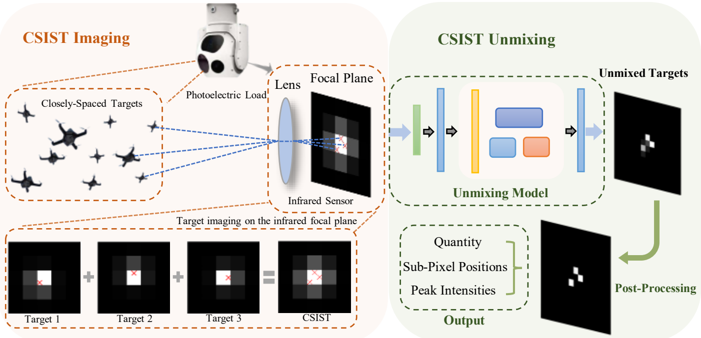
DISTA-Net: Dynamic Closely-Spaced Infrared Small Target Unmixing

Shengdong Han[1]* Shangdong Yang[1]* Yuxuan Li[2] Xin Zhang[2] Xiang Li[2,3] Jian Yang[2] Ming-Ming Cheng[2,3] Yimian Dai[2,3]†

[1]School of Computer Science, Nanjing University of Posts and Telecommunications [2]VCIP, CS, Nankai University [3]NKIARI, Futian, Shenzhen

Background.



CSIST unmixing, a key follow-up to infrared small target detection, remains underdeveloped due to:

1. Lack of public datasets;
2. Absence of tailored evaluation metrics;
3. Shortage of open-source toolkits.

Sparse Reconstruction.

Imaging Principle:

Distant point targets → Diffraction spreads energy → Modeled by 2D Gaussian PSF:

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

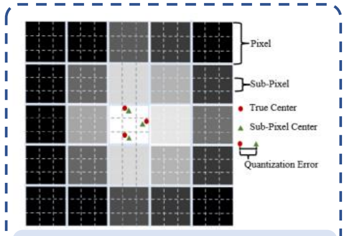
Each pixel integrates PSF over $U\times V$ array:

$$g_{i,j}(x_t,y_t)=\int_{x_{i,j}-1/2D}^{x_{i,j}+1/2D}\int_{y_{i,j}-1/2D}^{y_{i,j}+1/2D}p(x,y)dxdy, (2)$$

Measurement Model:

Vectorized form:

$$z=[g^c(x_1,y_1)g^c(x_2,y_2)\cdots g^c(x_K,y_K)]s+n=G(x,y)s+n, (3)$$



$n\times n$ grid per pixel → $L=UVn^2$ sub-pixels

Target positions → finite set $\Omega=\{(x_l,y_l)\}_{l=1,\cdots,L}$ (sparse subset)

Sufficient resolution → max deviation $\sqrt{2}D/2n$

Measurement model extends to Ω:

$$z=G(\Omega)\tilde{s}+w, (4)$$

Eq. (4) can be formalized with ℓ_1 regularization:

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

solution via iterative shrinkage-thresholding algorithm:

$$r^{(k)}=\tilde{s}^{(k-1)}-\rho G^\top(G\tilde{s}^{(k-1)}-z),$$

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

From ISTA to Deep Unfolding Network:

CSIST unmixing
→ sparse recovery problem
ISTA-Net: Replaces fixed transforms with learnable ones
Limitation: Static weights
→ poor adaptability in complex scenarios

Algorithm 1 DISTA-Net

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:

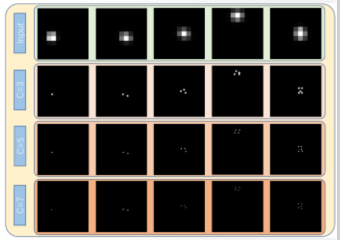
2: **for** $k=1$ **to** N **do**
3: $\quad\mathbf{r}^{(k)}\leftarrow\tilde{\mathbf{s}}^{(k-1)}-\rho^{(k)}\mathbf{G}^\top(\mathbf{G}\tilde{\mathbf{s}}^{(k-1)}-\mathbf{z})$
4: $\quad\acute{\mathbf{s}}^{(k)}\leftarrow\mathcal{F}_d(\tilde{\mathbf{s}}^{(k-1)},\mathbf{r}^{(k)})$
5: $\quad\theta_d^{(k)}\leftarrow\mathrm{DTG}^{(k)}(\acute{\mathbf{s}}^{(k)})$
6: $\quad\tilde{\mathbf{s}}^{(k)}\leftarrow\tilde{\mathcal{F}}(\mathrm{Soft}(\mathcal{F}_d(\mathbf{r}^{(k)}),\theta_d^{(k)}))$
7: **end for**

Results.

Method	#P ↓	FLOPs ↓	CSO-mAP mAP	AP-05	AP-10	AP-15	AP-20	AP-25	PSNR ↑	SSIM ↑
Traditional Optimization										
ISTA [7]	-	-	7.46	0.01	0.31	2.39	9.46	25.14	-	-
Image Super-Resolution										
ACTNet [45]	46.212M	62.80G	45.61	0.38	7.46	41.13	83.12	95.95	35.4526	99.70
CTNet [30]	0.400M	2.756G	45.11	0.38	7.53	40.39	82.11	95.14	35.1499	99.70
DCTLSA [38]	0.865M	13.69G	44.51	0.39	7.35	39.35	81.15	94.34	34.6314	99.65
EDSR [19]	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	<u>85.71</u>	95.08	34.5681	99.66
FeNet [31]	0.348M	2.578G	45.77	0.42	<u>8.19</u>	42.13	83.30	94.80	34.1531	99.66
RCAN [43]	1.079M	8.243G	45.87	0.42	7.96	41.81	83.61	95.57	35.2119	99.69
RDN [44]	22.306M	173.0G	45.81	0.35	7.11	41.07	84.07	96.43	36.4686	99.74
SAN [2]	4.442M	34.05G	45.95	0.36	7.35	41.17	84.32	<u>96.57</u>	<u>36.5037</u>	<u>99.74</u>
SRCNN [8]	0.019M	1.345G	29.06	0.23	4.10	21.65	49.95	69.39	28.7608	98.44
SRFBN [17]	0.373M	3.217G	46.05	<u>0.43</u>	**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
HiT-SNG [42]	0.952M	13.324G	45.01	0.39	7.34	40.19	81.98	95.17	35.1390	99.71
Deep Unfolding										
ISTA-Net [40]	0.171M	12.77G	45.16	0.41	7.71	40.57	82.58	94.53	33.9215	99.68
ISTA-Net+ [40]	0.337M	24.33G	<u>46.06</u>	0.42	7.66	41.58	84.46	96.17	36.0892	99.72
LAMP [7]	2.126M	0.278G	14.22	0.05	1.11	7.31	21.56	41.06	27.8299	96.89
LIHT [7]	21.10M	1.358G	10.35	0.06	0.92	4.99	14.74	30.5	27.5107	96.42
LISTA [7]	21.10M	1.358G	30.13	0.25	4.13	22.29	51.18	72.82	29.8936	99.12
FISTA-Net [34]	0.074M	18.96G	44.66	**0.45**	7.68	39.74	81.24	94.19	35.7519	99.67
TiLISTA [7]	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	<u>42.44</u>	**86.18**	**97.14**	**37.8747**	**99.79**

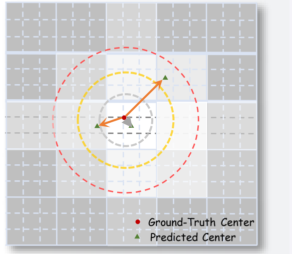
Ecosystem.

1 Dataset: CSIST-100K



- σ_{PSF}=0.5 pixel
- 1–5 overlapping targets
- Single-pixel centered
- Intensity: 220–250 (8-bit)
- Spacing ≥ 0.52R in a pixel
- Random in 11×11 space
- 100K samples (80K/10K/10K)

2 Metric: CSO-mAP



CSO-Aware Matching Criterion

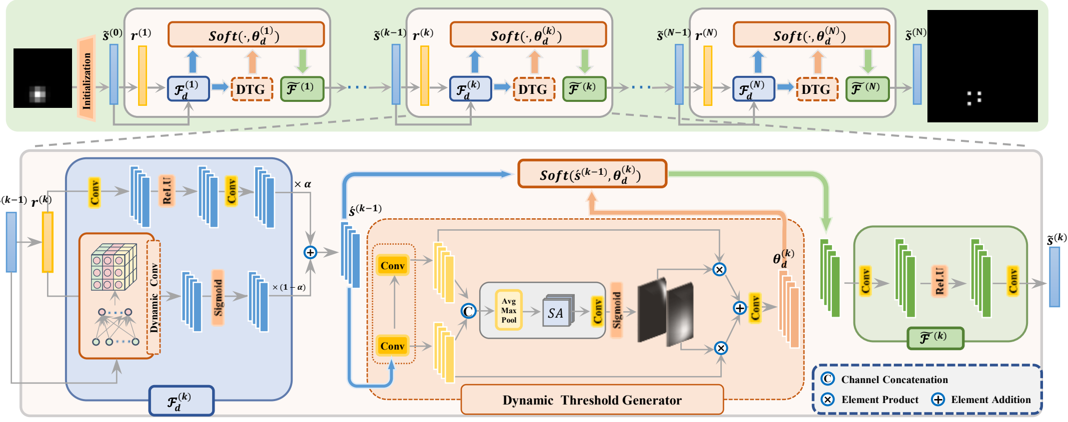
A predicted point is considered a true positive only if:

- It corresponds to a GT target
- That GT has not been matched with any other prediction exhibiting higher intensity

3 toolkit: GrokCSO

- **Pre-trained models & reproducibility**: Training scripts, logs, and model zoo
- **Tailored flexibility**: Adaptable backbones/necks, dataset loaders, attention modules
- **Specialized evaluation**: Metrics for closely-spaced targets

Method.



1. Dynamic Transformation Module $F_d(\cdot)$:

Employs a dual-branch structure with dynamic convolution kernels to adaptively enhance feature representation.

2. Dynamic Thresholding Module Soft$(\cdot,\Theta_d)$:

Incorporates dual convolutional layers to capture multi-scale features, significantly improving robustness to sparse vector perturbations during image reconstruction.

More.

paper code team wechat