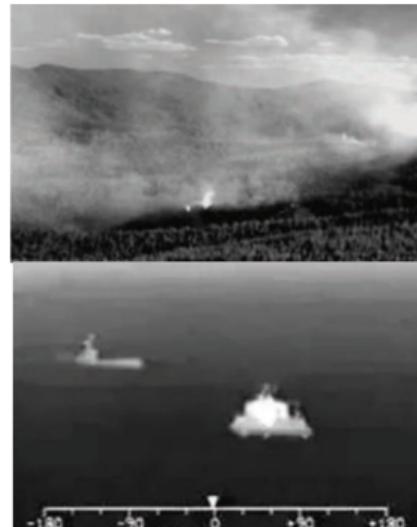
Feature Selection

**Infrared Small Target Detection**

Conclusions and Future Work

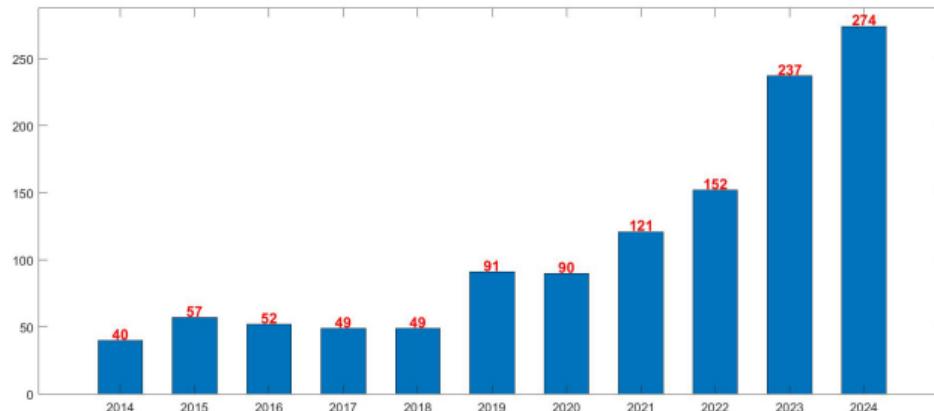
# ISTD

- ▶ 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

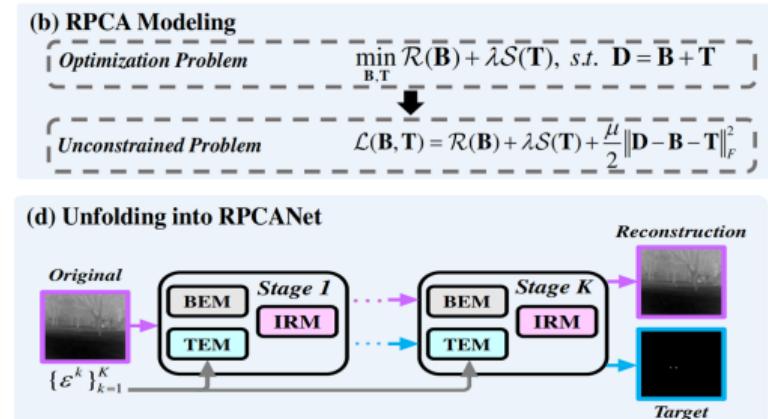
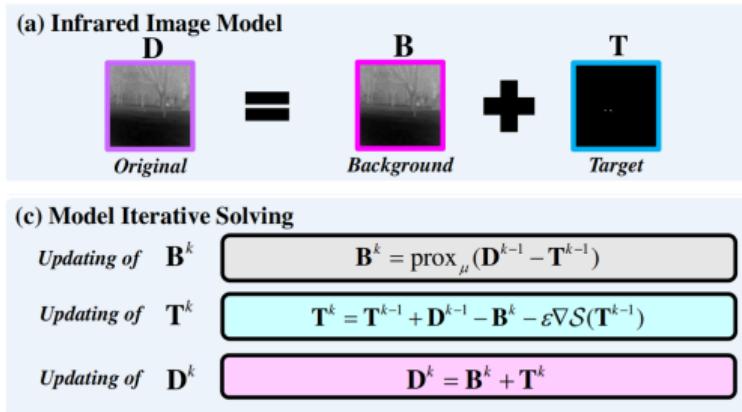


- ▶ Existing methods
  - ▶ Filter-based: Spatial domain, transformed domain
  - ▶ Local information-based: Local contrast, local entropy
  - ▶ Data structure-based: Subspace, dictionary, tensor representation
  - ▶ Deep learning-based

# Motivation

- Wu-Zhang-Li et al, RPCANet: Deep Unfolding RPCA Based Infrared Small Target Detection, [WACV](#), 2024

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



# Model

- ▶ From RPCANet to L-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)

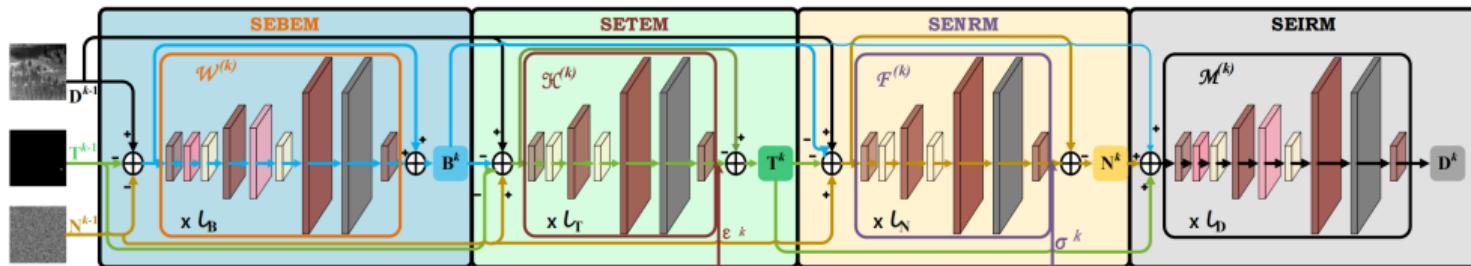
$$B^k = \arg \min_B \mathcal{R}(B) + \frac{\alpha}{2} \|D^{k-1} - B - T^{k-1} - N^{k-1}\|_F^2$$

↓

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

↓

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



## 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  $\gamma = 0.5$

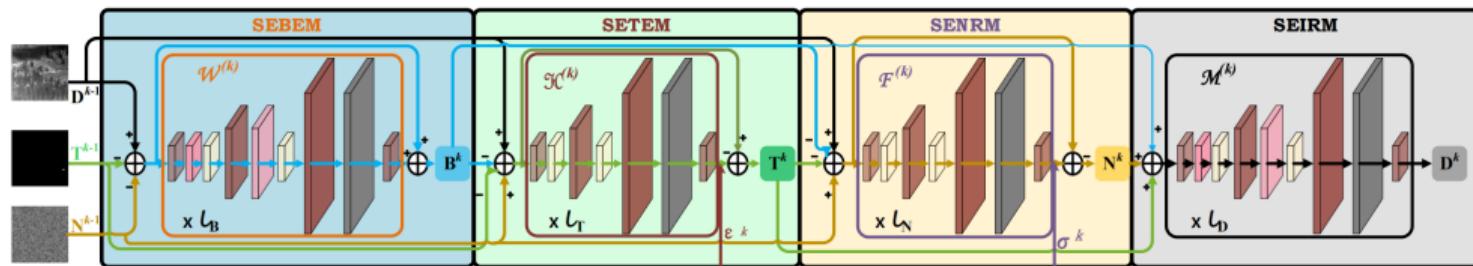
$$T^k = \gamma T^{k-1} + (1 - \gamma)(D^{k-1} - B^k - N^{k-1}) - \varepsilon \nabla S(T^{k-1})$$

↓

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

↓

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



# Update $N$

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

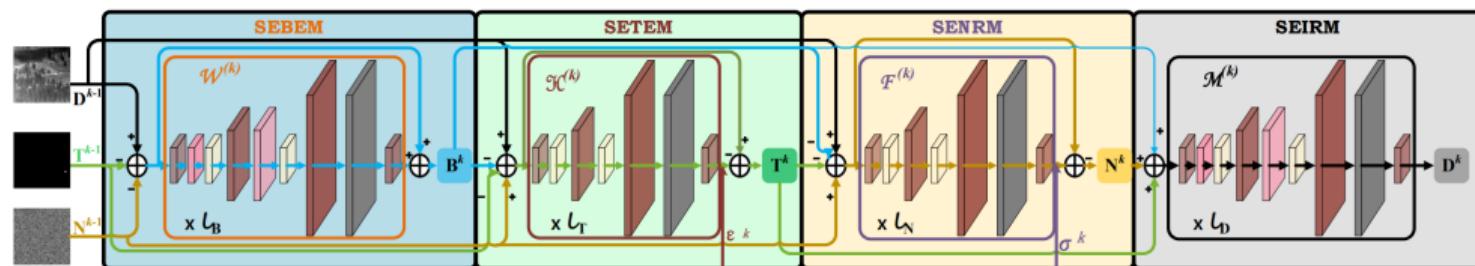
$$N^k = \arg \min_N \mu \mathcal{G}(N) + \frac{\alpha}{2} \|D^{k-1} - B^k - T^k - N\|_F^2$$

↓

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

↓

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



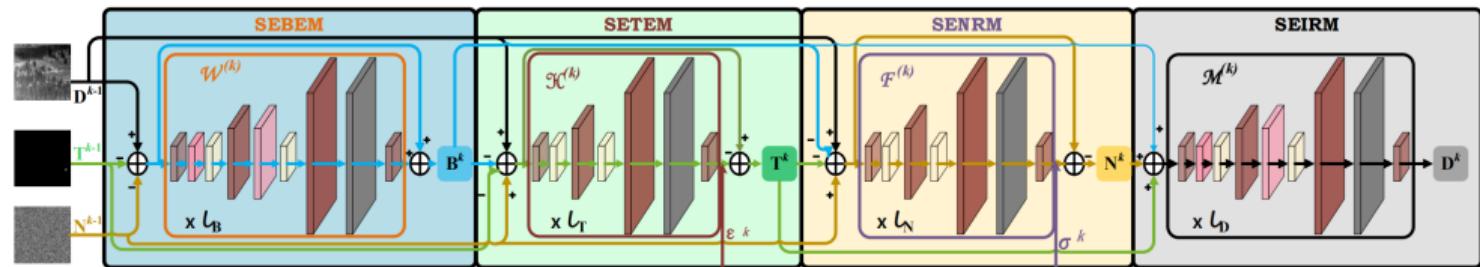
# Update $D$

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

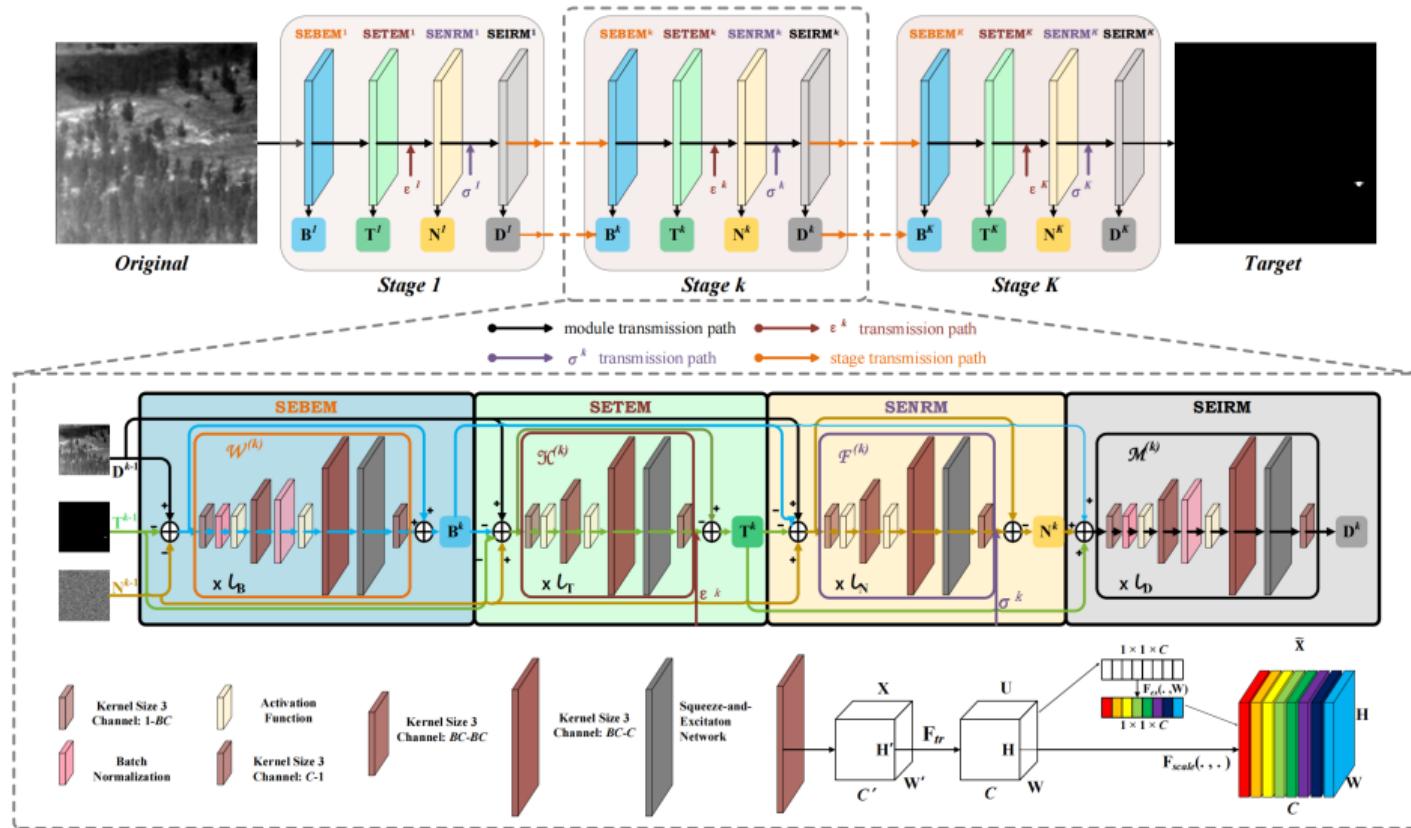
$$D^k = B^k + T^k + N^k$$

↓

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



# Architecture



# Experiments

- ▶ Compared methods
  - ▶ IPI: Gao-Meng-Yang et al, IEEE TIP, 2013
  - ▶ MPCM: Wei-You-Li, PR, 2016
  - ▶ PSTNN: Zhang-Peng, RS, 2019
  - ▶ AGPCNet: Zhang-Li-Cao et al, IEEE TAES, 2023
  - ▶ UIUNet: Wu-Hong-Chanussot, IEEE TIP, 2023
  - ▶ MSHNet: Liu-Liu-Zheng et al, CVPR, 2024
  - ▶ RPCANet: Wu-Zhang-Li et al, WACV, 2024
  - ▶ DRPCANet: Xiong-Zhou-Wu et al, IEEE TGRS, 2025
  - ▶ RPCANet++: Wu-Dai-Zhang et al, arXiv, 2025
- ▶ 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

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

# Experiments

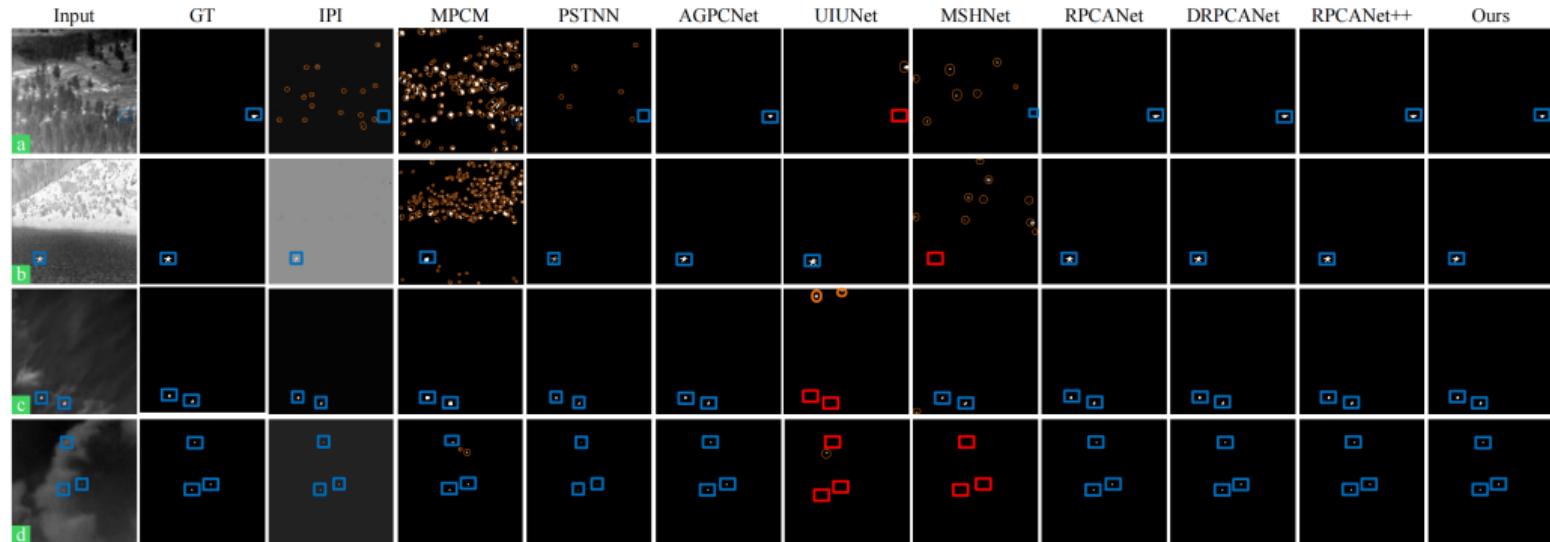
## ► Quantitative results

Methods	Params	NUDT-SIRST				SIRST-Aug				IRSTD-1k				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	21.90	35.97	80.36	<b>2.20</b>	18.67	31.48	78.54	11.11	3.0972/-
MPCM	–	25.96	40.78	78.59	7.91	19.49	33.00	93.58	3.04	14.81	25.93	69.03	6.51	0.0624/-
PSTNN	–	25.46	40.58	78.52	7.95	19.76	33.00	93.40	3.14	14.87	25.89	68.73	6.51	0.2249/-
AGPCNet	12.360M	85.31	92.45	97.90	4.77	72.36	83.83	99.03	35.56	61.00	75.75	89.35	5.34	-/0.0205
UIUNet	50.540M	88.71	94.01	91.43	1.89	71.80	83.59	98.35	28.29	63.06	77.35	<b>93.60</b>	6.57	-/0.0317
MSHNet	4.065M	89.99	93.57	96.07	2.63	71.64	84.16	90.78	23.09	64.50	77.55	91.68	4.46	-/0.0245
RPCANet	0.680M	89.31	94.35	97.14	2.87	72.54	84.08	98.21	34.14	63.21	77.45	88.31	4.39	-/0.0096
DRPCANet	1.169M	<b>93.12</b>	96.02	98.02	1.95	73.93	85.39	98.12	30.45	63.93	78.15	92.09	4.92	-/0.0101
RPCANet++	4.396M	92.46	96.05	98.05	<b>1.44</b>	73.14	84.39	97.36	32.48	64.03	77.26	89.35	<b>4.28</b>	-/0.0063
<b>Ours</b>	<b>0.216M</b>	92.37	<b>96.54</b>	<b>98.41</b>	1.79	<b>74.56</b>	<b>85.43</b>	<b>99.17</b>	29.78	<b>64.68</b>	<b>78.55</b>	89.39	4.66	-/0.0052

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

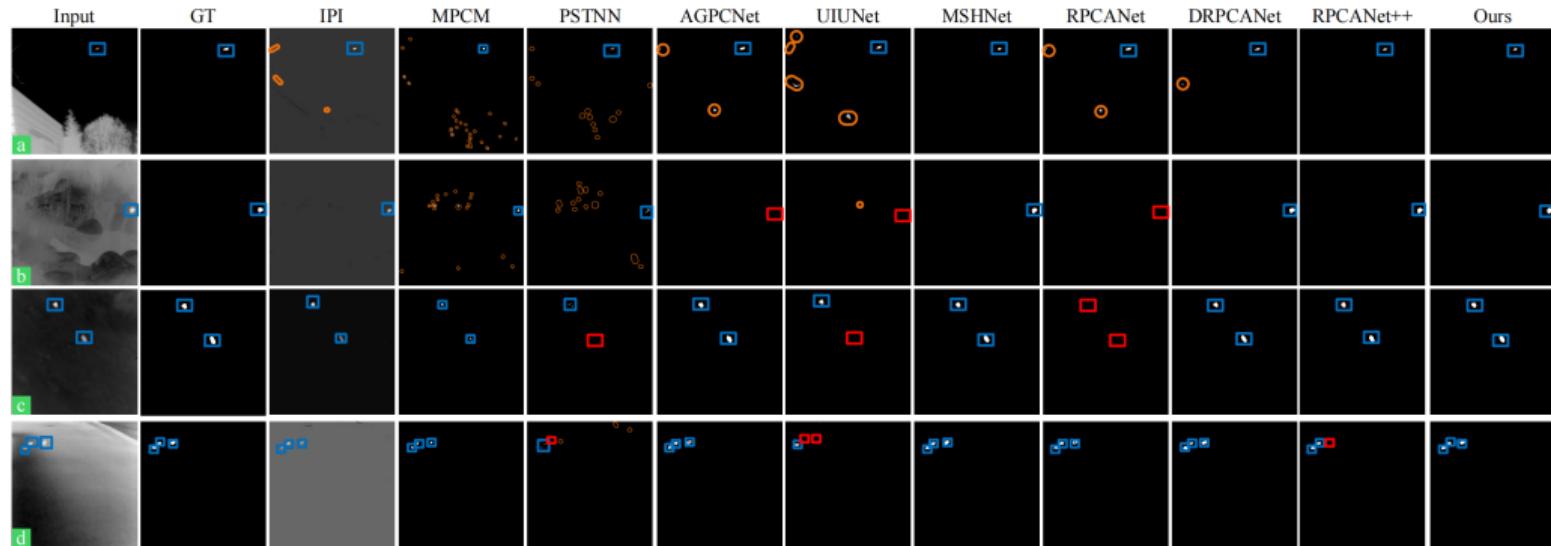
# Experiments

## ► Visualization on NUDT-SIRST



# Experiments

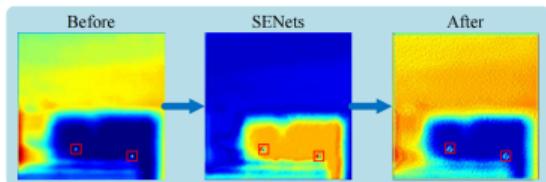
## ► Visualization on IRSTD-1k



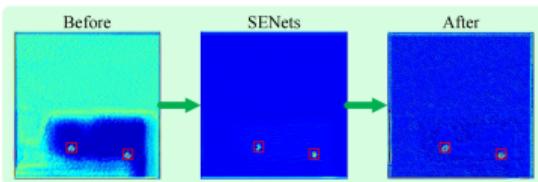
# Experiments

## ► Ablation studies on SENets

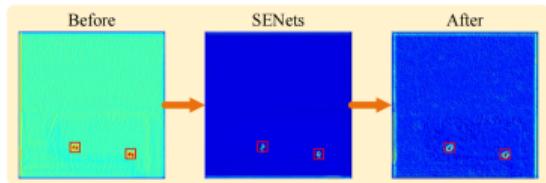
SENNets				NUDT-SIRST				SIRST-Aug				IRSTD-1k			
SEBEM	SETEM	SENRM	SEIRM	mIoU ↑	F <sub>1</sub> ↑	P <sub>d</sub> ↑	F <sub>a</sub> ↓	mIoU ↑	F <sub>1</sub> ↑	P <sub>d</sub> ↑	F <sub>a</sub> ↓	mIoU ↑	F <sub>1</sub> ↑	P <sub>d</sub> ↑	F <sub>a</sub> ↓
55	55	55	55	73.56	78.45	79.36	8.56	60.75	70.28	81.24	43.67	50.26	61.34	70.57	15.95
51	55	55	55	80.57	81.39	86.35	5.68	65.96	75.37	86.99	39.82	55.84	65.26	76.36	11.12
51	51	55	55	88.36	89.45	90.18	4.00	70.17	80.78	93.17	34.78	60.56	72.58	83.70	8.10
51	51	51	55	91.14	96.06	97.18	2.05	73.27	84.45	98.07	27.73	63.58	77.53	88.71	3.95
51	51	51	51	<b>92.37</b>	<b>96.54</b>	<b>98.41</b>	<b>1.79</b>	<b>74.56</b>	<b>85.43</b>	<b>99.17</b>	<b>29.78</b>	<b>64.68</b>	<b>78.55</b>	<b>89.39</b>	<b>4.66</b>



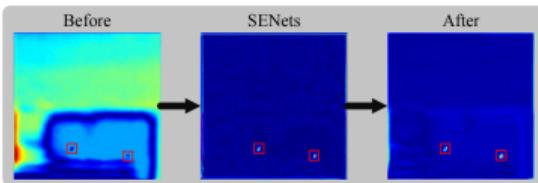
(a) SEBEM



(b) SETEM



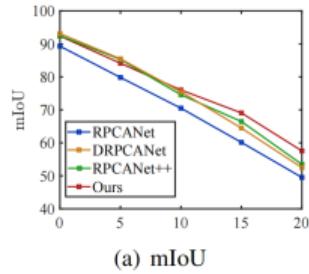
(c) SENRM



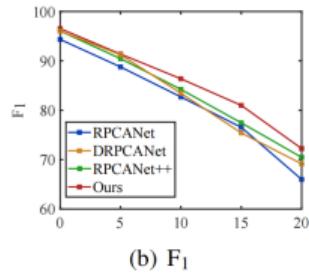
(d) SEIRM

# Experiments

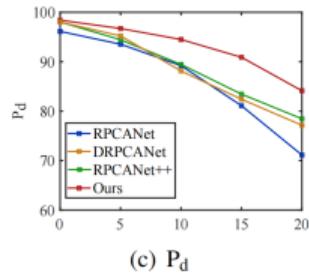
## ► Gaussian noise



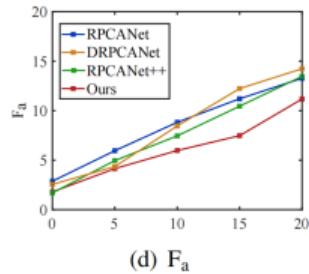
(a) mIoU



(b) F<sub>1</sub>

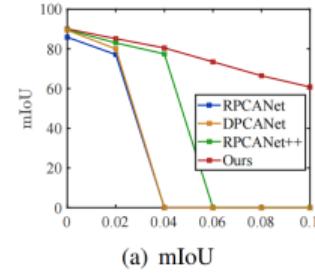


(c) P<sub>d</sub>

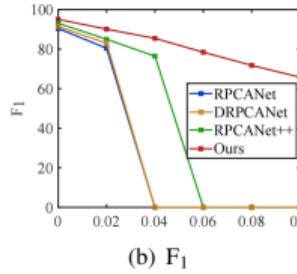


(d) F<sub>a</sub>

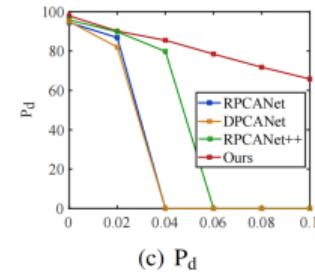
## ► Salt-and-pepper noise



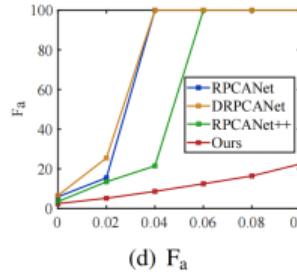
(a) mIoU



(b) F<sub>1</sub>



(c) P<sub>d</sub>



(d) F<sub>a</sub>

# Experiments

## ► Effects of the stage $K$

K	Params	NUDT-SIRST				SIRST-Aug				IRSTD-1k			
		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> ↓
1	0.0360M	91.39	95.51	98.52	2.16	72.83	84.28	98.07	29.14	61.26	75.98	92.12	6.29
2	0.0720M	91.61	95.62	97.88	2.08	72.96	84.37	98.9	34.82	61.59	76.24	86.99	5.12
3	0.1080M	91.53	95.58	97.98	2.00	73.58	84.78	99.17	32.83	62.60	77.00	88.7	5.21
4	0.1439M	90.06	95.06	97.88	1.95	73.81	84.93	98.21	28.51	61.98	76.53	87.33	4.35
5	0.1799M	91.64	95.64	98.31	2.01	74.28	85.24	98.76	28.06	63.45	77.63	88.36	5.45
6	0.216M	92.37	96.04	98.41	1.79	74.56	85.43	99.17	29.78	64.68	78.55	89.39	4.66
7	0.2519M	88.53	93.58	96.77	3.74	72.29	83.92	98.9	27.38	62.29	76.76	87.33	4.82

## ► Effects of the loss weight $\eta$

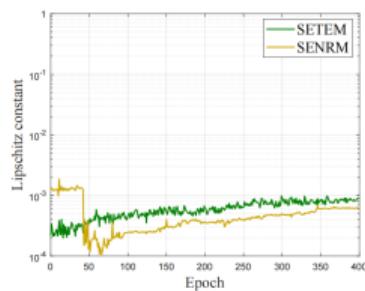
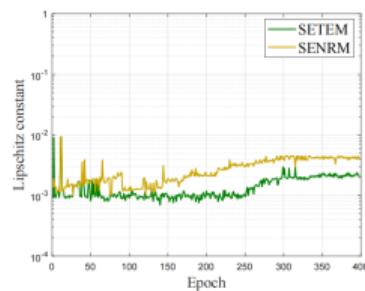
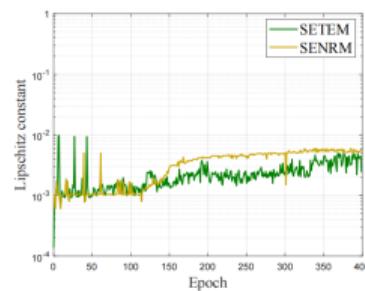
$\eta$	NUDT-SIRST				SIRST-Aug				IRSTD-1k			
	mIoU ↑	F <sub>1</sub> ↑	P <sub>d</sub> ↑	F <sub>a</sub> ↓	mIoU ↑	F <sub>1</sub> ↑	P <sub>d</sub> ↑	F <sub>a</sub> ↓	mIoU ↑	F <sub>1</sub> ↑	P <sub>d</sub> ↑	F <sub>a</sub> ↓
0.005	77.56	82.19	84.25	8.78	65.47	75.58	87.39	35.73	57.45	68.19	78.43	20.54
0.01	92.37	96.54	98.41	1.79	74.56	85.43	99.17	29.78	64.68	78.55	89.39	4.66
0.015	90.36	93.45	94.18	2.90	71.17	82.78	95.17	30.78	61.56	75.58	86.70	6.10
0.2	73.27	78.15	70.35	18.05	60.28	74.45	80.07	40.73	50.36	70.37	80.37	16.78

# Experiments

## ► Effects of channel numbers

Layers	Channels	NUDT-SIRST				SIRST-Aug				IRSTD-1k			
		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> ↓
BC = 4	C = 32	<b>92.37</b>	<b>96.04</b>	<b>98.41</b>	<b>1.79</b>	<b>74.56</b>	<b>85.43</b>	<b>99.17</b>	<b>29.78</b>	<b>64.68</b>	<b>78.55</b>	89.39	<b>4.66</b>
BC = 8	C = 32	91.03	95.58	97.78	2.00	72.71	84.21	98.35	29.45	62.63	77.02	90.75	5.59
BC = 16	C = 32	90.52	94.17	96.54	2.99	72.31	83.93	97.66	27.7	62.58	76.98	88.01	4.76
BC = 4	C = 40	90.06	94.06	97.38	2.34	72.54	83.08	97.21	30.23	62.60	77.00	88.7	5.21
BC = 4	C = 48	89.34	93.34	97.00	2.54	72.1	84.03	97.32	29.33	61.16	75.90	88.7	5.13
BC = 4	C = 56	89.09	92.85	97.31	3.01	70.34	83.78	96.39	31.14	60.98	75.76	91.44	5.99
BC = 4	C = 64	87.93	91.53	95.34	4.45	68.99	81.45	94.55	33.74	58.25	73.61	<b>92.81</b>	5.84

## ► Lipschitz conditions



(a) NUDT-SIRST

(b) SIRST-Aug

(c) IRSTD-1k

# Outline

Introduction

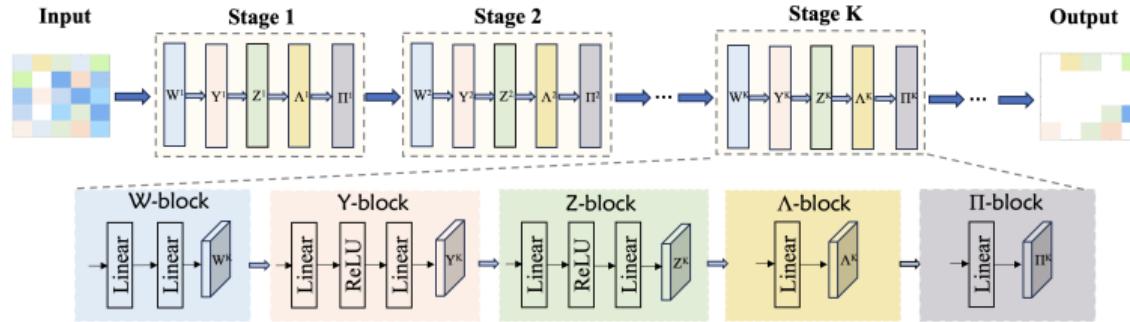
Feature Selection

Infrared Small Target Detection

Conclusions and Future Work

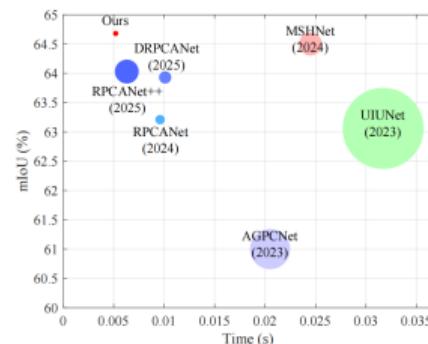
# Conclusions

- ▶ Chen-Xiu, Tuning-Free Structured Sparse PCA via Deep Unfolding Networks, CCC, 2025



- ▶ Liu-Han-Xiu et al, Lightweight Deep Unfolding Networks with Enhanced Robustness for Infrared Small Target Detection, arXiv, 2025

Methods	mIoU ↑	F <sub>1</sub> ↑	P <sub>d</sub> ↑	F <sub>a</sub> ↓
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>Ours</b>	<b>77.20</b>	<b>86.67</b>	<b>95.66</b>	12.08

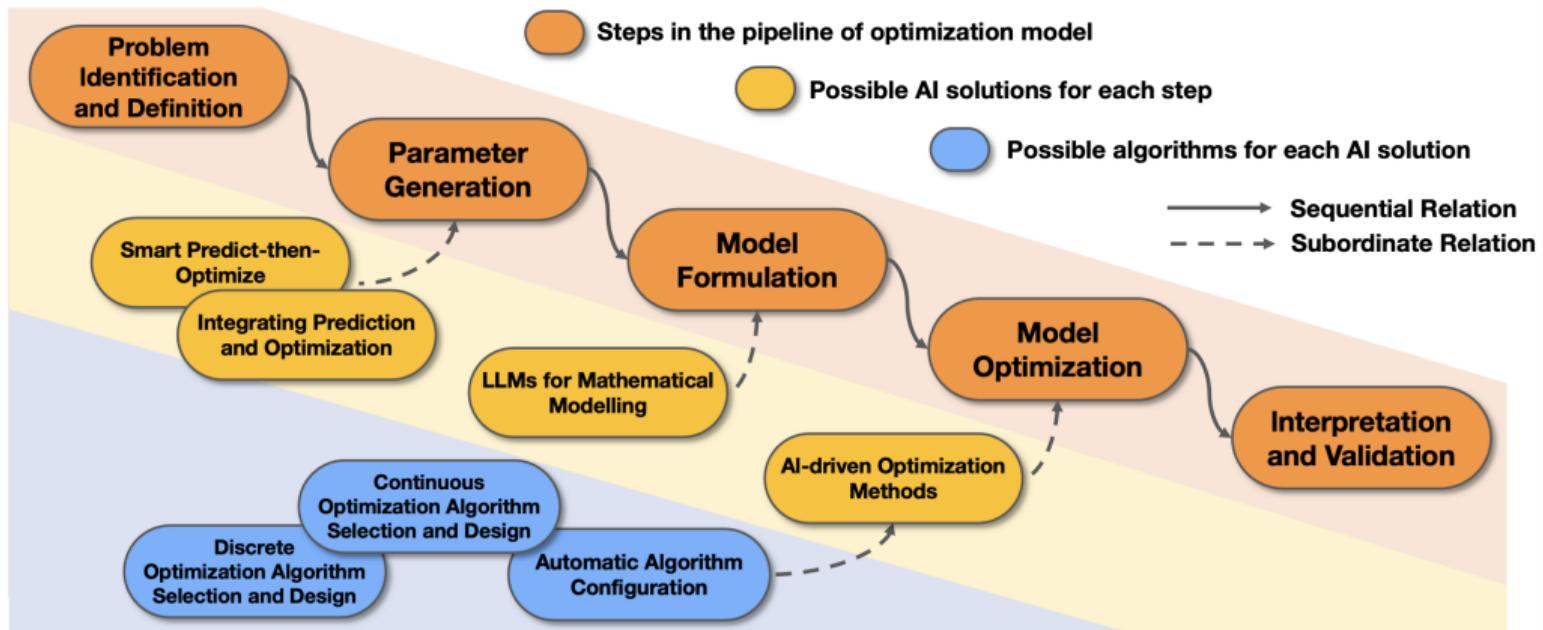


# Conclusions

- ▶ DUN + Generative AI
  - ▶ Chen-Zhang-Li et al, Invertible Diffusion Models for Compressed Sensing, [IEEE TPAMI](#), 2025
  - ▶ Liao-Shen-Li et al, Using Powerful Prior Knowledge of Diffusion Model in Deep Unfolding Networks for Image Compressive Sensing, [CVPR](#), 2025
  - ▶ Chang-Wang-Deng et al, WaterDiffusion: Learning a Prior-involved Unrolling Diffusion for Joint Underwater Saliency Detection and Visual Restoration, [AAAI](#), 2025
  - ▶ Ai-Cai-Zhang et al, Flow-Matching Guided Deep Unfolding for Hyperspectral Image Reconstruction, [arXiv](#), 2025
  - ▶ Tan-Mukherjee-Tang, From Image Denoisers to Regularizing Imaging Inverse Problems: An Overview, [arXiv](#), 2025
- ▶ DUN + Semi-Smooth Newton
  - ▶ Deng-Zhang-Jiang et al, DeepSN-Net: Deep Semi-Smooth Newton Driven Network for Blind Image Restoration, [IEEE TPAMI](#), 2025
  - ▶ Zhang-Deng-Xu et al, Deep Semi-Smooth Newton-Driven Unfolding Network for Multi-Modal Image Super-Resolution, [IEEE TIP](#), 2025

# Future Work

## ► Large Language Models for Optimization

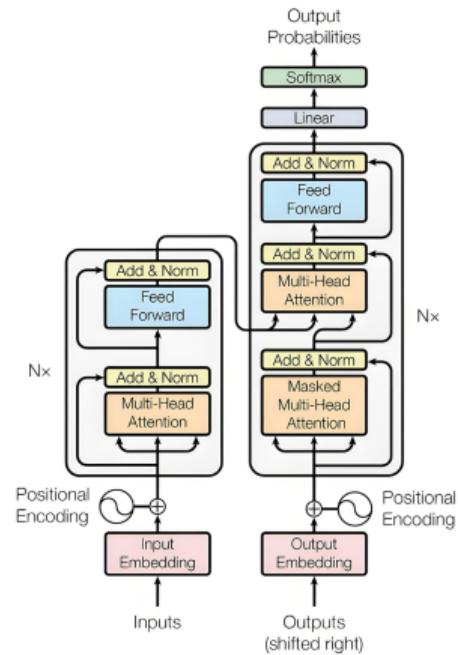
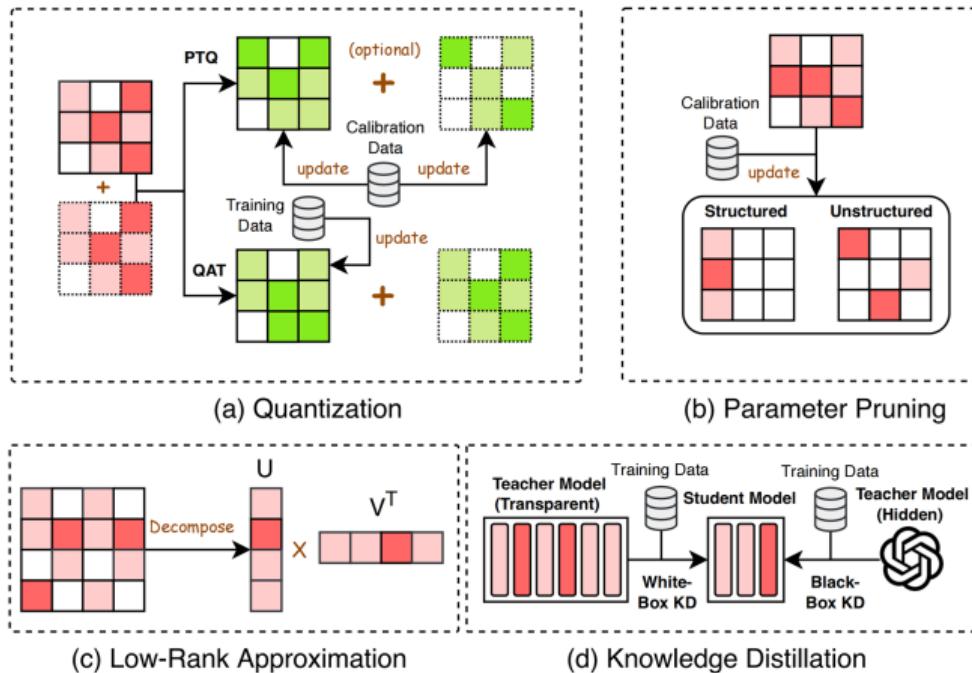


## Future Work

- ▶ Ramamonjison-Yu-Li et al, NL4Opt Competition: Formulating Optimization Problems Based on Their Natural Language Descriptions, [NeurIPS](#), 2022
- ▶ Yang-Wang-Lu et al, Large Language Models as Optimizers, [ICLR](#), 2024
- ▶ AhmadiTeshnizi-Gao-Udell, OptiMUS: Scalable Optimization Modeling with (MI)LP Solvers and Large Language Models, [ICML](#), 2024
- ▶ Gao-Jiang-Cai et al, StrategyLLM: Large Language Models as Strategy Generators, Executors, Optimizers, and Evaluators for Problem Solving, [NeurIPS](#), 2024
- ▶ Romera-Paredes-Barekatain et al, Mathematical Discoveries from Program Search with Large Language Models, [Nature](#), 2024
- ▶ Jiang-Shu-Qian et al, LLMOPT: Learning to Define and Solve General Optimization Problems from Scratch, [ICLR](#), 2025
- ▶ Huang-Tang-Hu et al, ORLM: A Customizable Framework in Training Large Models for Automated Optimization Modeling, [Operations Research](#), 2025
- ▶ Chen-Flores-Mantri et al, OptiChat: Bridging Optimization Models and Practitioners with Large Language Models, [INFORMS Journal on Data Science](#), 2025

# Future Work

## ► Optimization for Large Language Models



## Future Work

- ▶ Frantar-Alistarh, SparseGPT: Massive Language Models Can be Accurately Pruned in One-Shot, [ICML](#), 2023
- ▶ Ma-Fang-Wang, LLM-Pruner: On the Structural Pruning of Large Language Models, [NeurIPS](#), 2023
- ▶ Sun-Liu-Bair et al, A Simple and Effective Pruning Approach for Large Language Models, [ICLR](#), 2024
- ▶ Deng-Jiao-Liu, et al, DRPruning: Efficient Large Language Model Pruning through Distributionally Robust Optimization, [ACL](#), 2025
- ▶ Zhao-Hu-Li, et al, FISTAPruner: Layer-wise Post-training Pruning for Large Language Models, [EMNLP](#), 2025
- ▶ Liu-Liu-Wang et al, ARMOR: High-Performance Semi-Structured Pruning via Adaptive Matrix Factorization, [arXiv](#), 2025

Thank you for your attention!

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