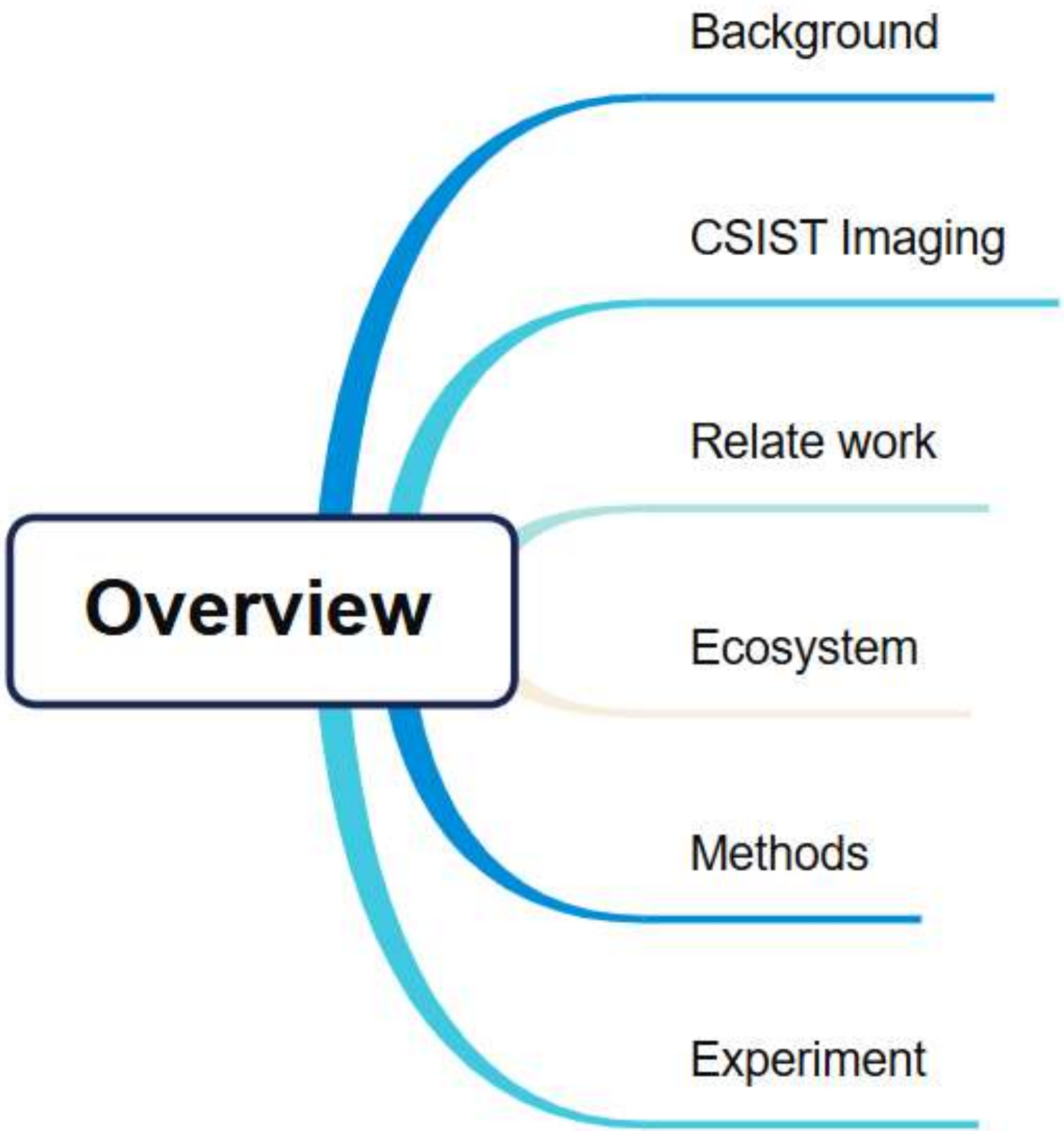
[ICCV 2025] DISTA-Net: Dynamic Closely-Spaced Infrared Small Target Unmixing

Shengdong Han

arXiv https://arxiv.org/abs/2505.19148

Code https://github.com/GrokCV/GrokCSO

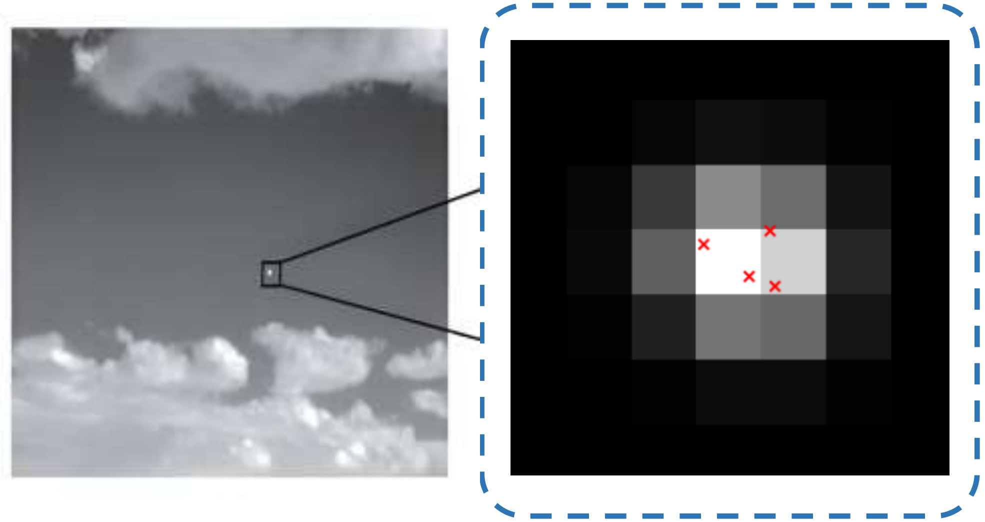
DISTA-Net: Dynamic Closely-Spaced Infrared Small Target Unmixing



Background

- Closely-Spaced Infrared Small Targets(CSIST)：Signal aliasing in high-density clusters of distant targets (drones/missiles) impedes accurate target counting, sub-pixel localization, and radiometric analysis.

(a) **Fixed-Wing Swarm UAV Platform**



(b) **Infrared Imaging of Closely-Spaced Targets**

Background—CSIST Unmixing

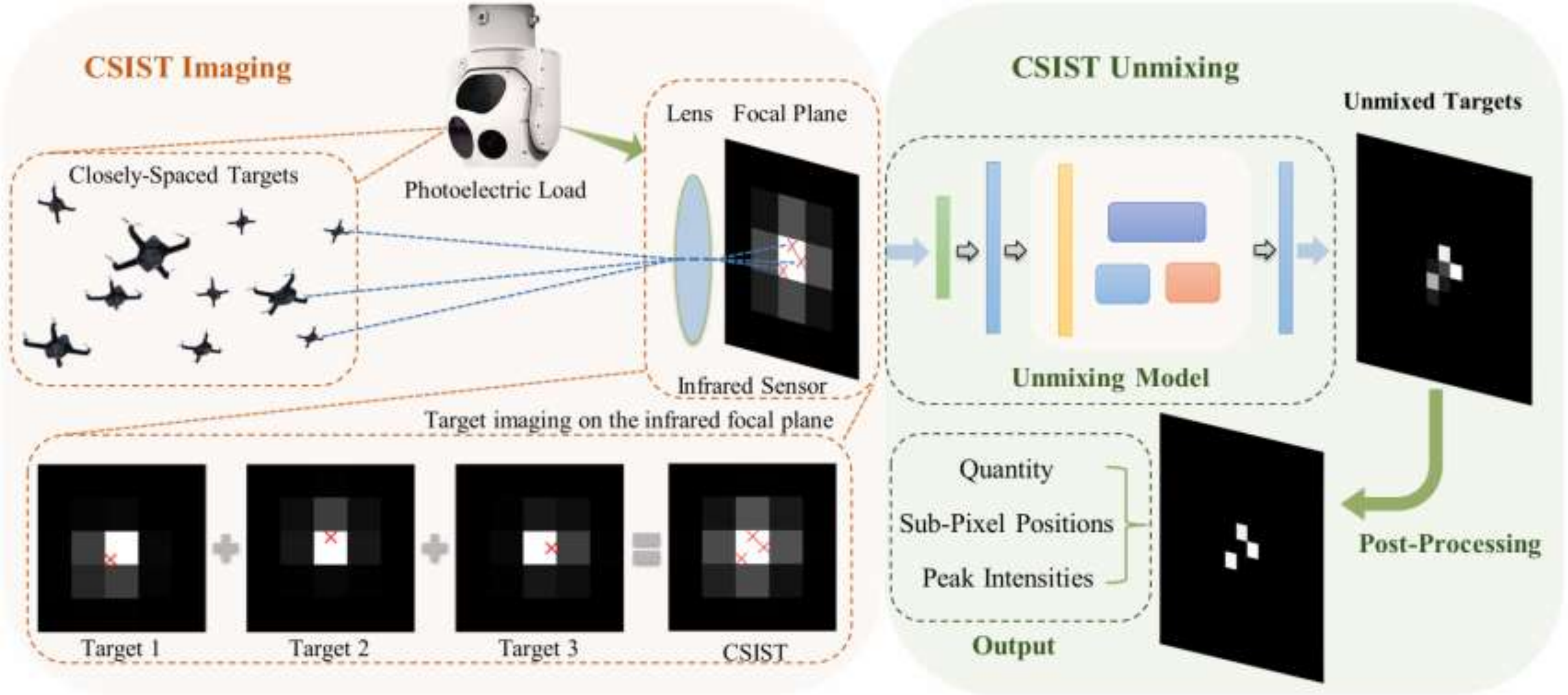
ISTD

1. Fundamental Detection Stage
2. Position Recognition
3. Binary Decision

VS

1. Post-processing Reconstruction Stage
2. Sub-pixel Localization
3. Multi-dimensional Analysis

CSIST

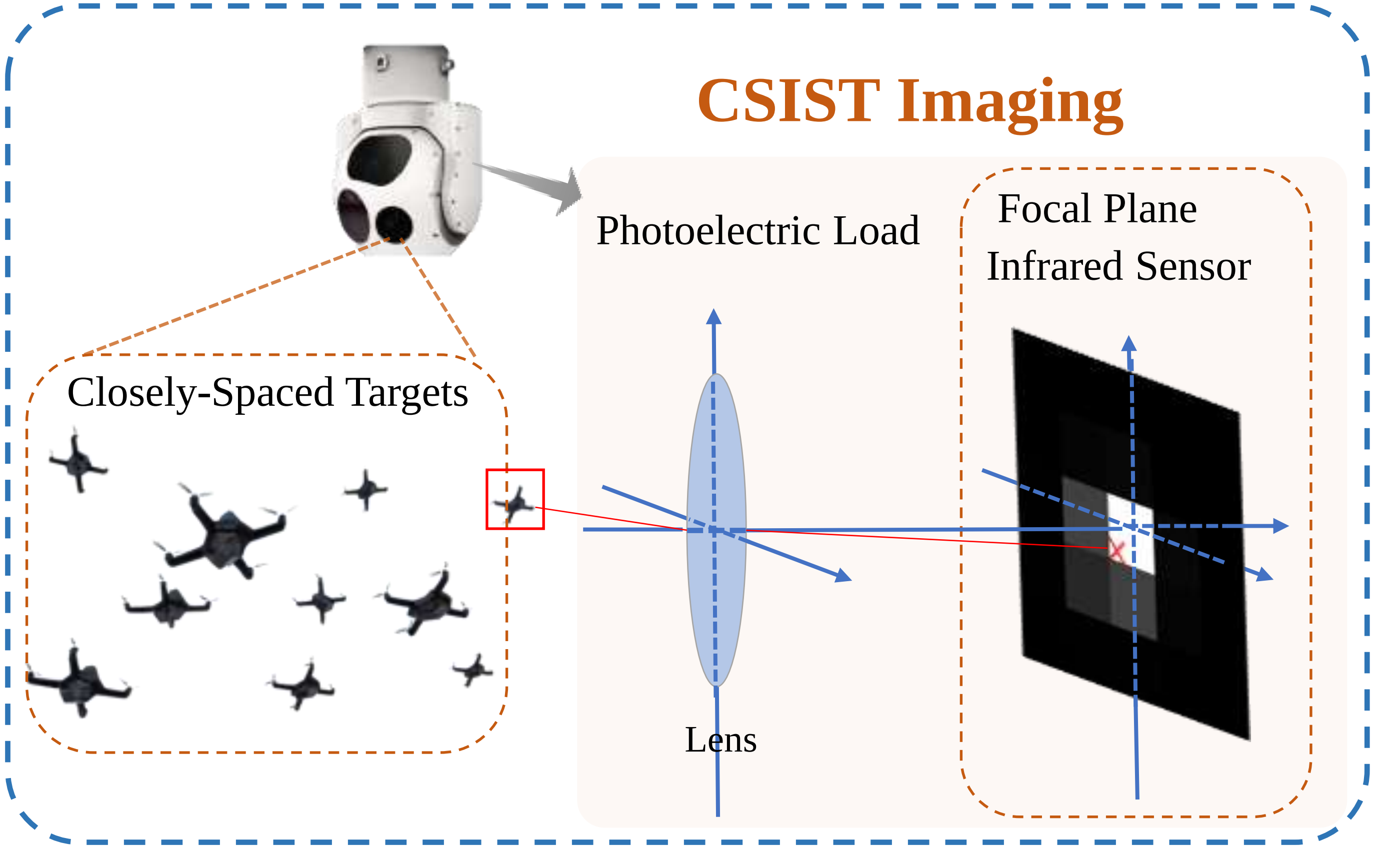


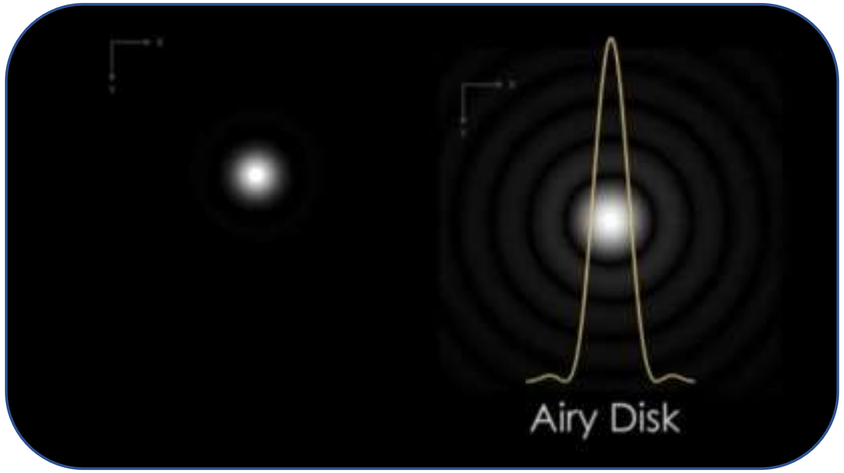
Unmixing Reconstruction

Ecosystem

CSIST Imaging

- The diffraction effect of an optical system causes the energy of a point target to spread over several adjacent pixels. This phenomenon is typically described by a two-dimensional Gaussian Point Spread Function (PSF).



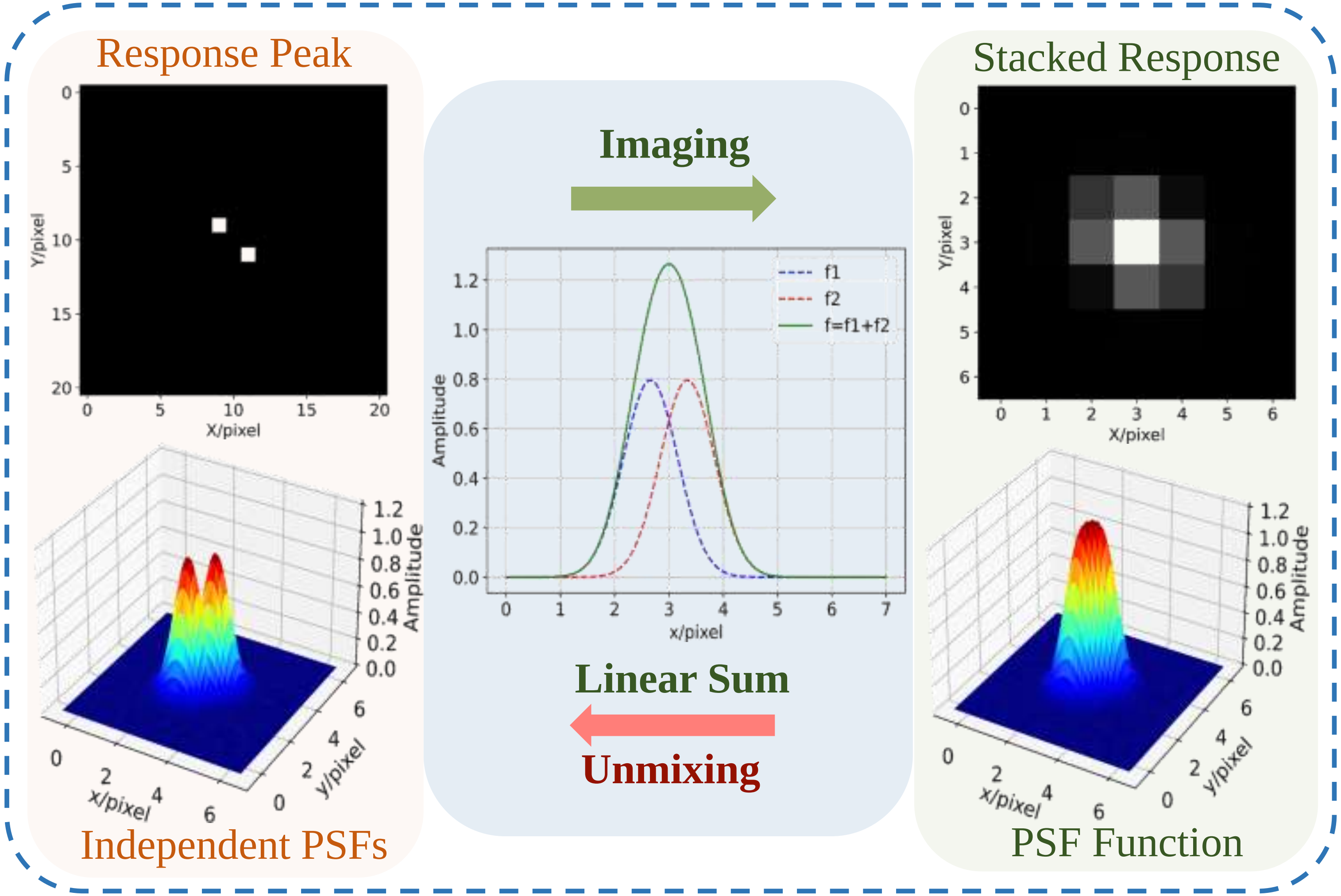


PSF: $$p(x,y)=\frac{1}{2\pi\sigma_{PSF}^2}exp[-\frac{(x-x_t)^2+(y-y_t)^2}{2\sigma_{PSF}^2}]$$

Pixel response: $$g_{i,j}(x_t,y_t)=\int_{x_{i,j}-\frac{d}{2}}^{x_{i,j}+\frac{d}{2}}\int_{y_{i,j}-\frac{d}{2}}^{y_{i,j}+\frac{d}{2}}P(x,y)dxdy$$

CSIST Imaging

- The linear superposition of individual PSFs, which occurs when multiple targets are closely spaced, constitutes the foundational principle for unmixing. The process aims to decompose this composite detector response back into its individual target components.



The image Z formed by N targets in a U×V image plane is:

$$Z = GS + n$$

$$G = [g(u, v, x_1, y_1), \ldots, g(u, v, x_N, y_N)]^{u=1,2,\ldots,u_{max}}_{v=1,2,\ldots,v_{max}}$$

S : Target radiation intensity vector，
n : Gaussian white noise.

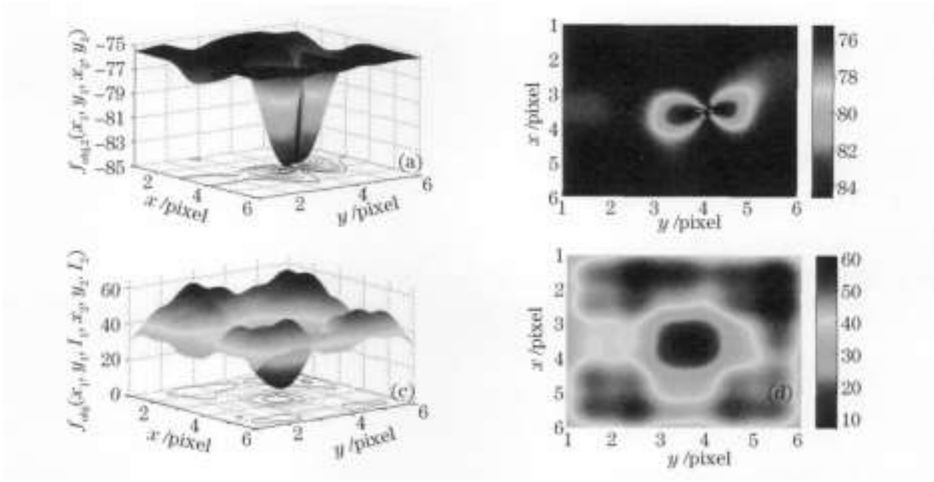
Key Challenge: Sub-1R Spacing
Severe aliasing below the sensor's resolution limit impedes target counting and precise localization.

Relate work

Traditional Optimization Methods

High reliance on hyperparameter tuning.

Poor generalization to varying target scenes.



深度学习应用的挑战

1 Divergence in task objectives

General Image SR:

Enhances image clarity by recovering high-frequency details, mapping a low-resolution image to a high-resolution one.

CSIST Unmixing

Decouples and accurately estimates the number, locations, and intensities of overlapping targets,

2 Lack of an open ecosystem

Key Missing Elements

- **Benchmark Datasets**
- **Customized Evaluation Metrics**
- **Open-Source Toolkits**

Ecosystem

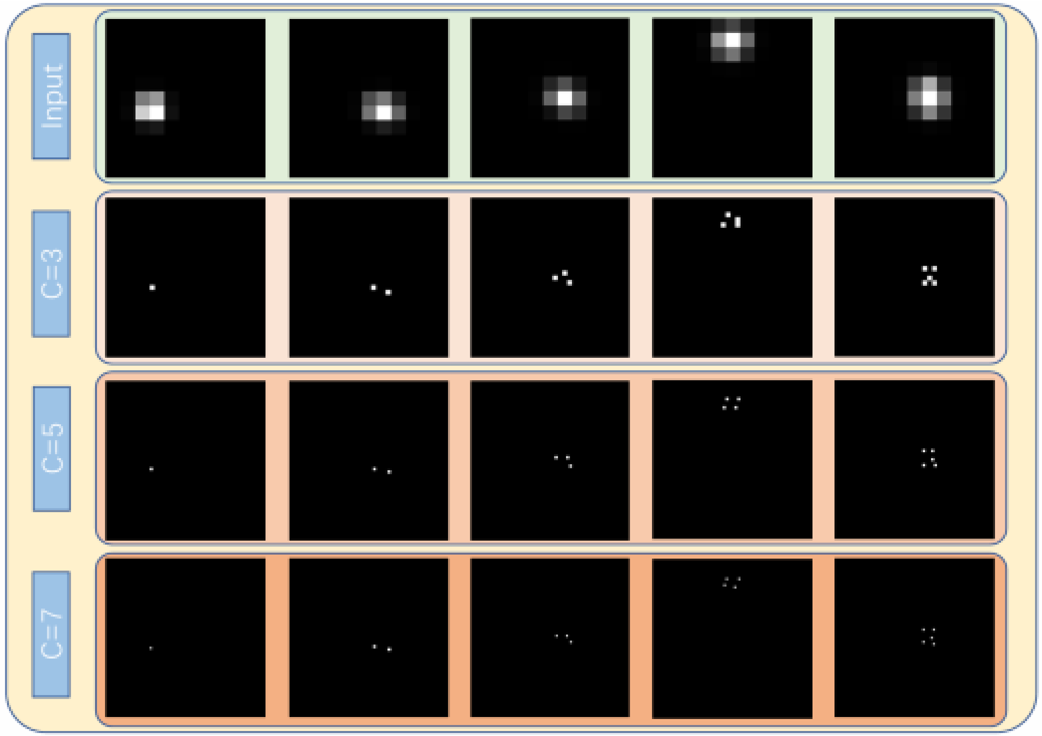
(1) CSIST-100K: Public Benchmark Dataset

(2) CSO-mAP: Dedicated Metric for Sub-Pixel Detection

(3) GrokCSO: Open-Source Toolkit

CSIST-100K

- Based on the optical diffraction effect and the Rayleigh criterion, our simulation generated 100,000 images with annotations, comprising the CSIST-100K dataset.



Simulation Parameters

Parameter	Value/Range
Imaging Size	11×11 pixels
σ_{PSF}	0.5 pixel
Targets per Image	1-5 (random)
Intensity Range	220-250 units
Spatial Constraints	Sub-pixel coordinates within a pixel + 0.52 R separation

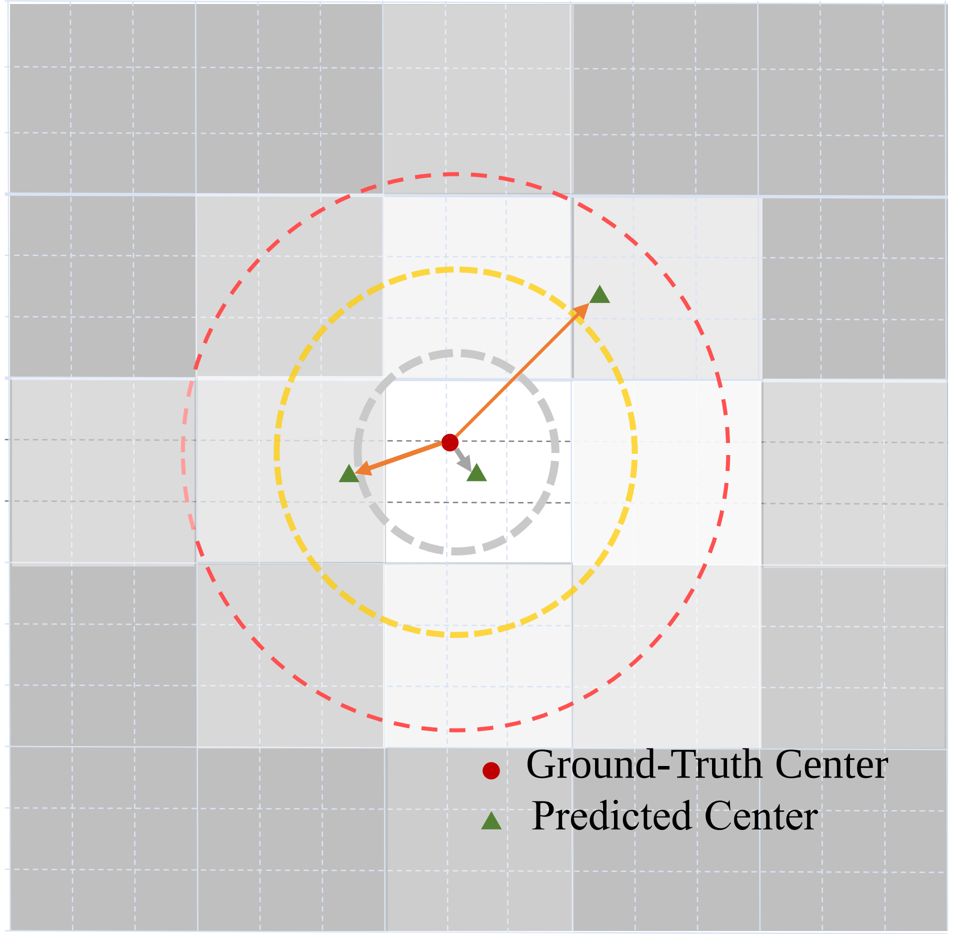
CSO-mAP

Similarity ∝ 1 / (Prediction-GT Distance)

Metric	Definition	Criterion
TP	A correct positive prediction.	distance < δ_k
TN	A correct negative prediction.	-
FP	An incorrect positive prediction.	distance > δ_k
FN	An incorrect negative prediction.	Undetected Targets

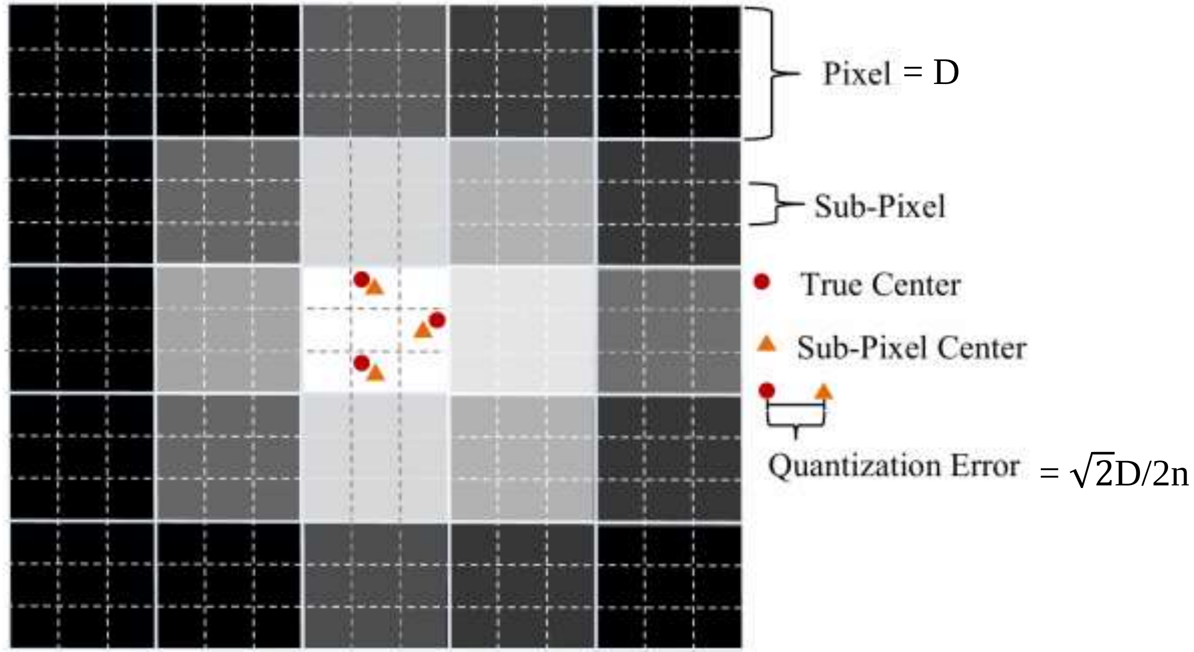
Metric	Definition	Formula
Precision	Fraction of correct positives among all predicted positives.	**TP/(TP+FP)**
Recall	Fraction of correct positives retrieved from all ground truth positives.	**TP/(TP+FN)**
AP	Area under the Precision-Recall curve across all thresholds.	**-**

$\delta_k \in \{0.05, 0.1, 0.15, 0.2, 0.25\}$

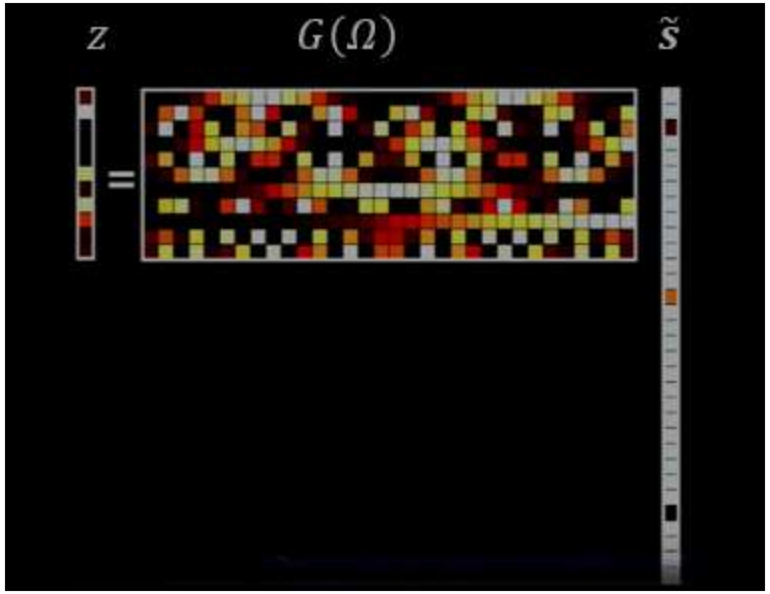


Sparse Reconstruction

- Transform CSIST Unmixing to Sparse Recovery by Sub-Pixel Partitioning



Sub-pixel division



Measurement model: $\min_{\tilde{\mathbf{s}}} ||\mathbf{z} - \mathbf{G}(\Omega)\tilde{\mathbf{s}}||_2^2 + \lambda||\tilde{\mathbf{s}}||_1$

$\Omega = \{(x_l, y_l)\}_{\{l=1,\cdots,L\}}$

ISTA–Net[1]

$$\min_{\tilde{\mathbf{s}}} \|\mathbf{z} - \mathbf{G}(\Omega)\tilde{\mathbf{s}}\|_2^2 + \lambda\|\tilde{\mathbf{s}}\|_1$$

Alternating steps:

Step（1） $\mathbf{r}^{(k)} = \tilde{\mathbf{s}}^{(k-1)} - \rho\mathbf{G}^\top\left(\mathbf{G}\tilde{\mathbf{s}}^{(k-1)} - \mathbf{z}\right)$

Step（2） $\tilde{\mathbf{s}}^{(k)} = \arg\min_{\tilde{\mathbf{s}}} \frac{1}{2}\left\|\tilde{\mathbf{s}} - \mathbf{r}^{(k)}\right\|_2^2 + \lambda\|\mathbf{\Psi}\tilde{\mathbf{s}}\|_1$

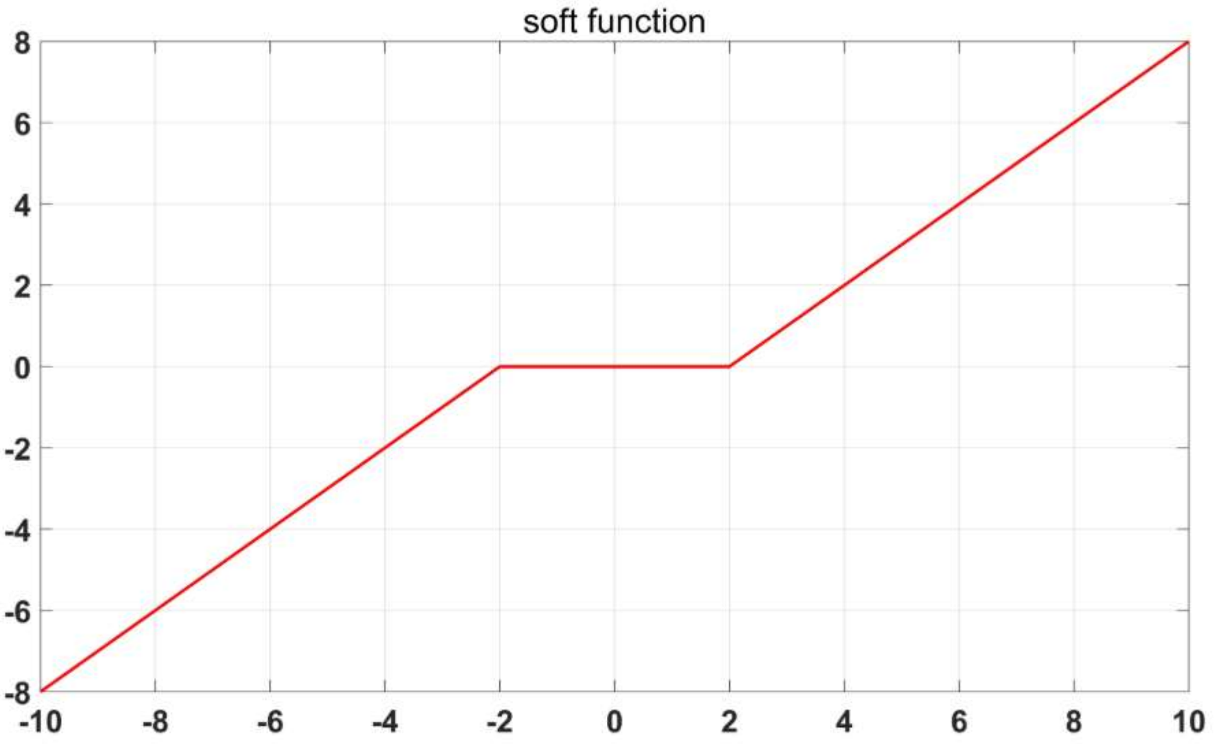
Step（2）solved by:

$$\mathrm{prox}_{\lambda\phi}(\mathbf{r}) = \arg\min_{\tilde{\mathbf{s}}} \frac{1}{2}\|\tilde{\mathbf{s}} - \mathbf{r}\|_2^2 + \lambda\phi(\tilde{\mathbf{s}})$$

Theorem 1 $\|\mathcal{F}(\tilde{\mathbf{s}}) - \mathcal{F}(r)\|_2^2 \approx \alpha\|\tilde{\mathbf{s}} - r\|_2^2$

Replace Ψ with a trainable non-linear transform $\mathcal{F}(\cdot)$

$$\tilde{\mathbf{s}}^{(k)} = \arg\min_{\tilde{\mathbf{s}}} \frac{1}{2}\left\|\mathcal{F}(\tilde{\mathbf{s}}) - \mathcal{F}\left(\mathbf{r}^{(k)}\right)\right\|_2^2 + \theta\|\mathcal{F}(\tilde{\mathbf{s}})\|_1$$



Sparse Representation Optimization: From Convex Optimization to Deep Unfolding
https://www.bilibili.com/video/BV1TV411X7ws/

Zhang J, Ghanem B. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing[C]//CVPR 2018: 1828-1837.

DISTA-Net

DISTA-Net introduces a dynamic mechanism via two key improvements:

Dynamic Transformation Module uses a dual-branch design with dynamic kernels for adaptive feature adjustment, enhancing representational power;

Dynamic Thresholding Module leverages dual convolutional layers to capture multi-scale features, improving robustness against sparse vector perturbations in reconstruction.

Algorithm 1 Dynamic Iterative Shrinkage Thresholding Network

Input: CSIST image $\mathbf{z}$, steering matrix $\mathbf{G}(\mathbf{x},\mathbf{y})$, initial matrix Q_{init}, number of stages N, step size $\{\rho^{(k)}\}_{k=1}^{N}$

Output: Reconstructed result $\tilde{\mathbf{s}}^{(N)}$

Learnable parameters:

$\{\rho^{(k)}\}_{k=1}^{N}, \{\mathrm{DTG}^{(k)}\}_{k=1}^{N}, \{\mathcal{F}_d^{(k)}\}_{k=1}^{N}, \{\tilde{\mathcal{F}}^{(k)}\}_{k=1}^{N}$

$(\tilde{\mathcal{F}}^{(k)} \circ \mathcal{F}_d^{(k)} = \mathbf{I})$

Initialization:

1: $\tilde{\mathbf{s}}^{(0)} \leftarrow Q_{\text{init}}\mathbf{z}$

Iterative reconstruction:

2: **for** $k = 1$ **to** N **do**

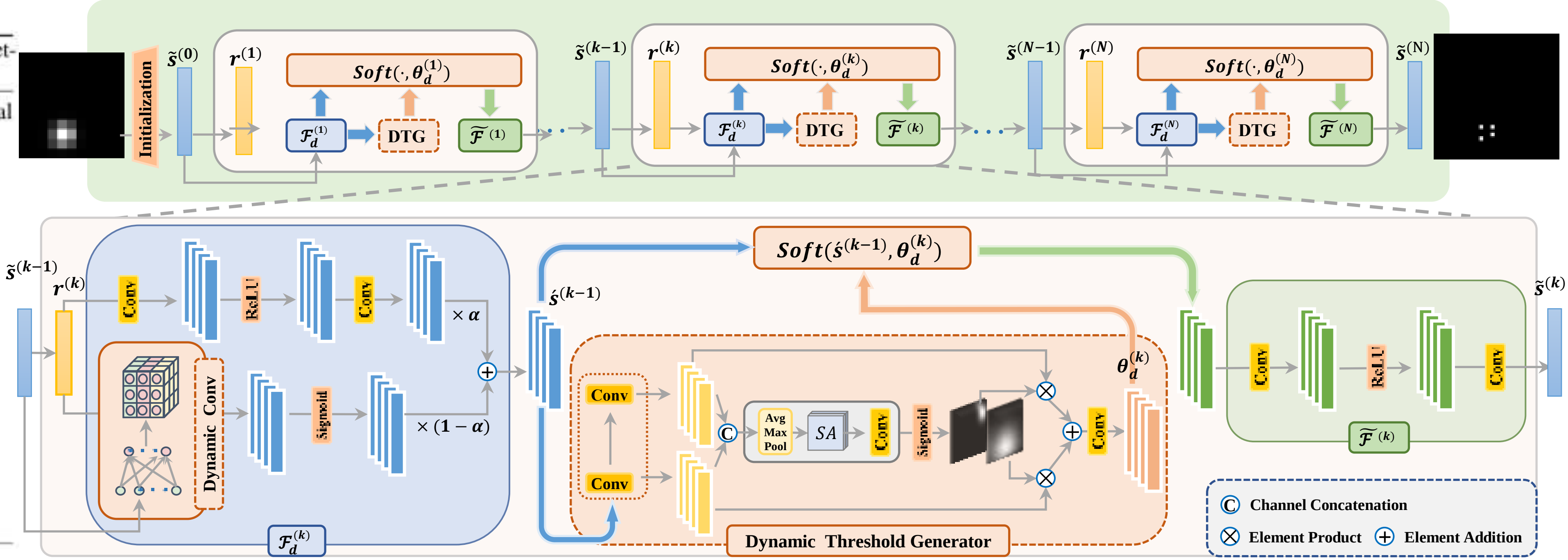
3: $\mathbf{r}^{(k)} \leftarrow \tilde{\mathbf{s}}^{(k-1)} - \rho^{(k)}\mathbf{G}^{\top}(\mathbf{G}\tilde{\mathbf{s}}^{(k-1)} - \mathbf{z})$

4: $\acute{\mathbf{s}}^{(k)} \leftarrow \mathcal{F}_d(\tilde{\mathbf{s}}^{(k-1)}, \mathbf{r}^{(k)})$

5: $\theta_d^{(k)} \leftarrow \mathrm{DTG}^{(k)}(\acute{\mathbf{s}}^{(k)})$

6: $\tilde{\mathbf{s}}^{(k)} \leftarrow \tilde{\mathcal{F}}(\mathrm{Soft}(\mathcal{F}_d(\mathbf{r}^{(k)}), \theta_d^{(k)}))$

7: **end for**



$$Q_{\text{init}} = \arg\min_{Q} \|QZ - S\|_F^2 = SZ^T(ZZ^T)^{-1}, \tilde{s}^{(0)} = Q_{\text{init}}Z.$$

$$\mathcal{L} = \frac{1}{MN_s}\sum_{i=1}^{M}\|\tilde{s}_i^{(N)} - s_i\|_2^2 + \frac{1}{MN_s}\sum_{i=1}^{M}\sum_{k=1}^{N}\|\tilde{\mathcal{F}}^{(k)}(\mathcal{F}_d^{(k)}(s_i)) - s_i\|_2^2.$$

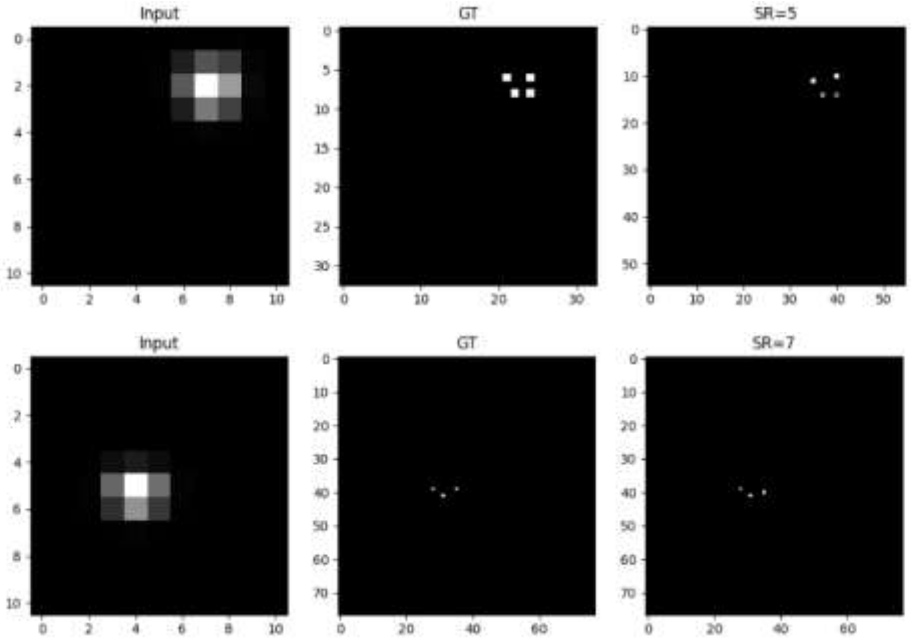
Experiment

Method	#P ↓	FLOPs ↓	CSO-mAP						PSNR ↑	SSIM ↑
			mAP	AP-05	AP-10	AP-15	AP-20	AP-25		
Traditional Optimization										
ISTA [8]	-	-	7.46	0.01	0.31	2.39	9.46	25.14	-	-
Image Super-Resolution										
ACTNet [40]	46.212M	62.80G	45.61	0.38	7.46	41.13	83.12	95.95	35.4526	99.70
CTNet [20]	0.400M	2.756G	45.11	0.38	7.53	40.39	82.11	95.14	35.1499	99.70
DCTLSA [35]	0.865M	13.69G	44.51	0.39	7.35	39.35	81.15	94.34	34.6314	99.65
EDSR [20]	1.552M	12.04G	45.32	0.33	7.07	40.58	83.24	95.41	35.3724	99.71
EGASR [25]	2.897M	17.73G	45.51	0.42	8.03	41.32	85.71	95.08	34.5681	99.66
FENet [1]	0.682M	5.289G	45.67	0.38	7.72	41.50	83.39	95.33	35.1887	99.69
RCAN [38]	1.079M	8.243G	45.87	0.42	7.96	41.81	83.61	95.57	35.2119	99.69
RDN [39]	22.306M	173.0G	45.81	0.35	7.11	41.07	84.07	96.43	36.4686	99.74
SAN [3]	4.442M	34.05G	45.95	0.36	7.35	41.17	84.32	96.57	36.5037	99.74
SRCNN [9]	0.019M	1.345G	29.06	0.23	4.10	21.65	49.95	69.39	28.7608	98.44
SRFBN [18]	0.373M	3.217G	46.05	0.43	8.31	42.83	83.72	94.95	34.0174	99.68
HAN [23]	64.342M	495.0G	45.70	0.39	7.46	40.90	83.61	96.17	35.2703	99.71
Deep Unfolding										
ISTA-Net [37]	0.171M	12.77G	45.16	0.41	7.71	40.57	82.58	94.53	33.9215	99.68
ISTA-Net+ [37]	0.337M	24.33G	46.06	0.42	7.66	41.58	84.46	96.17	36.0892	99.72
LAMP [8]	2.126M	0.278G	14.22	0.05	1.11	7.31	21.56	41.06	27.8299	96.89
LIHT [8]	21.10M	1.358G	10.35	0.06	0.92	4.99	14.74	30.5	27.5107	96.42
LISTA [8]	21.10M	1.358G	30.13	0.25	4.13	22.29	51.18	72.82	29.8936	99.12
FISTA-Net [31]	0.074M	18.96G	44.66	0.45	7.68	39.74	81.24	94.19	35.7519	99.67
TiLISTA [8]	2.126M	0.278G	14.95	0.06	1.23	7.72	22.50	46.23	27.7038	97.40
⋆ **DISTA-Net (Ours)**	2.179M	35.10G	**46.74**	0.38	7.54	42.44	**86.18**	**97.14**	**37.8747**	**99.79**

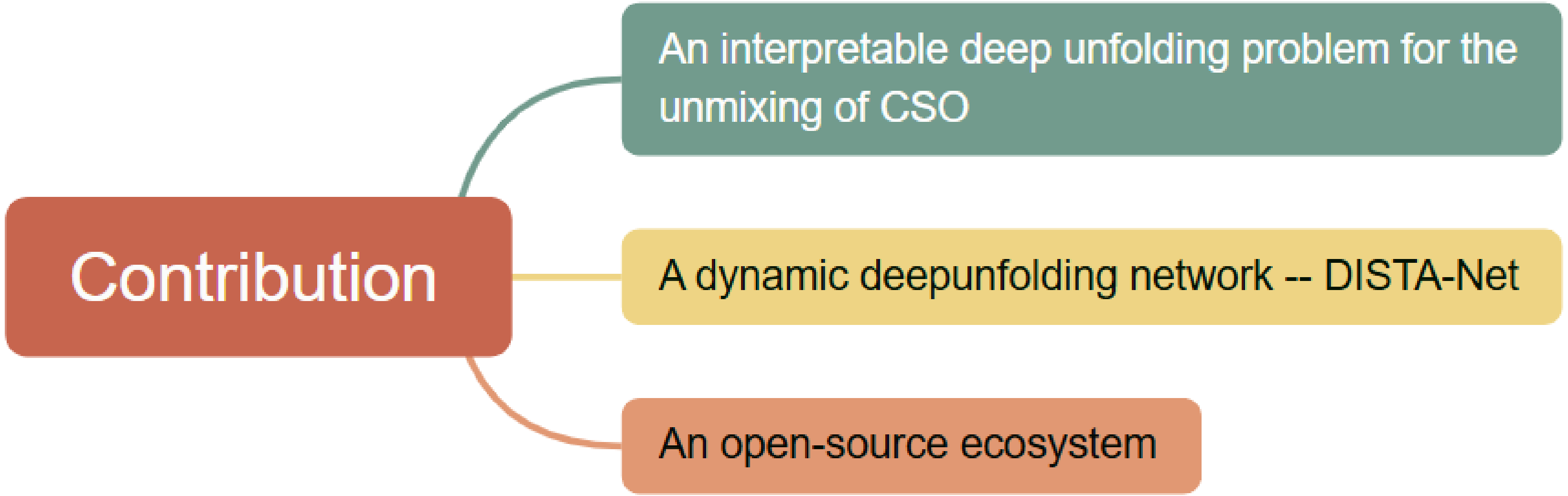
Method	CSO-mAP	AP-05	AP-10	AP-15	AP-20	AP-25
ISTA-Net [37]	45.16	0.41	7.71	40.57	82.58	94.53
DISTA-Net w/o DT	46.32	0.34	6.83	40.76	86.18	**97.50**
DISTA-Net w/o Thres.	46.17	**0.44**	**7.77**	42.18	84.67	95.79
⋆ **DISTA-Net (Ours)**	**46.74**	0.38	7.54	**42.44**	**86.18**	97.14

Method	CSO-mAP	AP-05	AP-10	AP-15	AP-20	AP-25
c=5						
ISTA-Net	66.90	5.60	56.73	87.26	91.55	93.35
ISTA-Net+	68.50	5.80	57.96	89.52	93.78	95.47
CFGN	67.95	**6.59**	58.08	88.35	92.60	94.15
⋆ **DISTA-Net**	**69.58**	6.28	**60.95**	**90.85**	**94.26**	**95.57**
c=7						
ISTA-Net	71.19	20.90	76.45	84.16	86.60	87.85
ISTA-Net+	71.09	18.37	74.90	84.90	87.88	89.31
CFGN	70.38	18.43	73.88	83.97	86.88	88.30
⋆ **DISTA-Net**	**72.84**	**21.32**	**78.47**	**86.09**	**88.55**	**89.76**

Visualization



Summary



arXiv https://arxiv.org/abs/2505.19148
code https://github.com/GrokCV/GrokCSO