Rethinking Sparse Optimization Through Deep Learning

Xianchao Xiu

Department of Automation

上海大学
Shanghai University

Mathematical Optimization Society, May 16-19, 2025

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

Outline

Sparse Optimization

- Sparse optimization considers

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

$$\Updownarrow$$

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

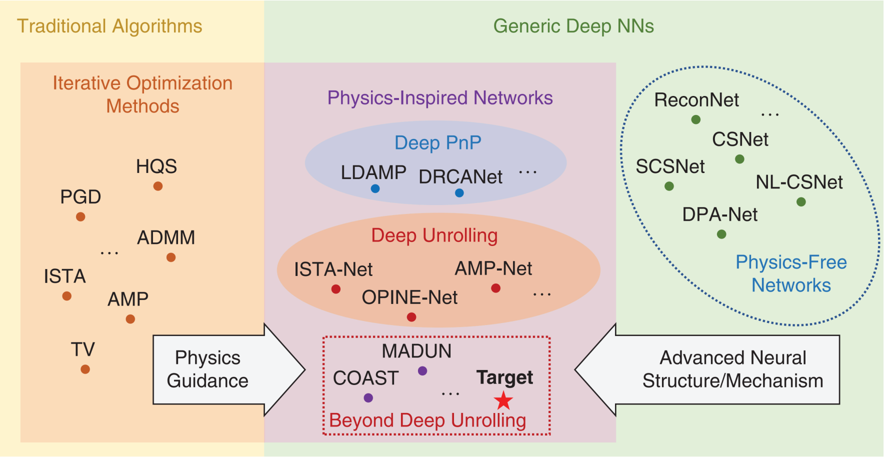
  - x can be extended to matrices and tensors
  - $f(x)$ may be nonsmooth even nonconvex
  - $\|x\|_0$ counts the number of nonzeros
  - λ and s are parameters
- Also compressed sensing, variable selection
- Borad applications: Machine learning, pattern recognition, engineering
- Efficient algorithms: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

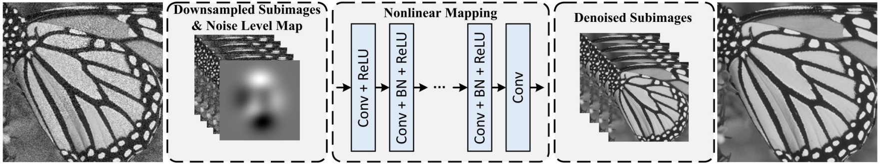
Deep Learning

- From iterative optimization to deep learning



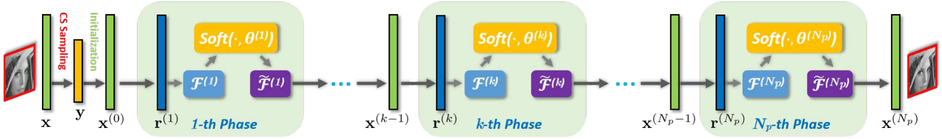
PnP

- 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

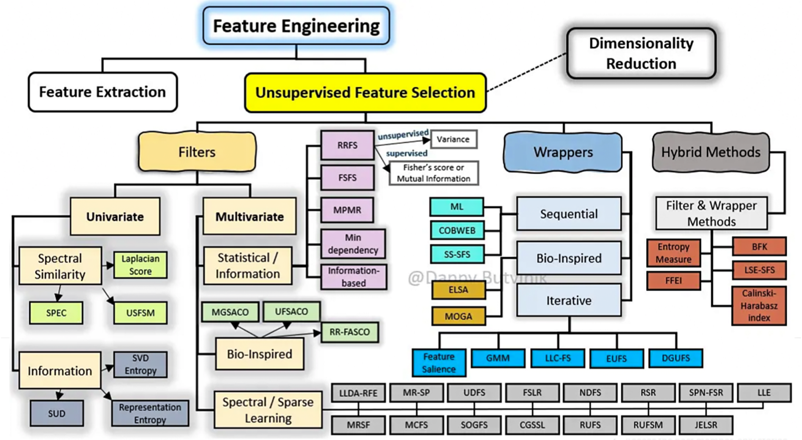
- Gregor-LeCun, Learning Fast Approximations of Sparse Coding, ICML, 2010
- Zhang-Ghanem, ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing, CVPR, 2018
- Yang-Sun-Li et al, ADMM-CSNet: A Deep Learning Approach for Image Compressive Sensing, IEEE TPAMI, 2020
- Chen-Chen-Chen 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, 2025



Outline

UFS

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



PCA

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

$$\begin{aligned} \min_{W\in\mathbb{R}^{d\times p}} \quad & \frac{1}{2}\|X - WW^\top X\|_{\mathrm{F}}^2 \\ \text{s.t.} \quad & W^\top W = I_p \end{aligned}$$

$$\Updownarrow$$

$$\begin{aligned} \min_{W\in\mathbb{R}^{d\times p}} \quad & -\operatorname{Tr}(W^\top XX^\top W) \\ \text{s.t.} \quad & W^\top W = I_p \end{aligned}$$

- Unsupervised feature selection by sparse PCA

$$\begin{aligned} \min_{W\in\mathbb{R}^{d\times p}} \quad & -\operatorname{Tr}(W^\top XX^\top W) \\ \text{s.t.} \quad & W^\top W = I_p,\ \|W\|_{2,0} \leq s \end{aligned}$$

- The i-th feature can be measured by $\|\mathbf{w}^i\|$ since $\mathbf{z}_i = (\mathbf{w}^{1\top},\ \mathbf{w}^{2\top},\ \ldots\ , \mathbf{w}^{d\top})\mathbf{x}_i$
- 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

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

- Li-Sun-Zhang, Unsupervised Feature Selection via Nonnegative Orthogonal Constrained Regularized Minimization, arXiv, 2024

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

- Xiu-Huang-Shang et al, Bi-Sparse Unsupervised Feature Selection, IEEE TIP, 2025

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

Model

- Consider structured sparse PCA

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

- Alternating direction method of multipliers (ADMM)

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

$$\Downarrow$$

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

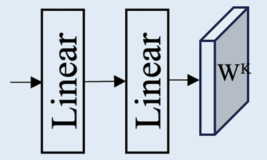
SPCA-Net

- Update W-block

$$\begin{aligned} \min_{W} \quad & f(W) := \frac{1}{2}\|X - WW^\top X\|_{\mathrm{F}}^2 + \frac{\alpha}{2}\|W - Y^k + \Lambda^k/\alpha\|_{\mathrm{F}}^2 + \frac{\beta}{2}\|W - Z^k + \Pi^k/\beta\|_{\mathrm{F}}^2 \\ \text{s.t.} \quad & W^\top W = I \end{aligned}$$

$$\Downarrow$$

$$\begin{aligned} \min_{W} \quad & f(W^k) + \langle \nabla f(W^k), W - W^k \rangle + \frac{1}{2\eta}\|W - W^k\|_{\mathrm{F}}^2 \\ \text{s.t.} \quad & W^\top W = I \end{aligned}$$



$$\Downarrow$$

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

$$\Downarrow$$

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

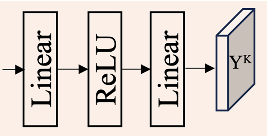
SPCA-Net

- Update Y-block

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

$$\Downarrow$$

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



$$\Downarrow$$

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

$$\Downarrow$$

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

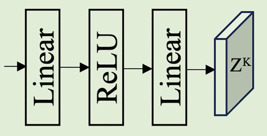
SPCA-Net

- Update Z-block

$$\min_{Z} \ \mu\|Z\|_1 + \frac{\beta}{2}\|X^{k+1} - Z + \Pi^k/\beta\|_{\mathrm{F}}^2$$

$$\Downarrow$$

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



$$\Downarrow$$

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

$$\Downarrow$$

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

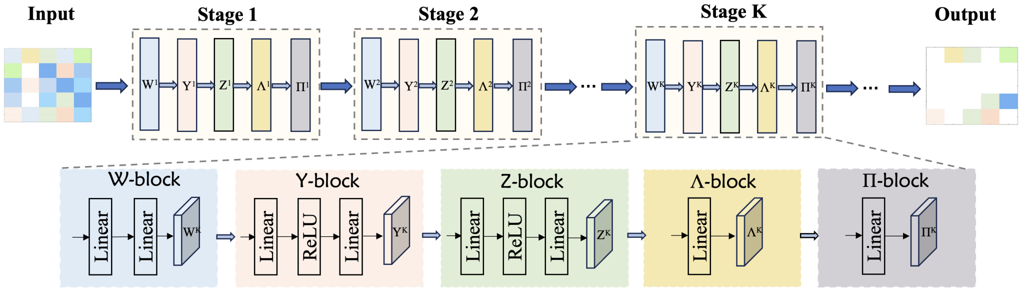
SPCA-Net

- Input: $X,\ \lambda, \mu, \alpha, \beta$
- Initialize: $(W^0,\ Y^0,\ Z^0,\ \Lambda^0,\ \Pi^0)$
- While $k = 1, \ldots, K$ do
  - Update W^{k+1} by
  $$W^{k+1} = \mathbf{LargNet}(U, V^\top)$$
  - Update Y^{k+1} by
  $$Y^{k+1} = \mathbf{GSoftNet}(X^{k+1} + \Lambda^k/\alpha, \lambda/\alpha)$$
  - Update Z^{k+1} by
  $$Z^{k+1} = \mathbf{SoftNet}(X^{k+1} + \Pi^k/\beta, \mu/\beta)$$
  - Update Λ^{k+1}, Π^{k+1} by
  $$\Lambda^{k+1} = \mathbf{Linear}(W^{k+1}, Y^{k+1}, \Lambda^k, \alpha),\ \ \Pi^{k+1} = \mathbf{Linear}(W^{k+1}, Z^{k+1}, \Pi^k, \beta)$$
- Output: Trained W

Architecture

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

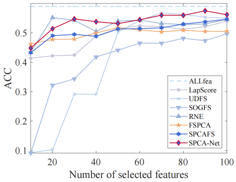
$$\textsc{Loss} = \frac{1}{2}\|X - \bar{W}\bar{W}^\top X\|_{\mathrm{F}}^2 + \lambda\|\bar{W}\|_{2,1} + \mu\|\bar{W}\|_1$$



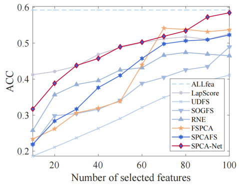
Experiments

- Real datasets: Accuracy (ACC) ↑

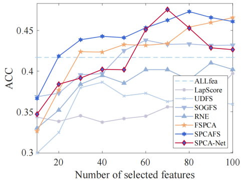
Datasets	ALLfea	LapScore	UDFS	SOGFS	RNE	FSPCA	SPCAFS	SPCA-Net
COIL20	58.97±4.99 (10)	53.91±3.61 (100)	**56.70±3.09 (70)**	49.66±3.63 (100)	55.16±3.35 (20)	51.71±3.05 (50)	54.63±3.64 (100)	**57.46±2.76 (90)**
Isolet	59.18±3.19 (10)	52.55±2.83 (100)	41.11±1.71 (100)	48.93±2.69 (100)	47.39±2.91 (80)	**54.15±2.69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



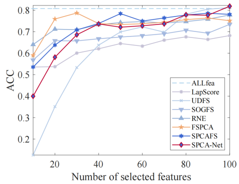
(a) COIL20



(b) Isolet



(c) UMIST

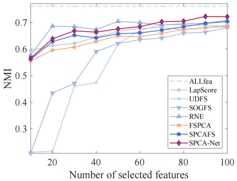


(d) MSTAR

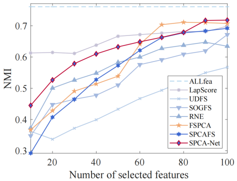
Experiments

- Real datasets: Normalized mutual information (NMI) ↑

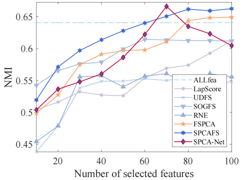
Datasets	ALLfea	LapScore	UDFS	SOGFS	RNE	FSPCA	SPCAFS	SPCA-Net
COIL20	76.04±1.69 (10)	69.01±1.53 (100)	69.12±1.17 (80)	68.03±1.59 (100)	**70.76±2.07 (100)**	68.41±1.60 (100)	70.29±1.31 (100)	**72.21±2.68 (90)**
Isolet	76.09±1.77 (10)	69.86±1.26 (100)	56.73±1.05 (100)	67.15±1.45 (100)	64.74±1.28 (90)	**71.12±1.11 (80)**	69.18±1.33 (100)	**71.80±1.59 (100)**
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)	**66.26±1.74 (100)**	**66.62±7.52 (70)**
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)	**79.62±2.30 (100)**	**80.67±3.47 (90)**



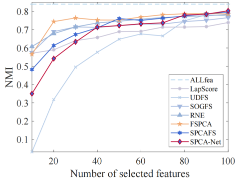
(a) COIL20



(b) Isolet



(c) UMIST



(d) MSTAR

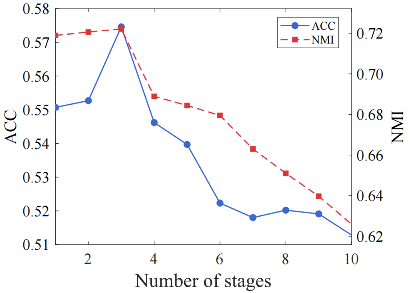
Experiments

- Ablation studies

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

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

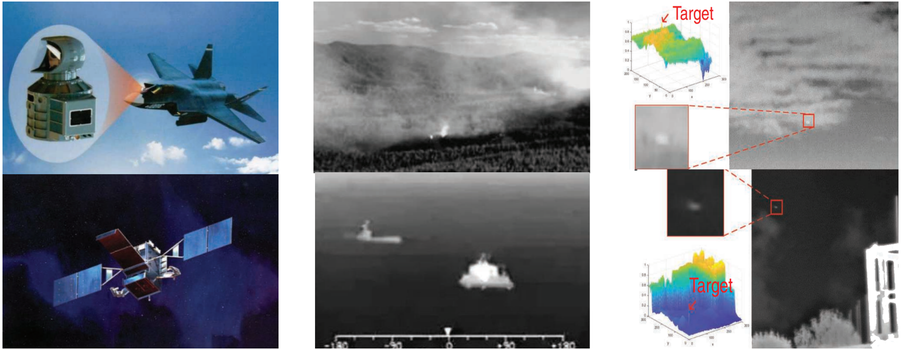
- Effect of deep unfolding stages



Outline

ISTD

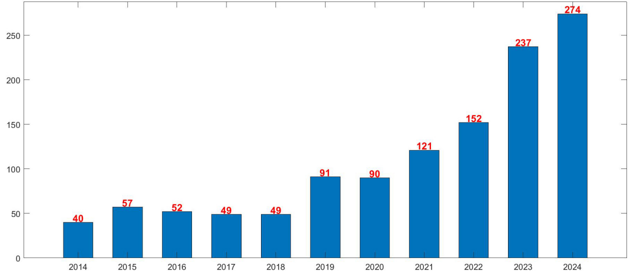
- Infrared small target detection (ISTD)



- Difficulties: Small size, low SNR, weak contrast
- Advantages: Strong concealment, good anti-interference

ISTD

- Trends of "infrared small target detection" on Google Scholar

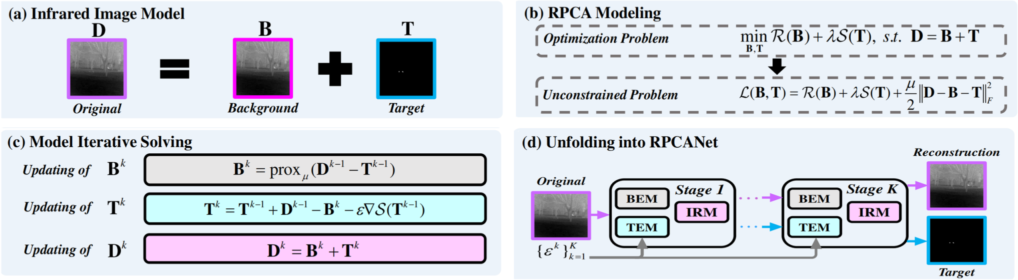


- Existing methods
  - Filter-based: Spatial domain, transformed domain
  - Local information-based: Local contrast, local entropy
  - Data structure-based: Subspace, dictionary, tensor representation
  - Deep learning-based

Motivation

- Wu-Zhang-Li et al, RPCANet: Deep Unfolding RPCA Based Infrared Small Target Detection, WACV, 2024

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



Model

- From RPCANet to L-RPCANet

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

$$\Downarrow$$

$$\begin{aligned} \min_{B,T,N} \quad & \|B\|_* + \lambda \|T\|_1 + \mu \|N\|_F^2 \\ \text{s.t.} \quad & D = B + T + N \end{aligned}$$

$$\Downarrow$$

$$\begin{aligned} \min_{B,T,N} \quad & \mathcal{R}(B) + \lambda \mathcal{S}(T) + \mu \mathcal{G}(N) \\ \text{s.t.} \quad & D = B + T + N \end{aligned}$$

- Unconstrained version

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

Update B

- Bcakgound estimation module + squeeze-and-excitation network (SEBEM)

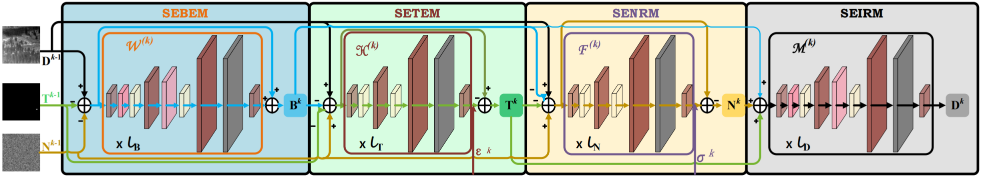
$$B^k = \arg\min_B \ \mathcal{R}(B) + \frac{\alpha}{2}\|D^{k-1} - B - T^{k-1} - N^{k-1}\|_F^2$$

$$\Downarrow$$

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

$$\Downarrow$$

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



Update T

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

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

$$\Downarrow$$

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

$$\Downarrow$$

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

$$\Downarrow$$

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

Update T

- Set $\gamma = 0.5$

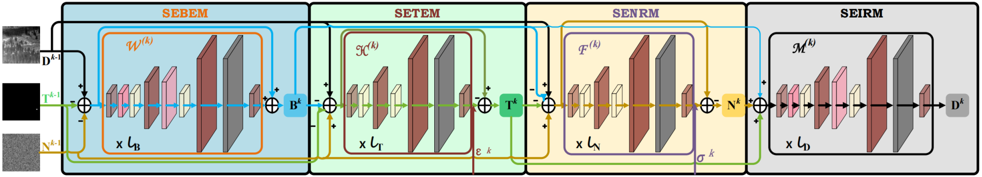
$$T^k = \gamma T^{k-1} + (1-\gamma)(D^{k-1} - B^k - N^{k-1}) - \varepsilon \nabla \mathcal{S}(T^{k-1})$$

$$\Downarrow$$

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

$$\Downarrow$$

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



Update N

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

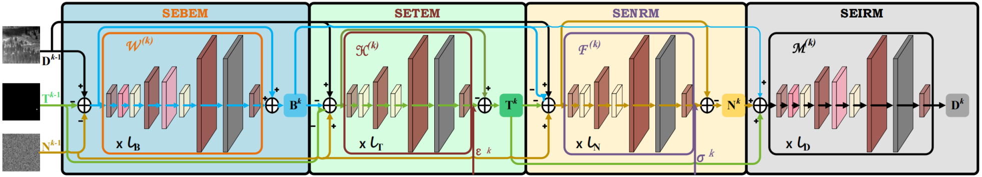
$$N^k = \arg\min_N \ \mu\mathcal{G}(N) + \frac{\alpha}{2}\|D^{k-1} - B^k - T^k - N\|_{\mathrm{F}}^2$$

$$\Downarrow$$

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

$$\Downarrow$$

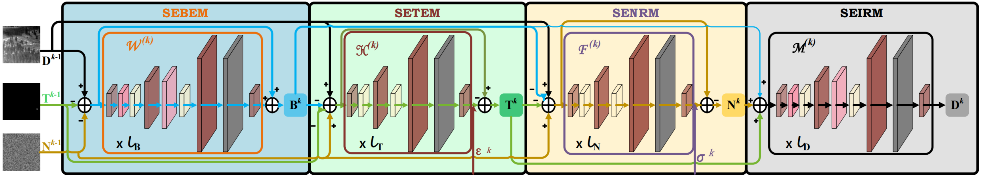
$$N^k = N^{k-1} + D^{k-1} - B^k - T^k - \sigma^k\mathcal{F}^k(N^{k-1} + D^{k-1} - B^k - T^k)$$



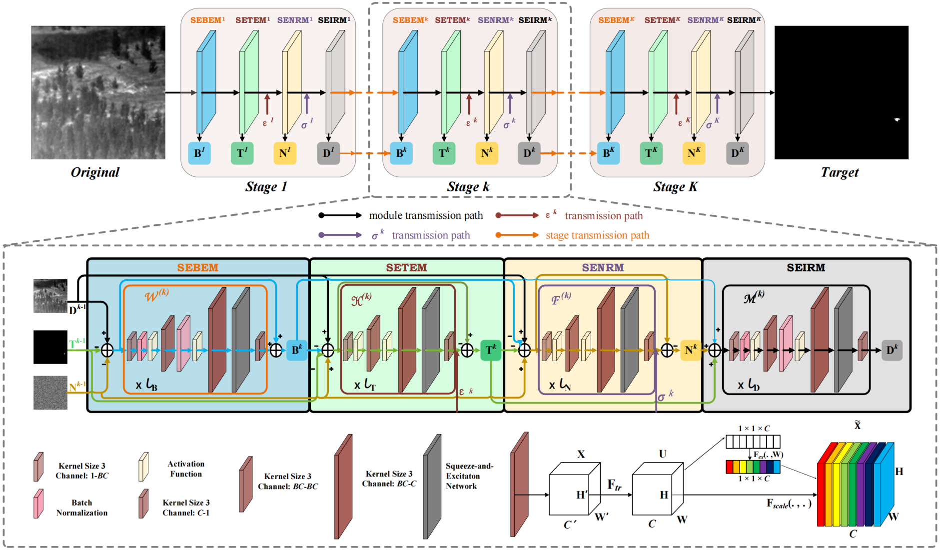
Update D

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

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



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-Pu et al, IEEE TAES, 2023
  - UIUNet: Wu-Hong-Chanussot, IEEE TIP, 2023
  - MSHNet: Liu-Liu-Zheng et al, CVPR, 2024
  - RPCANet: Wu-Zhang-Li-Huang et alg, 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 ($\mathrm{F}_1 \uparrow$), Probability of detection ($\mathrm{P}_\mathrm{d} \uparrow$)
  - False alarm rate ($\mathrm{F}_\mathrm{a} \downarrow$)
- Loss function

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

Experiments

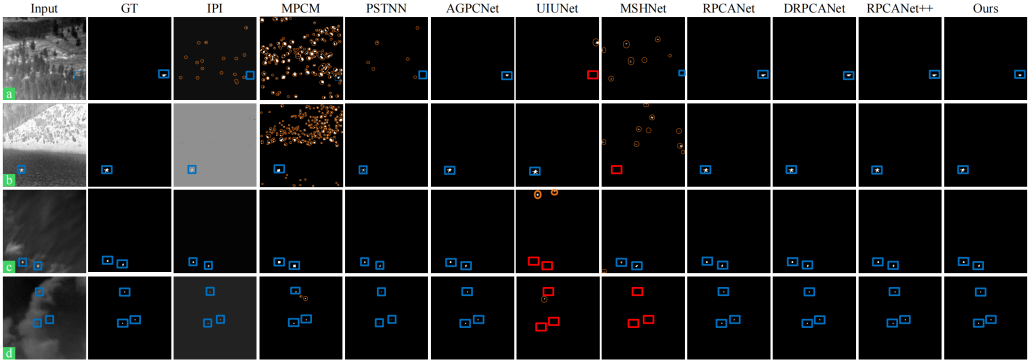
Quantitative results

Methods	Params	NUDT-SIRST				SIRST-Aug				IRSTD-1k				Time (s)
		mIoU ↑	F_1 ↑	P_d ↑	F_a ↓	mIoU ↑	F_1 ↑	P_d ↑	F_a ↓	mIoU ↑	F_1 ↑	P_d ↑	F_a ↓	CPU/GPU
IPI	–	34.83	51.49	92.58	7.14	21.90	35.97	80.36	**2.20**	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	**93.60**	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	**93.12**	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	**1.44**	73.14	84.39	97.36	32.48	64.03	77.26	89.35	**4.28**	-/0.0063
Ours	**0.216M**	92.37	**96.54**	**98.41**	1.79	**74.56**	**85.43**	**99.17**	29.78	**64.68**	**78.55**	89.39	4.66	-/0.0052

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

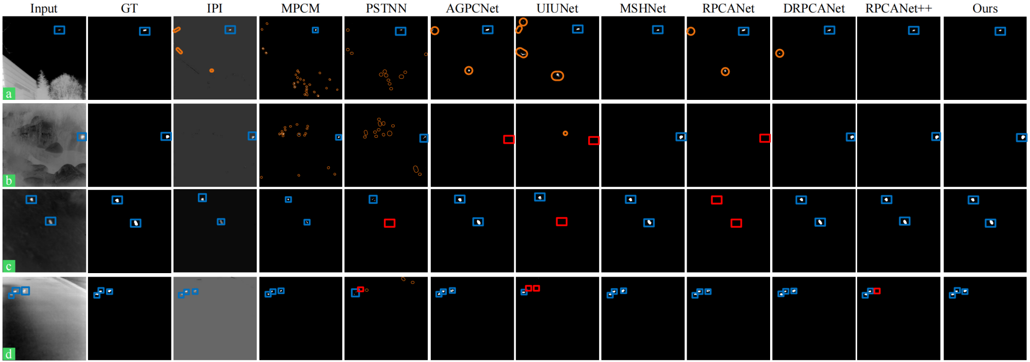
Experiments

- Visualization on NUDT-SIRST



Experiments

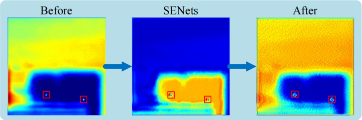
- Visualization on IRSTD-1k



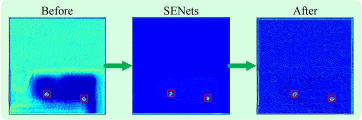
Experiments

▶ Ablation studies on SENets

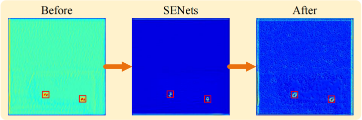
SENets				NUDT-SIRST				SIRST-Aug				IRSTD-1k			
SEBEM	SETEM	SENRM	SEIRM	mIoU ↑	F_1 ↑	P_d ↑	F_a ↓	mIoU ↑	F_1 ↑	P_d ↑	F_a ↓	mIoU ↑	F_1 ↑	P_d ↑	F_a ↓
55	55	55	55	73.56	78.45	79.36	8.56	60.75	70.28	81.24	43.67	50.26	61.34	70.57	15.95
51	55	55	55	80.57	81.39	86.35	5.68	65.96	75.37	86.99	39.82	55.84	65.26	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	**92.37**	**96.54**	**98.41**	**1.79**	**74.56**	**85.43**	**99.17**	**29.78**	**64.68**	**78.55**	**89.39**	**4.66**



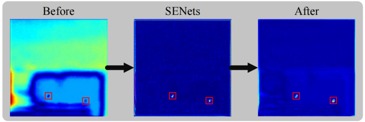
(a) SEBEM



(b) SETEM



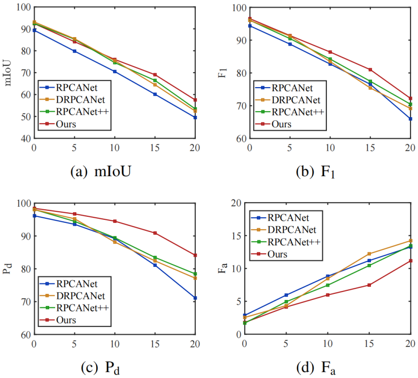
(c) SENRM



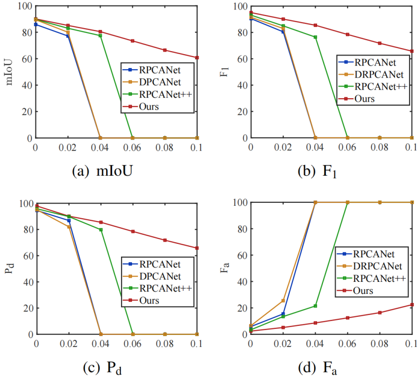
(d) SEIRM

Experiments

- Gaussian noise



- Salt-and-pepper noise



Experiments

- Effects of the stage K

K	Params	NUDT-SIRST				SIRST-Aug				IRSTD-1k			
		mIoU ↑	F_1 ↑	P_d ↑	F_a ↓	mIoU ↑	F_1 ↑	P_d ↑	F_a ↓	mIoU ↑	F_1 ↑	P_d ↑	F_a ↓
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 η

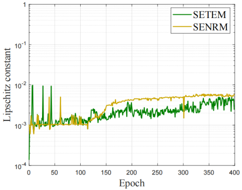
η	NUDT-SIRST				SIRST-Aug				IRSTD-1k			
	mIoU ↑	F_1 ↑	P_d ↑	F_a ↓	mIoU ↑	F_1 ↑	P_d ↑	F_a ↓	mIoU ↑	F_1 ↑	P_d ↑	F_a ↓
0.005	77.56	82.19	84.25	8.78	65.47	75.58	87.39	35.73	57.45	68.19	78.43	20.54
0.01	**92.37**	**96.54**	**98.41**	**1.79**	**74.56**	**85.43**	**99.17**	**29.78**	**64.68**	**78.55**	**89.39**	**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

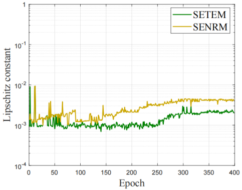
- Effects of channel numbers

Layers	Channels	NUDT-SIRST				SIRST-Aug				IRSTD-1k			
		mIoU ↑	F_1 ↑	P_d ↑	F_a ↓	mIoU ↑	F_1 ↑	P_d ↑	F_a ↓	mIoU ↑	F_1 ↑	P_d ↑	F_a ↓
$BC = 4$	$C = 32$	**92.37**	**96.04**	**98.41**	**1.79**	**74.56**	**85.43**	**99.17**	**29.78**	**64.68**	**78.55**	89.39	**4.66**
$BC = 8$	$C = 32$	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	**92.81**	5.84

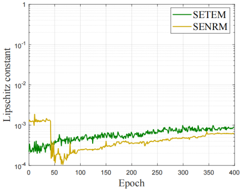
- Lipschitz conditions



(a) NUDT-SIRST



(b) SIRST-Aug

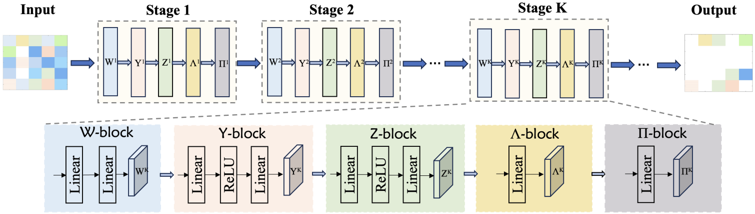


(c) IRSTD-1k

Outline

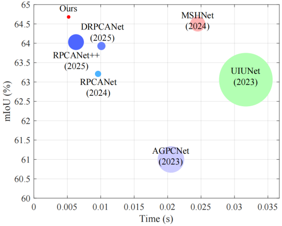
Conclusions

- 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_1 ↑	P_d ↑	F_a ↓
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	**5.82**
Ours	**77.20**	**86.67**	**95.66**	12.08

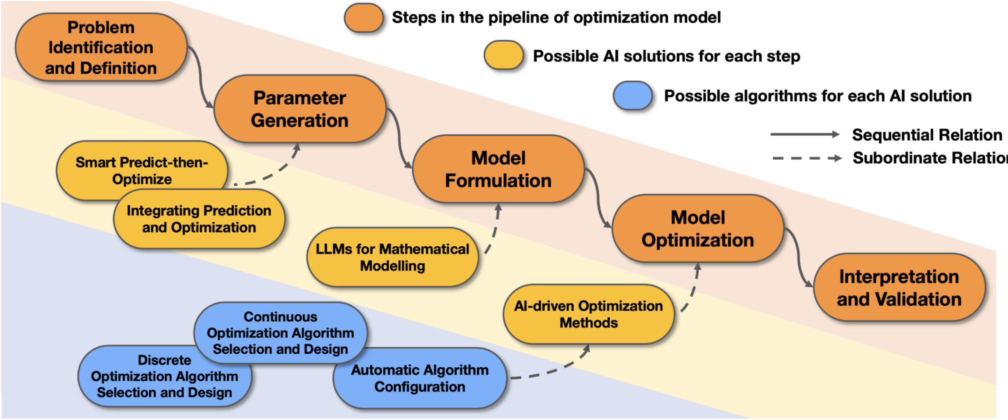


Conclusions

- 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

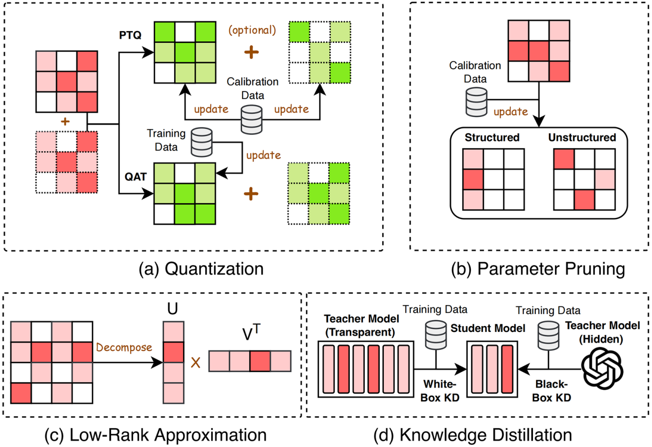


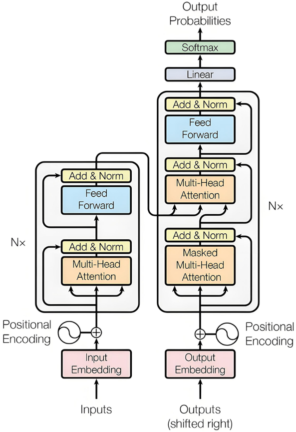
Future Work

- 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