Data-Driven Optimization: Theory, Algorithms, and Applications

Xianchao Xiu

Department of Automation

上海大学
Shanghai University

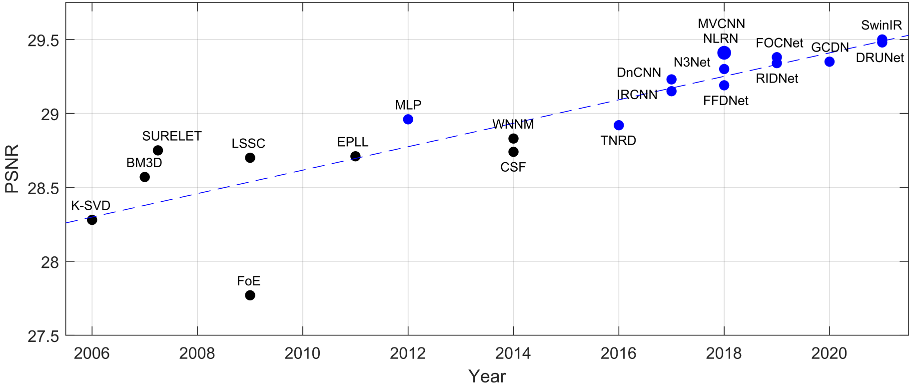
Mathematical Optimization Society, May 16-19, 2025

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

Outline

Why

- Elad-Kawar-Vaksman, Image Denoising: The Deep Learning Revolution and Beyond, SIIMS, 2023



How

- Special Issue: Physics-Driven Machine Learning for Computational Imaging
  - Zha-Wen-Yuan et al, Learning Nonlocal Sparse and Low-Rank Models for Image Compressive Sensing: Nonlocal Sparse and Low-Rank Modeling, IEEE SPM, 2023
  - Dong-Valzania-Maillard et al, Phase Retrieval: From Computational Imaging to Machine Learning, IEEE SPM, 2023
  - Zhang-Chen-Xiong et al, Physics-Inspired Compressive Sensing: Beyond Deep Unrolling, IEEE SPM, 2023
  - Dong-Wu-Li et al, Bayesian Deep Learning for Image Reconstruction: From Structured Sparsity to Uncertainty Estimation, IEEE SPM, 2023
  - Kamilov-Bouman-Buzzard et al, Plug-and-Play Methods for Integrating Physical and Learned Models in Computational Imaging: Theory, Algorithms, and Applications, IEEE SPM, 2023
  - Chen-Davies-Ehrhardt et al, Imaging With Equivariant Deep Learning: From Unrolled Network Design to Fully Unsupervised Learning, IEEE SPM, 2023
  - Zhao-Ye-Bresler, Generative Models for Inverse Imaging Problems: From Mathematical Foundations to Physics-Driven Applications, IEEE SPM, 2023
  - Mukherjee-Hauptmann-Öktem et al, Learned Reconstruction Methods With Convergence Guarantees, IEEE SPM, 2023
- `https://github.com/xianchaoxiu/Deep-Learning-for-Optimization`

How

- Iterative shrinkage thresholding algorithm (ISTA)

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|Ax - b\|^2 + \lambda\|x\|_1$$

  - Step 1: Compute gradient

$$r^{k+1} = x^k - \rho A^\top (Ax^k - b)$$

  - Step 2: Compute shrinkage

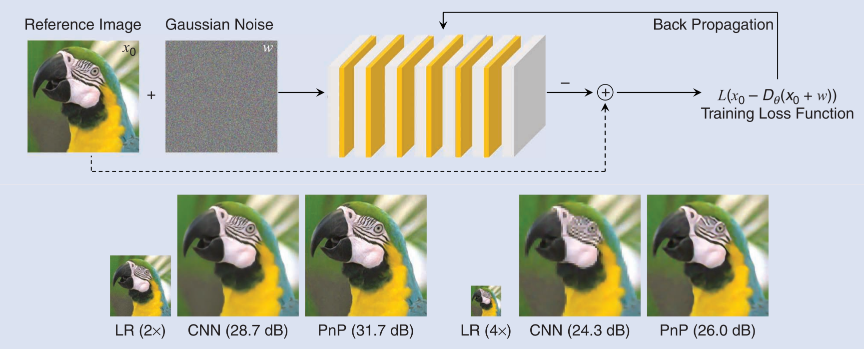
$$\begin{aligned} x^{k+1} &= \arg\min_x \frac{1}{2}\|x - r^{k+1}\|^2 + \lambda\|x\|_1 \\ &= \mathtt{prox}_{\lambda\|\cdot\|_1}(x^k - \rho A^\top (Ax^k - b)) \end{aligned}$$

- Two important issues need to be considered
  - How to design regularizations $\Rightarrow$ Deep PnP/Deep Unrolling
  - How to choose parameters $\Rightarrow$ Deep Unrolling

Deep PnP

- Venkatakrishnan-Bouman-Wohlberg, Plug-and-Play Priors for Model Based Reconstruction, IEEE GlobalSIP, 2013

$$\min_x \frac{1}{2}\|Ax-b\|^2+\lambda\|x\|_1 \quad\Rightarrow\quad \min_x \frac{1}{2}\|Ax-b\|^2+\lambda\phi(x)$$



Related Works

- Ryu-Liu-Wang et al, Plug-and-Play Methods Provably Converge with Properly Trained Denoisers, ICML, 2019
- Wei-Rivero-Liang et al, Tuning-free Plug-and-Play Proximal Algorithm for Inverse Imaging Problems, ICML, 2020
- Sun-Wu-Xu et al, Scalable Plug-and-Play ADMM With Convergence Guarantees, IEEE TCI, 2021
- Cohen-Elad-Milanfar, Regularization by Denoising via Fixed-Point Projection (RED-PRO), SIIMS, 2021
- Huraul-Leclaire-Papadakis, Proximal Denoiser for Convergent Plug-and-Play Optimization with Nonconvex Regularization, ICML, 2022
- Tan-Mukherjee-Tang et al, Provably Convergent Plug-and-Play Quasi-Newton Methods, SIIMS, 2024
- Wang-Zhang-Li et al, DMPlug: A Plug-in Method for Solving Inverse Problems with Diffusion Models, NeurIPS, 2024
- Zheng-Chu-Zhang et al,InverseBench: Benchmarking Plug-and-Play Diffusion Priors for Inverse Problems in Physical Sciences, ICLR, 2025

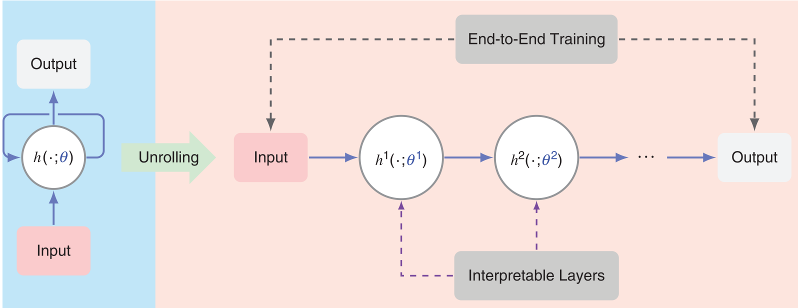
Deep Unrolling

- Gregor-LeCun, Learning Fast Approximations of Sparse Coding, ICML, 2010

$$x^{k+1} = \mathtt{prox}_{\lambda\|\cdot\|_1}(x^k - \rho A^\top(Ax^k - b))$$

$$\Downarrow$$

$$x^{k+1} = h_{\theta^k}(x^k - \rho A^\top(Ax^k - b))$$



Related Works

- Chen-Liu-Wang et al, Theoretical Linear Convergence of Unfolded ISTA and Its Practical Weights and Thresholds, NeurIPS, 2018
- Chen-Liu-Wang et al, ALISTA: Analytic Weights Are As Good As Learned Weights in LISTA, ICLR, 2019
- 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
- Xiang-Dong-Yang, FISTA-Net: Learning a Fast Iterative Shrinkage Thresholding Network for Inverse Problems in Imaging, IEEE TMI, 2021
- Wu-Zhang-Li et al, RPCANet: Deep Unfolding RPCA Based Infrared Small Target Detection, WACV, 2024
- Zhang-Deng-Xu et al, Deep Semi-Smooth Newton-Driven Unfolding Network for Multi-Modal Image Super-Resolution, IEEE TIP, 2025
- Shlezinger-Segarra-Zhang et al, Deep Unfolding: Recent Developments, Theory, and Design Guidelines, arXiv, 2025

NeurIPS 2018

- Chen-Liu-Wang et al, Theoretical Linear Convergence of Unfolded ISTA and Its Practical Weights and Thresholds, NeurIPS, 2018

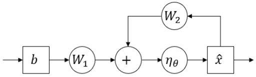
$$\min_x \ \frac{1}{2}\|Ax - b\|^2 + \lambda\|x\|_1$$

$$\Downarrow$$

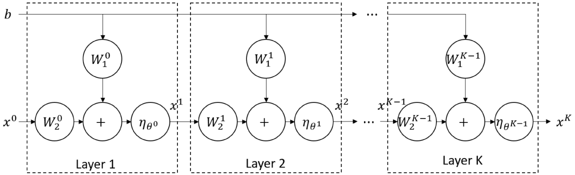
$$x^{k+1} = \eta_{\theta^k}(x^k - \rho A^\top(Ax^k - b))$$

$$= \eta_{\theta^k}(W_1^k b + W_2^k x^k)$$

- Here $W_1 = \rho A^\top, \ W_2 = I - \rho A^\top A, \ \theta = \rho\lambda$ and $k = 0, 1, \cdots, K-1$



(a) RNN structure of ISTA.



(b) Unfolded learned ISTA Network.

NeurIPS 2018

- (Assumption) Signal x^* and the observation noise ε are sampled from

$$
\begin{aligned}
(x^*, \varepsilon) \in \mathcal{X}(\alpha, s, \sigma) \\
&= \{(x^*, \varepsilon) \mid |x_i^*| \leq \alpha, \forall i, \|x^*\|_0 \leq s, \|\varepsilon\|_1 \leq \sigma\}
\end{aligned}
$$

- (Necessary Condition) Given $\{W_1^k, W_2^k, \theta^k\}_{k=0}^{\infty}$ and $x^0 = 0$, let $\{x^k\}_{k=1}^{\infty}$ be generated layer-wise by LISTA. For any $(x^*, \varepsilon) \in \mathcal{X}(\alpha, s, 0)$, if

$$
\begin{aligned}
x^k(\{W_1^\tau, W_2^\tau, \theta^\tau\}_{\tau=0}^{k-1}, b, x^0) \to x^* \quad \text{as} \quad k \to \infty \\
\|W_2^k\|_2 \leq \alpha_W, \quad \forall k = 0, 1, 2, \cdots
\end{aligned}
$$

then $\{W_1^k, W_2^k, \theta^k\}_{k=0}^{\infty}$ must satisfy

$$
\begin{aligned}
& W_2^k - (I - W_1^k A) \to 0 \quad \text{as} \quad k \to \infty \\
& \theta^k \to 0 \quad \text{as} \quad k \to \infty
\end{aligned}
$$

NeurIPS 2018

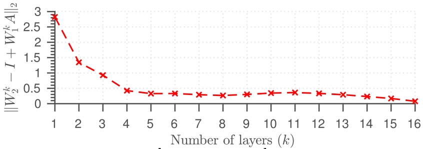
- Let $W^k = (W_1^k)^\top$, thus

$$
\begin{aligned}
\text{(LISTA)} \quad x^{k+1} &= \eta_{\theta^k}(x^k - \rho A^\top(Ax^k - b)) \\
&= \eta_{\theta^k}(W_1^k b + W_2^k x^k)
\end{aligned}
$$

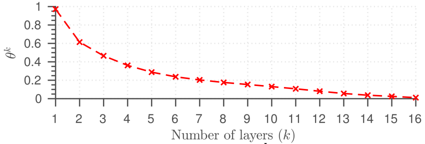
$$\Downarrow$$

$$\text{(LISTA-CP)} \quad x^{k+1} = \eta_{\theta^k}(x^k - (W^k)^\top(Ax^k - b))$$

- Fewer parameters, but better performance



(a) Weight $W_2^k \to I - W_1^k A$ as $k \to \infty$.



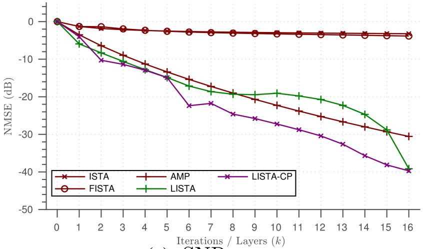
(b) The threshold $\theta^k \to 0$.

NeurIPS 2018

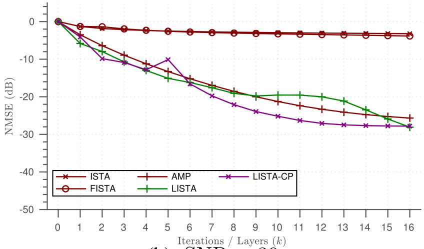
- (Convergence of LISTA-CP) Given $\{W^k, \theta^k\}_{k=0}^{\infty}$ and $x^0 = 0$, if s is sufficiently small, then there exists $\{W^k, \theta^k\}$ such that, for all $(x^*, \varepsilon) \in \mathcal{X}(\alpha, s, \sigma)$, the error bound satisfies

$$\|x^k - x^*\|_2 \leq s\alpha \exp(-ck) + C\sigma, \quad \forall k = 1, 2, \cdots$$

where $c > 0$, $C > 0$ are constants that depend only on A and s



(a) SNR $= \infty$



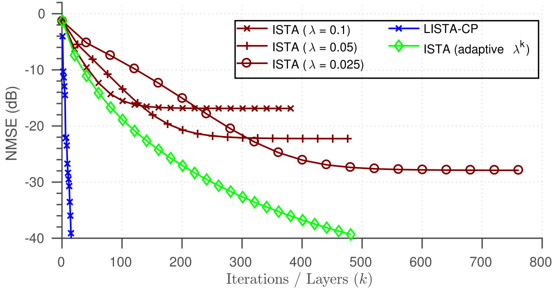
(b) SNR $= 30$

NeurIPS 2018

- Denote $\bar{x}(\lambda)$ be the solution of LASSO, then

$$x^k \to \bar{x}(\lambda) \text{ sublinearly}, \quad \|\bar{x}(\lambda) - x^*\| = O(\lambda), \quad \lambda > 0$$
$$x^k \to \bar{x}(\lambda) \text{ linearly}, \quad \|\bar{x}(\lambda) - x^*\| = O(\lambda), \quad \lambda \text{ large enough}$$

- A larger λ leads to faster convergence but a less accurate solution, and vice versa



ICLR 2019

- Chen-Liu-Wang et al, ALISTA: Analytic Weights Are As Good As Learned Weights in LISTA, ICLR, 2019

$$\min_{x} \ \frac{1}{2}\|Ax-b\|^2 + \lambda\|x\|_1$$

$$\Downarrow$$

$$x^{k+1} = \eta_{\theta^k}(x^k - (W^k)^\top (Ax^k - b))$$

- (Assumption) Signal x^* is sampled from

$$x^* \in \mathcal{X}(\alpha, s) = \{x^* \mid |x_i^*| \leq \alpha, \forall i, \|x^*\|_0 \leq s\}$$

- (Definition) Given $A \in \mathbb{R}^{n\times m}$ with each of its column normalized, define the generalized mutual coherence

$$\tilde{\mu}(A) = \inf_{W\in\mathbb{R}^{n\times m}, (W_{:,i})^\top A_{:,i}=1, 1\leq i\leq m} \{\max_{i\neq j, 1\leq i\leq m} (W_{:,i})^\top A_{:,j}\}$$

If $W \in \mathcal{W}(A) = \{W \in \mathbb{R}^{n\times m} \mid W \text{ attains the infimum}\}$, it is called "good"

ICLR 2019

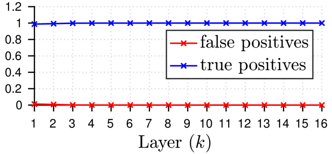
- (Recovery Error Upper Bound) Take any $x^* \in \mathcal{X}(\alpha, s)$, any $W \in \mathcal{W}(A)$, and any sequence $\gamma^k \in (0, \frac{2}{2\tilde{\mu}s - \tilde{\mu} + 1})$. Using them, define the parameters as $\{W^k, \theta^k\}$

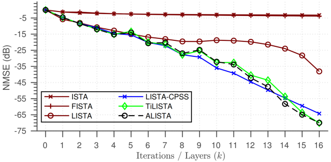
$$W^k = \gamma^k W, \quad \theta^k = \gamma^k \tilde{\mu}(A) \sup_{x^* \in \mathcal{X}(\alpha,s)} \{\|x^k - x^*\|_1\}$$

Let Assumption holds with any $\alpha > 0$ and $s < (1 + 1/\tilde{\mu})/2$. Then, we have

$$\mathrm{support}(x^k) \subset \mathbb{S}, \quad \|x^k - x^*\|_2 \leq s\alpha \exp\left(-\sum_{\tau=0}^{k-1} c^\tau\right)$$

where $\mathbb{S}$ is the support of x^* and $c^k = -log((2\tilde{\mu}s - \tilde{\mu})\gamma^k + |1 - \gamma^k|)$ is a positive constant





ICLR 2019

- Tied LISTA (TiLISTA)

$$x^{k+1} = \eta_{\theta^k}(x^k - (W^k)^\top (Ax^k - b))$$

$$\Downarrow$$

$$x^{k+1} = \eta_{\theta^k}(x^k - \gamma^k W^\top (Ax^k - b))$$

- Analytic LISTA (ALISTA)
  - Step 1: Solve a convex quadratic programming

$$\tilde{W} \in \arg\min_{W \in \mathbb{R}^{n \times m}} \|W^\top A\|_{\mathrm{F}}^2 \quad \text{s.t.} \quad \mathrm{diag}(W^\top A) = 1$$

  - Step 2: End-to-end learning $\mathcal{O}(K + mn) \Rightarrow \mathcal{O}(K)$

$$x^{k+1} = \eta_{\theta^k}(x^k - \gamma^k \tilde{W}^\top (Ax^k - b))$$

CVPR 2018

- Zhang-Ghanem, ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing, CVPR, 2018

$$\min_x \frac{1}{2}\|Ax - b\|^2 + \lambda\|\Psi x\|_1$$

  - Step 1: Compute gradient

$$r^k = x^{k-1} - \rho^k A^\top (Ax^{k-1} - b)$$

  - Step 2: Compute shrinkage

$$x^k = \arg\min_x \frac{1}{2}\|x - r^k\|^2 + \lambda\|\mathcal{F}(x)\|_1$$

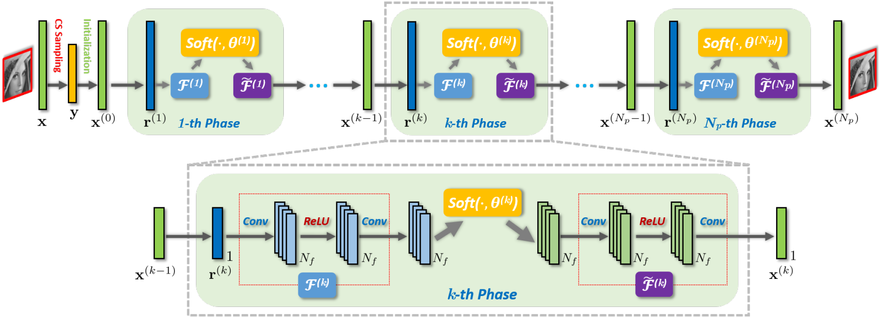
- (Approximation Theory) Let $X_1, \cdots, X_n$ be independent normal random variables with common zero mean and variance σ^2. If $\vec{X} = [X_1, \cdots, X_n]^\top$ and given any matrices A and B, define $\vec{Y} = B\mathrm{ReLU}(A\vec{X}) = B\max(0, A\vec{X})$. Then

$$\mathbb{E}[\|\vec{Y} - \mathbb{E}[\vec{Y}]\|^2] = \alpha\mathbb{E}[\|\vec{X} - \mathbb{E}[\vec{X}]\|^2]$$

CVPR 2018

- Suppose that r^k and $\mathcal{F}(r^k)$ are the mean values of x and $\mathcal{F}(x)$ respectively, then

$$
\begin{aligned}
x^k &= \arg\min_x \ \frac{1}{2}\|x - r^k\|^2 + \lambda\|\mathcal{F}(x)\|_1 \\
&\approx \arg\min_x \ \frac{1}{2}\|\mathcal{F}(x) - \mathcal{F}(r^k)\|^2 + \theta\|\mathcal{F}(x)\|_1 \\
&= \tilde{\mathcal{F}}(soft(\mathcal{F}(r^k), \theta^k))
\end{aligned}
$$



CVPR 2018

- From ISTA-Net to ISTA-Net$^+$

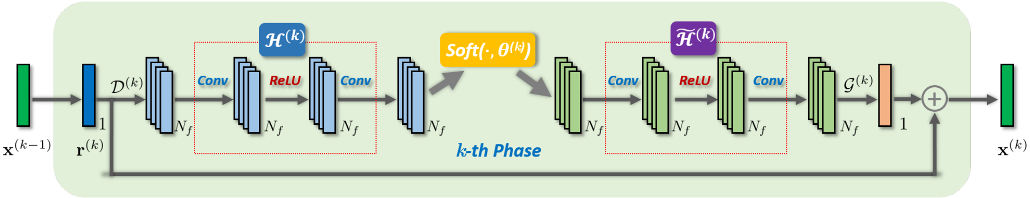
$$\min_x \quad \frac{1}{2}\|\mathcal{F}(x) - \mathcal{F}(r^k)\|^2 + \theta\|\mathcal{F}(x)\|_1$$

$$\Downarrow$$

$$\min_x \quad \frac{1}{2}\|\mathcal{H}(\mathcal{D}(x)) - \mathcal{H}(\mathcal{D}(r^k))\|^2 + \theta\|\mathcal{H}(\mathcal{D}(x))\|_1$$

- Step 2 becomes

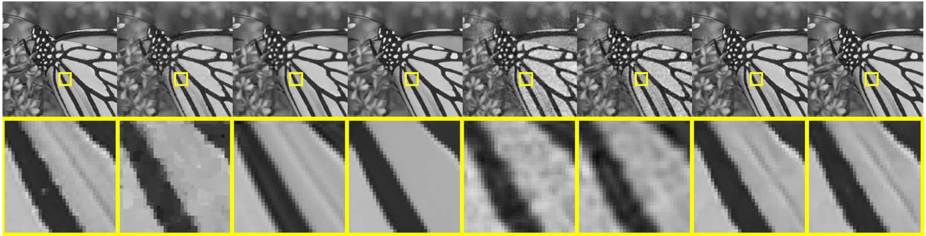
$$x^k = r^k + \mathcal{G}(\tilde{\mathcal{H}}(soft(\mathcal{H}(\mathcal{D}(r^k)), \theta)))$$



CVPR 2018

- Comparison with different CS ratios

Ground Truth | TVAL3 | D-AMP | IRCNN | SDA | ReconNet | ISTA-Net | ISTA-Net$^+$



(PSNR) | (27.83 dB) | (26.80 dB) | (29.94 dB) | (24.53 dB) | (24.55 dB) | (32.31 dB) | **(33.52 dB)**

Algorithm	CS Ratio 50%	40%	30%	25%	10%	4%	1%	Time CPU/GPU
TVAL3 [22]	33.55	31.46	29.23	27.92	22.99	18.75	16.43	3.135s/———
D-AMP [28]	35.92	33.56	30.39	28.46	22.64	18.40	5.21	51.21s/———
IRCNN [49]	36.23	34.06	31.18	30.07	24.02	17.56	7.70	———/68.42s
SDA [30]	28.95	27.79	26.63	25.34	22.65	20.12	17.29	———/0.0032s
ReconNet [21]	31.50	30.58	28.74	25.60	24.28	20.63	17.27	———/0.016s
ISTA-Net	37.43	35.36	32.91	31.53	25.80	21.23	17.30	0.923s/0.039s
ISTA-Net$^+$	**38.07**	**36.06**	**33.82**	**32.57**	**26.64**	**21.31**	**17.34**	1.375s/0.047s

Algorithm	CS Ratio 20%	30%	40%	50%	Time GPU
ADMM-Net	37.17	39.84	41.56	43.00	0.9535s
ISTA-Net	38.30	40.52	42.12	43.60	0.1246s
ISTA-Net$^+$	**38.73**	**40.89**	**42.52**	**44.09**	0.1437s

TIP 2025

- Zhang-Deng-Xu et al, Deep Semi-Smooth Newton-Driven Unfolding Network for Multi-Modal Image Super-Resolution, IEEE TIP, 2025

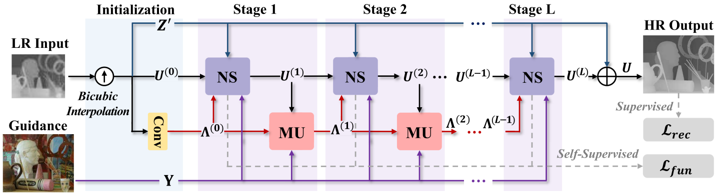
$$\min_{U} \quad \frac{1}{2}\|DU - Z\|_{\mathrm{F}}^2 + \frac{\alpha}{2}\|UH - Y\|_{\mathrm{F}}^2 + \beta\phi(U) + \gamma\psi(U|Y)$$

$$\Downarrow$$

$$(U^l, A^l, B^l) = \arg\min_{U,A,B} \{\mathcal{L}(U, A, B, \Lambda_1^{l-1}, \Lambda_2^{l-1})\}$$

$$\Lambda_1^l = \Lambda_1^{l-1} + \sigma_1(U^l - A^l)$$

$$\Lambda_2^l = \Lambda_2^{l-1} + \sigma_2(U^l - B^l)$$



TIP 2025

- First-order optimality conditions

$$
\begin{aligned}
&0 \in \beta\partial\phi(A) - \Lambda_1 + \sigma_1(A-U) \quad \Rightarrow \quad A = S_{\phi,\hat{\beta}}(\Lambda_1/\sigma_1 + U)\\
&0 \in \gamma\partial\psi(B|Y) - \Lambda_2 + \sigma_2(B-U) \quad \Rightarrow \quad B = S_{\psi,\hat{\gamma},Y}(\Lambda_2/\sigma_2 + U)\\
&0 = D^\top(DU-Z) + \alpha(UH-Y)H^\top + \Lambda_1 + \sigma_1(U-A) + \Lambda_2 + \sigma_2(U-B)
\end{aligned}
$$

- Zhao-Sun-Toh, A Newton-CG Augmented Lagrangian Method for Semidefinite Programming, SIOPT, 2010

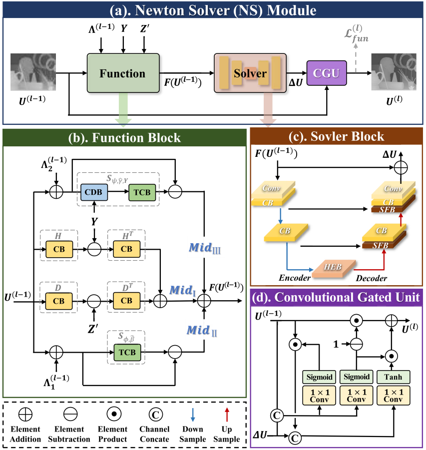
$$
\begin{aligned}
F(U) = D^\top(DU-Z) + \alpha(UH-Y)H^\top + \Lambda_1 + \sigma_1(U - S_{\phi,\hat{\beta}}(\Lambda_1/\sigma_1 + U))&\\
+ \Lambda_2 + \sigma_2(U - S_{\psi,\hat{\gamma},Y}(\Lambda_2/\sigma_2 + U)) = 0&
\end{aligned}
$$

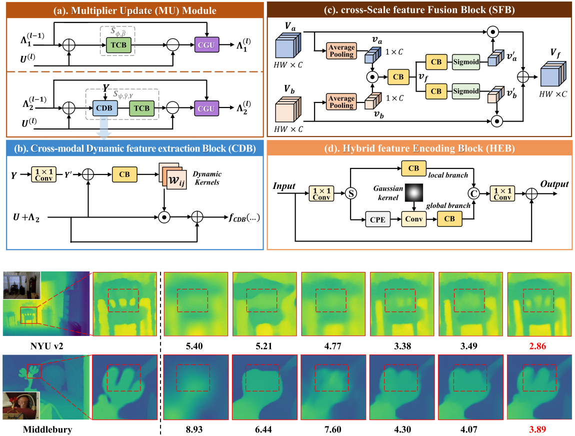
$$\Downarrow$$

$$
\begin{aligned}
&V(U_{m-1})\triangle U_m = -F(U_{m-1})\\
&U_m = U_{m-1} + \eta\triangle U_m
\end{aligned}
$$

TIP 2025

▶ Detailed structures and visual results

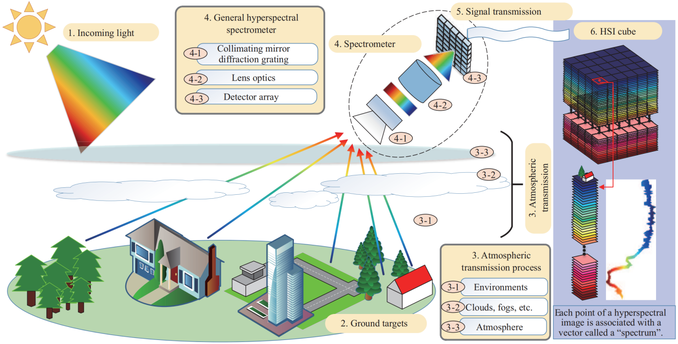




Outline

HSI

- Hyperspectral image (HSI): Rich spectral and spatial information

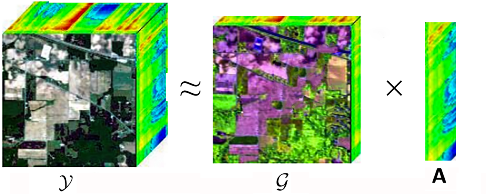


- Board applications: Agriculture, astronomy, geosciences, medicine
- Various tasks: Denoising, classification, detection, fusion, unmixing
- https://github.com/xianchaoxiu/Hyperspectral-Imaging

Motivation

- General subspace tensor representation framework

$$
\begin{aligned}
(\mathrm{P}) \quad & \min_{\mathcal{G},A} \ \frac{1}{2}\|\mathcal{Y} - \mathcal{G} \times_3 A\|_{\mathrm{F}}^2 + \lambda\Omega(\mathcal{G}) \\
& \mathrm{s.t.} \ \ A^\top A = I
\end{aligned}
$$



- What are the advantages of tensor modeling?
  - Excellent data representation
  - Various tensor decomposition
  - (possibly) Low computational complexity

Motivation

- Xiong-Zhou-Tao et al, SMDS-Net: Model Guided Spectral-Spatial Network for Hyperspectral Image Denoising, IEEE TIP, 2022

$$
\begin{aligned}
(\mathrm{P}) \quad \min_{\mathcal{G},A} \quad & \frac{1}{2}\|\mathcal{Y} - \mathcal{G} \times_3 A\|_{\mathrm{F}}^2 + \lambda\Omega(\mathcal{G}) \\
\text{s.t.} \quad & A^\top A = I
\end{aligned}
$$

$$\Downarrow$$

$$
\begin{aligned}
\min_{\mathcal{G},\mathcal{B}_i,A} \quad & \frac{1}{2}\|\mathcal{Y} - \mathcal{G} \times_3 A\|_{\mathrm{F}}^2 + \lambda \sum_i (\phi(\mathcal{G}, \mathcal{B}_i) + \gamma_1 \|\mathcal{B}_i\|_1) \\
\text{s.t.} \quad & A^\top A = I
\end{aligned}
$$

- Multidimentional representation

$$\phi(\mathcal{G}, \mathcal{B}_i) = \frac{1}{2}\|\mathcal{R}_i\mathcal{G} - \mathcal{B}_i \times_1 D_1 \times_2 D_2 \times_3 D_3\|_{\mathrm{F}}^2$$

- How to characterize priors? How to develop algorithms?

Model

- Sparse tensor aided representation (STAR)

$$
\begin{aligned}
\text{(P)} \quad & \min_{\mathcal{G},A} \ \frac{1}{2}\|\mathcal{Y} - \mathcal{G} \times_3 A\|_{\mathrm{F}}^2 + \lambda \Omega(\mathcal{G}) \\
& \text{s.t.} \quad A^\top A = I
\end{aligned}
$$

$$\Downarrow$$

$$
\begin{aligned}
\text{(STAR)} \quad & \min_{\mathcal{G},\mathcal{B}_i,A} \ \frac{1}{2}\|\mathcal{Y} - \mathcal{G} \times_3 A\|_{\mathrm{F}}^2 + \lambda \sum_i \left(\phi(\mathcal{G},\mathcal{B}_i) + \gamma_1\|\mathcal{B}_i\|_1 + \gamma_2\|\mathcal{B}_i\|_*\right) \\
& \text{s.t.} \quad A^\top A = I
\end{aligned}
$$

$$\Downarrow$$

$$
\begin{aligned}
\text{(STAR-S)} \quad & \min_{\mathcal{G},\mathcal{S},\mathcal{B}_i,A} \ \frac{1}{2}\|\mathcal{Y} - \mathcal{G} \times_3 A - \mathcal{S}\|_{\mathrm{F}}^2 + \mu\|\mathcal{S}\|_1 \\
& \qquad + \lambda \sum_i \left(\phi(\mathcal{G},\mathcal{B}_i) + \gamma_1\|\mathcal{B}_i\|_1 + \gamma_2\|\mathcal{B}_i\|_*\right) \\
& \text{s.t.} \quad A^\top A = I
\end{aligned}
$$

Algorithm

- Alternating direction method of multipliers (ADMM)

$$\begin{aligned} \min_{\mathcal{G},\mathcal{B}_i,\mathcal{L}_i,A} \quad & \frac{1}{2}\|\mathcal{Y} - \mathcal{G} \times_3 A\|_{\mathrm{F}}^2 + \lambda \sum_i \left(\phi(\mathcal{G}, \mathcal{B}_i) + \gamma_1 \|\mathcal{B}_i\|_1 + \gamma_2 \|\mathcal{L}_i\|_*\right) \\ \text{s.t.} \quad & A^\top A = I, \ \mathcal{L}_i = \mathcal{B}_i \end{aligned}$$

$$\Downarrow$$

$$\begin{aligned} L_\beta(\mathcal{G}, \mathcal{B}_i, \mathcal{L}_i, A, \mathcal{P}_i) = & \frac{1}{2}\|\mathcal{Y} - \mathcal{G} \times_3 A\|_{\mathrm{F}}^2 \\ & + \lambda \sum_i \left(\phi(\mathcal{G}, \mathcal{B}_i) + \gamma_1 \|\mathcal{B}_i\|_1 + \gamma_2 \|\mathcal{L}_i\|_*\right) \\ & + \langle \mathcal{P}_i, \mathcal{L}_i - \mathcal{B}_i \rangle + \frac{\beta}{2}\|\mathcal{L}_i - \mathcal{B}_i\|_{\mathrm{F}}^2 \end{aligned}$$

- Note that $A^\top A = I$ is preserved

STAR-Net

- Update $\mathcal{G}$-block

$$\begin{aligned}\mathcal{G}^{k+1} = \arg\min_{\mathcal{G}} \ & \frac{1}{2}\|\mathcal{Y} - \mathcal{G} \times_3 A^k\|_{\mathrm{F}}^2 \\ & + \frac{\lambda}{2}\sum_i \|\mathcal{R}_i\mathcal{G} - \mathcal{B}_i^k \times_1 D_1 \times_2 D_2 \times_3 D_3\|_{\mathrm{F}}^2\end{aligned}$$

$$\Downarrow$$

$$\mathcal{G}^{k+1} = (\mathcal{I} + \lambda \sum_i \mathcal{R}_i^\top \mathcal{R}_i)^{-1}(\lambda \sum_i \mathcal{R}_i^\top \mathcal{B}_i^k \times_1 D_1 \times_2 D_2 \times_3 D_3 + \mathcal{Y} \times_3 A^{k\top})$$

$$\Downarrow$$

$$\mathcal{G}^{k+1} = \mathcal{E}_1 * \mathcal{E}_2$$

$$\Downarrow$$

$$\mathcal{G}^{k+1} = \mathrm{LargNet}(\mathcal{E}_1, \mathcal{E}_2)$$

STAR-Net

- Update $\mathcal{B}_i$-block

$$\mathcal{B}_i^{k+1} = \arg\min_{\mathcal{B}_i} \frac{\lambda}{2}\|\mathcal{R}_i\mathcal{G}^{k+1} - \mathcal{B}_i \times_1 D_1 \times_2 D_2 \times_3 D_3\|_{\mathrm{F}}^2 + \frac{\beta}{2}\|\mathcal{L}_i^k - \mathcal{B}_i + \mathcal{P}_i^k/\beta\|_{\mathrm{F}}^2 + \lambda\gamma_1\|\mathcal{B}_i\|_1$$

$$\Downarrow$$

$$\mathcal{B}_i^{k+1} = \arg\min_{\mathcal{B}_i} \frac{1}{2}\|(\beta\mathcal{I} + \lambda\mathcal{I} \times_1 D_1 \times_2 D_2 \times_3 D_3)\mathcal{B}_i - (\lambda\mathcal{R}_i\mathcal{G}^{k+1} + \beta\mathcal{L}_i^k + \mathcal{P}_i^k)\|_{\mathrm{F}}^2 + \lambda\gamma_1\|\mathcal{B}_i\|_1$$

$$\Downarrow$$

$$\mathcal{B}_i^{k+1} = \mathrm{sgn}(\mathcal{F}_i) \circ \mathrm{ReLU}(|\mathcal{F}_i| - \lambda\gamma_1/\nu)$$

$$\Downarrow$$

$$\mathcal{B}_i^{k+1} = \mathrm{ShrinkNet}(\mathcal{F}_i, \lambda\gamma_1/\nu)$$

STAR-Net

- Update A-block

$$A^{k+1} = \arg \min_{A^\top A = I} \frac{1}{2} \|\mathcal{Y} - \mathcal{G}^{k+1} \times_3 A\|_{\mathrm{F}}^2$$
$$\Downarrow$$
$$A^{k+1} = UV^\top$$
$$\Downarrow$$
$$A^{k+1} = \mathrm{LargNet}(U, V^\top)$$

- Update $\mathcal{P}_i$-block

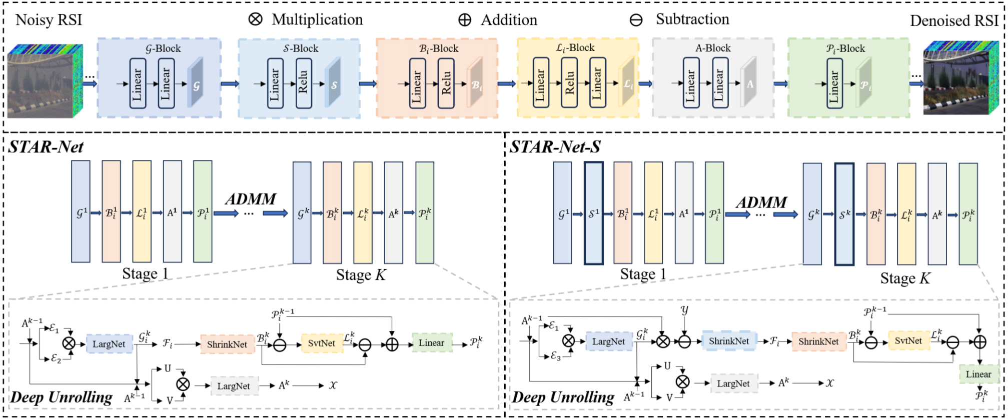
$$\mathcal{P}_i^{k+1} = \mathcal{P}_i^k + \beta(\mathcal{L}_i^{k+1} - \mathcal{B}_i^{k+1})$$
$$\Downarrow$$
$$\mathcal{P}_i^{k+1} = \mathrm{Linear}(\Theta_i)$$

STAR-Net

- Input: Noisy HSI $\mathcal{Y}$, parameters $\lambda, \beta, \gamma_1, \gamma_2, \nu$
- Initialize: $(\mathcal{G}^0, \mathcal{B}_i^0, \mathcal{L}_i^0, A^0, \mathcal{P}_i^0)$
- While $k = 1, \ldots, K$ do
  - Update $\mathcal{G}$-block by
    $$\mathcal{G}^{k+1} = \mathrm{LargNet}(\mathcal{E}_1, \mathcal{E}_2)$$
  - Update $\mathcal{B}_i$-block by
    $$\mathcal{B}_i^{k+1} = \mathrm{ShrinkNet}(\mathcal{F}_i, \lambda\gamma_1/\nu)$$
  - Update $\mathcal{L}_i$-block by
    $$\mathcal{L}_i^{k+1} = \mathrm{SvtNet}(\mathcal{B}_i^{k+1} - \mathcal{P}_i^k/\beta, \lambda\gamma_2/\beta)$$
  - Update A-block by
    $$A^{k+1} = \mathrm{LargNet}(U, V^\top)$$
  - Update $\mathcal{P}_i$-block by
    $$\mathcal{P}_i^{k+1} = \mathrm{Linear}(\Theta_i)$$
- Output: Denoised HSI $\mathcal{X} = \mathcal{G}^{k+1} \times_3 A^{k+1}$

Architecture

- Our proposed STAR-Net and STAR-Net-S



Setup

- Compared methods
  - BM4D: Maggioni-Katkovnik-Egiazarian et al, IEEE TIP, 2012
  - LLRT: Chang-Yan-Zhong, CVPR, 2017
  - LRTDTV: Wang-Chen-Han et al, RS, 2017
  - NGMeet: He-Yao-Li-Yokoya et al, IEEE TPAMI, 2022
  - NLSSR: Zha-Wen-Yuan et al, IEEE TGRS, 2023
  - FastHyMix: Zhuang-Ng, IEEE TNNLS, 2023
  - HSI-SDeCNN: Maffei-Haut-Paoletti et al, IEEE TGRS, 2020
  - SMDS-Net: Xiong-Zhou-Tao et al, IEEE TIP, 2022
  - Eigen-CNN: Zhuang-Ng-Gao et al, IEEE TGRS, 2024
  - RCILD: Peng-Wang-Cao et al, IEEE TGRS, 2024
- Implementation details
  - Training loss: $L = \|\mathrm{STAR\text{-}Net}(\mathcal{Y}) - \mathcal{X}\|_{\mathrm{F}}^2$
  - Training dataset: 100 HSIs from ICVL, data augmentation
  - Testing dataset: ICVL, Washington DC Mall
  - Hyperparameters: Learning rate=0.005, batch=2, epoch=300, K=6, dictionary=[9, 9, 9]
  - Training parameters: $\lambda, \beta, \mu, \gamma_1, \gamma_2, D_1, D_2, D_3$

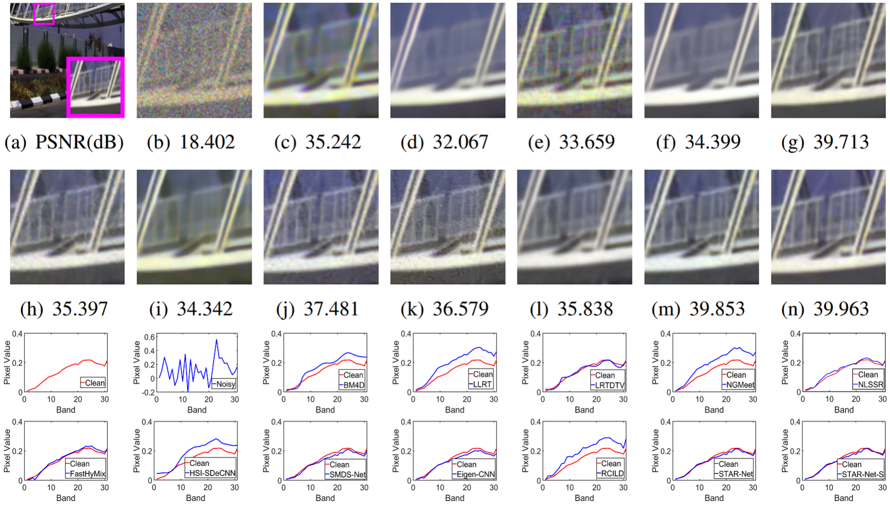
Synthetic

- Quantitative comparisons on ICVL

σ	Index	Noisy	BM4D	LLRT	LRTDTV	NGMeet	NLSSR	FastHyMix	HSI-SDeCNN	SMDS-Net	Eigen-CNN	RCILD	STAR-Net	STAR-Net-S
10	PSNR ↑	29.018	42.987	39.810	43.882	42.383	45.928	43.628	41.519	46.371	**47.321**	42.458	47.286	**47.345**
	SSIM ↑	0.521	0.973	0.962	0.979	0.968	0.984	**0.988**	0.969	0.985	**0.989**	0.987	**0.988**	**0.989**
	SAM ↓	0.229	0.080	0.045	0.077	0.074	0.066	0.035	0.075	**0.028**	0.032	0.044	**0.025**	**0.025**
	ERGAS ↓	243.021	36.420	59.279	44.893	34.764	28.026	24.893	61.289	20.056	25.355	25.800	**18.124**	**17.951**
30	PSNR ↑	21.591	37.630	34.250	38.245	36.791	41.629	38.286	36.840	42.337	41.491	38.514	**42.435**	**42.500**
	SSIM ↑	0.146	0.930	0.921	0.877	0.915	0.968	0.966	0.926	**0.972**	0.963	**0.971**	**0.972**	**0.972**
	SAM ↓	0.535	0.142	0.084	0.149	0.139	0.084	0.068	0.124	0.040	0.060	0.067	**0.039**	**0.038**
	ERGAS ↓	729.026	62.662	40.725	69.464	60.101	41.388	48.675	103.084	32.393	49.458	49.362	**32.056**	**31.862**
50	PSNR ↑	18.402	35.242	32.067	33.659	34.399	39.713	35.397	34.342	37.481	36.579	35.838	**39.853**	**39.963**
	SSIM ↑	0.042	0.888	0.899	0.862	0.887	**0.955**	0.941	0.893	0.907	0.879	0.951	**0.956**	**0.956**
	SAM ↓	0.779	0.190	0.107	0.195	0.177	0.109	0.096	0.134	0.066	0.106	0.092	**0.050**	**0.047**
	ERGAS ↓	1215.105	81.133	54.869	110.351	80.585	53.811	68.681	136.304	55.786	85.426	68.945	**43.362**	**42.923**
70	PSNR ↑	18.126	33.586	30.746	30.565	32.389	**37.450**	33.377	32.794	37.197	32.194	33.980	37.342	**38.237**
	SSIM ↑	0.038	0.844	0.852	0.762	0.858	**0.934**	0.915	0.855	0.923	0.752	0.930	**0.943**	**0.943**
	SAM ↓	0.897	0.231	0.214	0.304	0.217	0.128	0.120	0.186	0.066	0.154	0.098	**0.058**	**0.055**
	ERGAS ↓	1701.060	97.267	66.690	163.788	97.309	65.893	88.790	173.430	58.223	126.017	89.340	**57.341**	**52.261**
Average	PSNR ↑	21.784	37.361	34.218	36.588	36.490	41.180	37.672	36.374	40.846	39.396	37.697	**41.729**	**42.011**
	SSIM ↑	0.187	0.909	0.909	0.870	0.907	**0.960**	0.953	0.911	0.947	0.896	**0.960**	**0.965**	**0.965**
	SAM ↓	0.610	0.161	0.113	0.181	0.152	0.097	0.080	0.130	0.050	0.088	0.075	**0.043**	**0.041**
	ERGAS ↓	972.053	69.370	55.391	97.124	68.190	47.280	57.760	118.527	41.614	71.564	58.362	**37.721**	**36.249**

Synthetic

- Visualization results



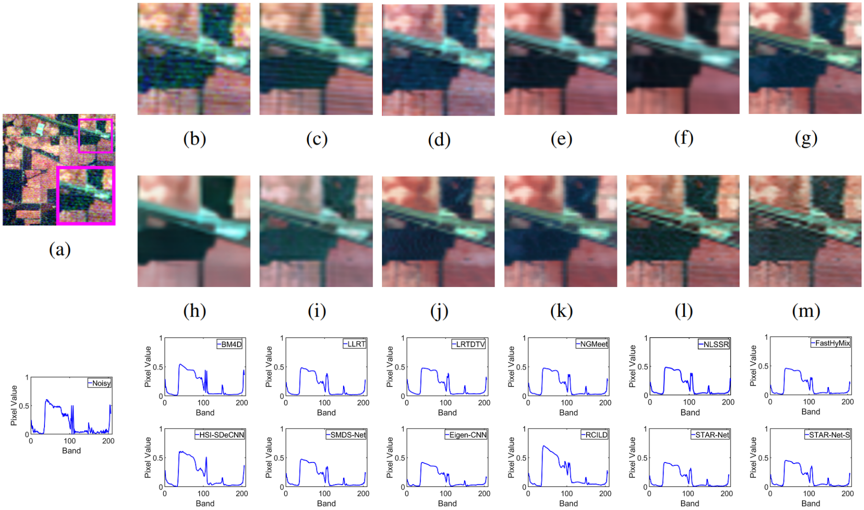
Synthetic

- Quantitative comparisons on PaviaU

σ	Index	Noisy	BM4D	LLRT	LRTDTV	NGMeet	NLSSR	FastHyMix	HSI-SDeCNN	SMDS-Net	Eigen-CNN	RCILD	STAR-Net	STAR-Net-S
10	PSNR ↑	29.181	34.044	30.671	31.796	36.493	35.419	39.646	38.509	35.882	39.660	40.509	40.881	**41.172**
	SSIM ↑	0.654	0.902	0.821	0.856	0.933	0.925	**0.968**	0.953	0.938	0.969	0.962	**0.971**	**0.971**
	SAM ↓	0.221	0.108	0.159	0.141	0.084	0.092	0.060	0.067	0.068	0.060	**0.056**	**0.049**	**0.049**
	ERGAS ↓	157.975	77.328	116.590	102.736	60.251	67.076	44.100	48.351	67.165	44.016	44.690	**40.944**	**39.927**
30	PSNR ↑	21.409	28.398	30.394	31.756	30.448	33.835	32.960	32.236	30.010	33.028	**35.947**	35.637	**35.976**
	SSIM ↑	0.237	0.746	0.809	0.855	0.818	0.901	0.919	0.849	0.929	0.922	0.899	**0.935**	**0.936**
	SAM ↓	0.580	0.202	0.165	0.141	0.610	0.111	0.125	0.134	0.077	0.124	0.090	**0.074**	**0.073**
	ERGAS ↓	473.723	147.139	120.494	103.199	119.247	80.904	96.429	98.539	73.304	95.860	69.178	**68.764**	**66.131**
50	PSNR ↑	18.760	26.159	27.022	31.648	27.756	31.750	31.909	29.198	31.748	32.031	31.416	**33.227**	**33.243**
	SSIM ↑	0.114	0.650	0.666	0.850	0.722	0.856	0.897	0.757	0.878	**0.899**	0.831	0.898	**0.902**
	SAM ↓	0.816	0.259	0.242	0.196	0.220	0.140	0.138	0.182	0.098	0.136	0.150	**0.093**	**0.092**
	ERGAS ↓	789.831	188.949	177.238	104.514	162.735	103.138	105.438	138.135	104.315	104.304	115.949	**89.223**	**88.743**
70	PSNR ↑	16.705	24.940	26.626	30.644	26.281	30.180	31.087	27.435	30.989	31.097	30.560	**31.658**	**31.694**
	SSIM ↑	0.064	0.595	0.643	0.820	0.659	0.814	**0.876**	0.694	0.859	**0.872**	0.779	0.868	0.868
	SAM ↓	0.971	0.298	0.254	0.160	0.265	0.166	0.150	0.215	**0.113**	0.150	0.157	**0.105**	**0.105**
	ERGAS ↓	1105.483	216.652	185.492	117.070	193.479	122.756	114.277	169.228	113.503	114.060	126.980	**106.249**	**105.060**
Average	PSNR ↑	21.514	28.385	28.678	31.461	30.245	32.796	33.901	31.844	32.157	33.954	34.608	**35.351**	**35.521**
	SSIM ↑	0.267	0.723	0.735	0.845	0.783	0.874	0.915	0.813	0.901	0.915	0.868	**0.918**	**0.919**
	SAM ↓	0.647	0.217	0.205	0.159	0.295	0.127	0.118	0.149	**0.089**	0.118	0.113	**0.080**	**0.080**
	ERGAS ↓	631.753	157.517	149.953	106.880	133.928	93.468	90.061	113.563	89.572	89.560	89.199	**76.295**	**74.965**

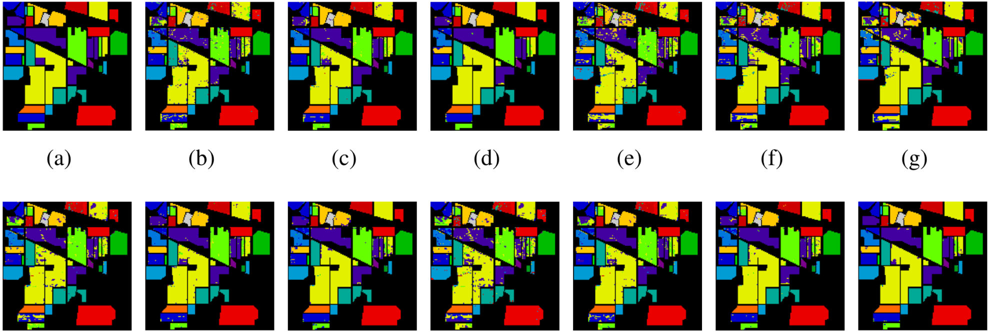
Real-World

- Case study on Indian Pines



Real-World

- Classification



(a) (b) (c) (d) (e) (f) (g)

(h) (i) (j) (k) (l) (m) (n)

Discussion

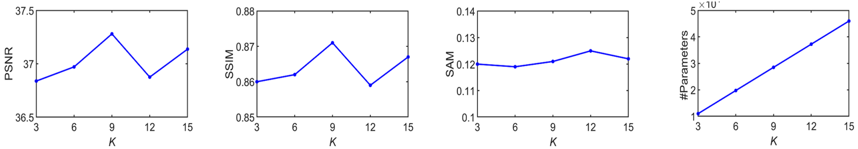
Testing runtime

Datasets	BM4D	LLRT	LRTDTV	NGMeet	NLSSR	FastHyMix	HSI-SDeCNN	SMDS-Net	Eigen-CNN	RCILD	STAR-Net	STAR-Net-S
ICVL	561.307	888.305	136.796	335.265	206.964	**8.555**	11.093	105.109	**6.478**	24.743	107.552	107.958
PaviaU	1123.815	2581.424	354.911	716.722	491.568	82.029	60.522	349.292	**11.361**	**30.706**	337.133	341.220

Number of parameters

Methods	FastHyMix	HSI-SDeCNN	SMDS-Net	Eigen-CNN	RCILD	STAR-Net	STAR-Net-S
#Parameters	/	1892100	**5103**	/	2892288	**27702**	28487

Number of iterations

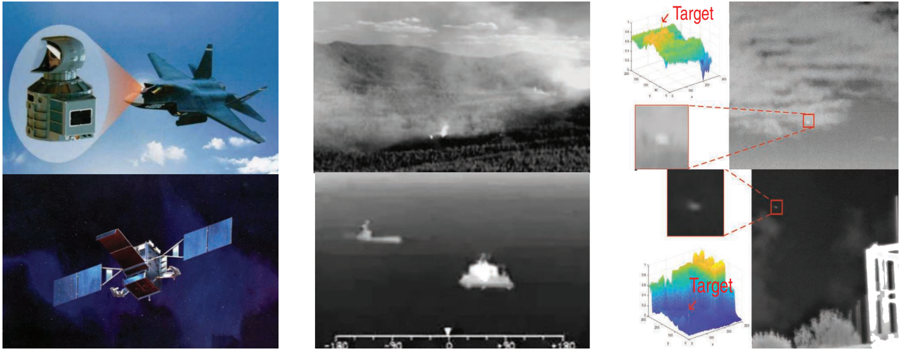


(a) PSNR (b) SSIM (c) SAM (d) #Parameters

Outline

ISTD

- Infrared small target detection (ISTD)

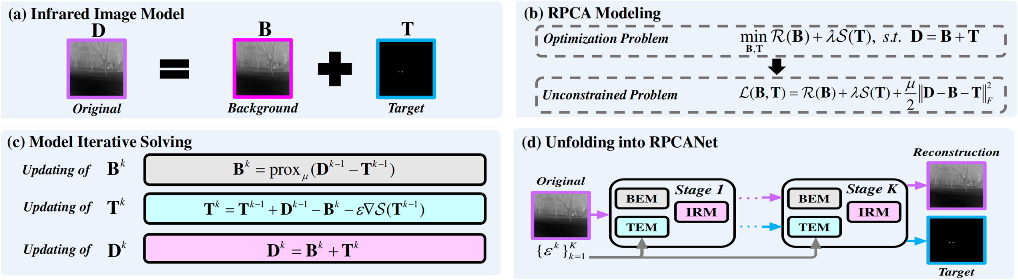


- Difficulties: Small size, low SNR, weak contrast
- Advantages: Strong concealment, good anti-interference
- It is an urgent need in aerospace

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)

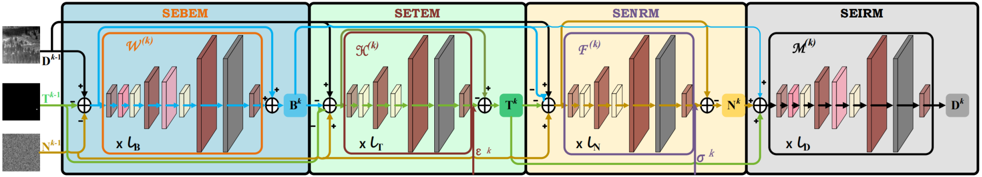
$$B^k = \arg\min_B \ \mathcal{R}(B) + \frac{\alpha}{2}\|D^{k-1} - B - T^{k-1} - N^{k-1}\|_F^2$$

$$\Downarrow$$

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

$$\Downarrow$$

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



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 $\gamma = 0.5$

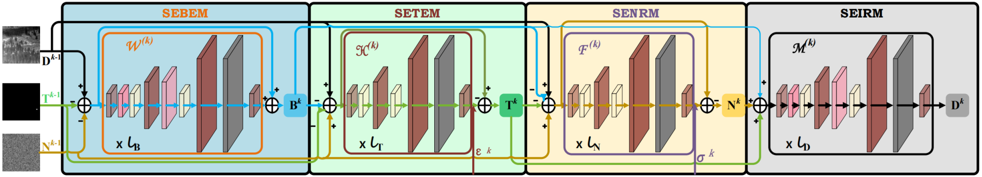
$$T^k = \gamma T^{k-1} + (1-\gamma)(D^{k-1} - B^k - N^{k-1}) - \varepsilon \nabla \mathcal{S}(T^{k-1})$$

$$\Downarrow$$

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

$$\Downarrow$$

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



Update N

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

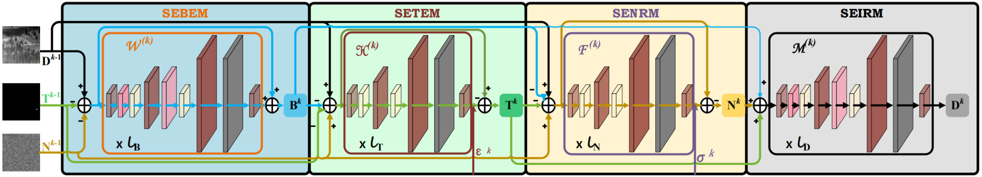
$$N^k = \arg\min_N \ \mu\mathcal{G}(N) + \frac{\alpha}{2}\|D^{k-1} - B^k - T^k - N\|_{\mathrm{F}}^2$$

$$\Downarrow$$

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

$$\Downarrow$$

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



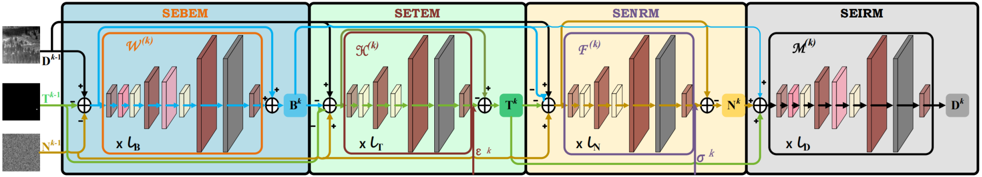
Update D

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

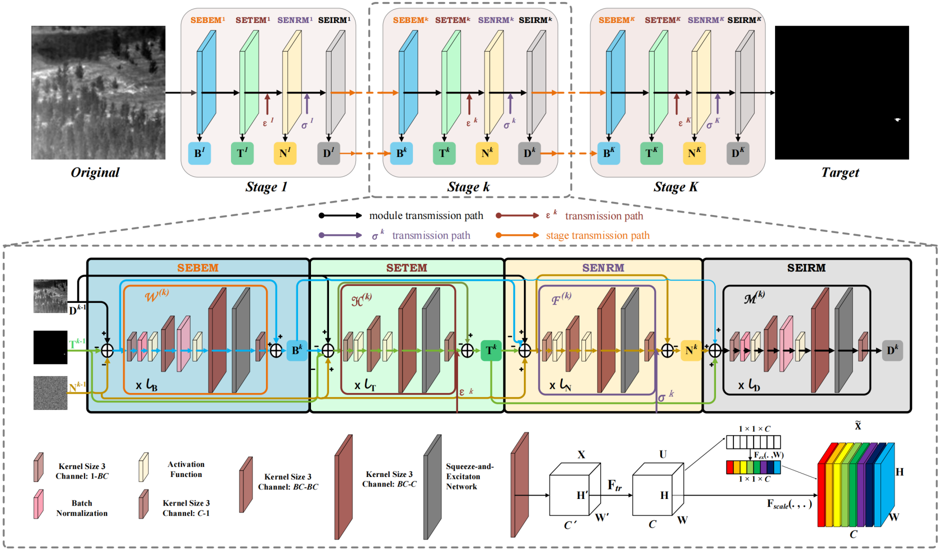
$$D^k = B^k + T^k + N^k$$

$$\Downarrow$$

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



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 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 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 (F_1 $\uparrow$), Probability of detection (P_d $\uparrow$)
  - False alarm rate (F_a $\downarrow$)
- Loss function

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

Experiments

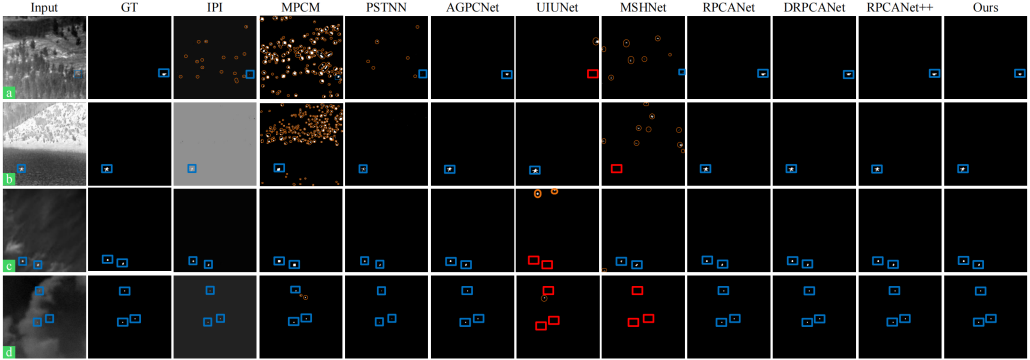
Quantitative results

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

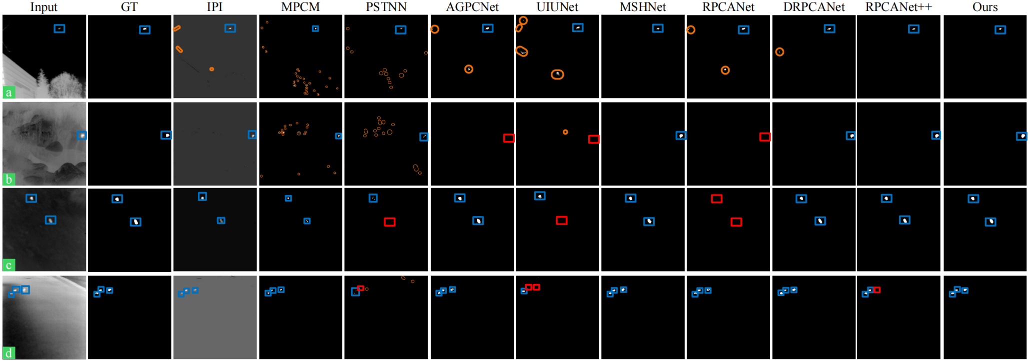
Experiments

- Visualization on NUDT-SIRST



Experiments

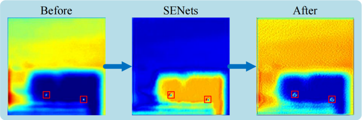
- Visualization on IRSTD-1k



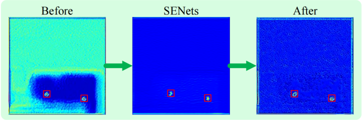
Experiments

▶ Ablation studies on SENets

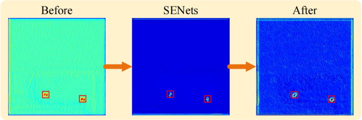
SENets				NUDT-SIRST				SIRST-Aug				IRSTD-1k			
SEBEM	SETEM	SENRM	SEIRM	mIoU ↑	F_1 ↑	P_d ↑	F_a ↓	mIoU ↑	F_1 ↑	P_d ↑	F_a ↓	mIoU ↑	F_1 ↑	P_d ↑	F_a ↓
✗	✗	✗	✗	73.56	78.45	79.36	8.56	60.75	70.28	81.24	43.67	50.26	61.34	70.57	15.95
✓	✗	✗	✗	80.57	81.39	86.35	5.68	65.96	75.37	86.99	39.82	55.84	65.26	76.36	11.12
✓	✓	✗	✗	88.36	89.45	90.18	4.00	70.17	80.78	93.17	34.78	60.56	72.58	83.70	8.10
✓	✓	✓	✗	91.14	96.06	97.18	2.05	73.27	84.45	98.07	27.73	63.58	77.53	88.71	3.95
✓	✓	✓	✓	**92.37**	**96.54**	**98.41**	**1.79**	**74.56**	**85.43**	**99.17**	**29.78**	**64.68**	**78.55**	**89.39**	**4.66**



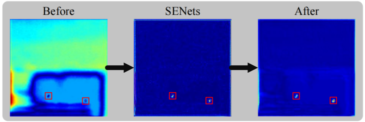
(a) SEBEM



(b) SETEM



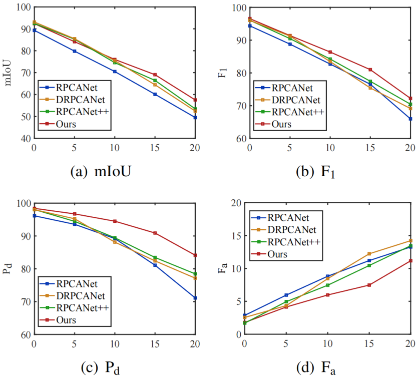
(c) SENRM



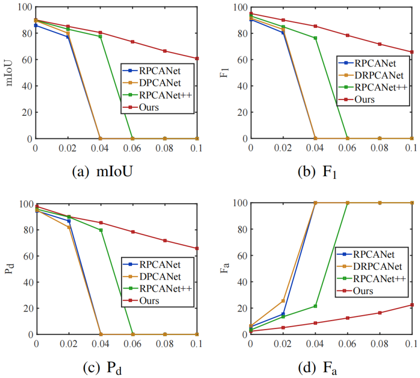
(d) SEIRM

Experiments

- Gaussian noise



- Salt-and-pepper noise

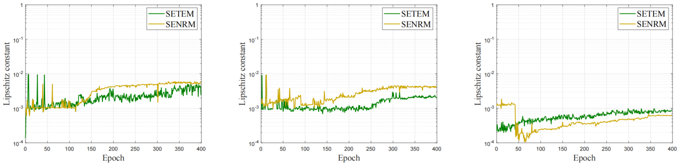


Experiments

- 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

- Lipschitz conditions



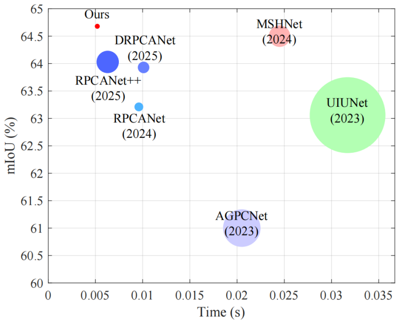
(a) NUDT-SIRST (b) SIRST-Aug (c) IRSTD-1k

Outline

Conclusions

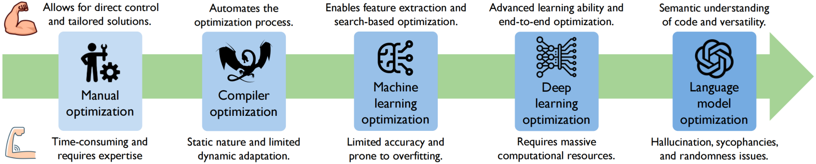
- Liu-Jin-Xiu et al, STAR-Net: An Interpretable Model-Aided Network for Remote Sensing Image Denoising, Pattern Recognition, 172: 112496, 2026
- Liu-Han-Xiu et al, Lightweight Deep Unfolding Networks with Enhanced Robustness for Infrared Small Target Detection, submitted to IEEE TIP, 2025

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



Future Work

- Data-driven optimization (DDO)
  - How to reduce complexity
  - How to guarantee convergence
- Further comments
  - When it outperform traditional optimization methods
  - When it outperform large language models



Thank you for your attention!

xcxiu@shu.edu.cn