ICJK
Chinese Journal of Information Fusion
http://dx.doi.org/10.62762/CJIF.2026.903328

RESEARCH ARTICLE

DA-DISTA-Net: Degradation-Aware Unfolding Network for Closely-Spaced Infrared Small Target Unmixing

Wenjun Yang[1], **Yaxiu Chen**[1], **Yang Wang**[1], **Bo Lv**[1], **Jungang Yang**[1,*] **and Jiangwei Zou**[1]

[1] College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China

Abstract

Deep unfolding networks for closely-spaced infrared small target (CSIST) unmixing are usually trained and tested with a single, known point spread function (PSF). When the PSF at test time differs from the training PSF, their accuracy drops noticeably. We treat unmixing under an unknown PSF as an information fusion problem: the low-resolution observation contains both the target configuration and cues about the blur that produced it, and the second kind of information can be used to steer the reconstruction. Based on this idea, we propose DA-DISTA-Net, a degradation-aware unfolding network. A lightweight Global Degradation Estimator (GDE) maps the low-resolution input to a compact degradation embedding once, before unfolding starts. At every stage, the embedding drives a Degradation-Aware Step Modulator (DASM) and a Degradation-Aware Threshold Modulator (DATM), which adjust the gradient step size and the soft-thresholding level, respectively. No PSF labels are needed. For evaluation, we re-render the images of CSIST-100K from their original annotations with a PSF standard deviation σ drawn uniformly from $[0.2, 0.9]$ pixels for each image. On this random-degradation test set, DA-DISTA-Net reaches a mean average precision (mAP) of 43.29%, which is 0.78 percentage points higher than the strongest baseline, at the cost of a 3.2% increase in parameters.

Keywords: infrared small target detection, closely-spaced target unmixing, deep unfolding network, sparse reconstruction, degradation-aware inference, information fusion.

1 Introduction

Closely-spaced infrared small target (CSIST) unmixing is an important and difficult task in infrared imaging. Because of optical diffraction and the limited resolution of infrared focal plane arrays (IRFPAs), distant targets appear on the detector as small, texture-free blobs that are well approximated by Gaussians. When several targets are close to each other, their point spread function (PSF) responses overlap and form a single energy cluster that conventional detectors cannot reliably split. The task can be written as a sparse recovery problem: given a low-resolution observation $\mathbf{y} = \mathbf{\Phi x} + \mathbf{n}$, we want to recover the sub-pixel positions and intensities of all targets at the same time.

Submitted: 12 January 2026
Revised: 18 September 2026
Accepted: 22 September 2026
Published: 19 November 2026

Vol. 3, **No.** 4, 2026.
10.62762/CJIF.2026.903328

***Corresponding author:**
✉ Jungang Yang
yangjungang@nudt.edu.cn

Citation
Yang, W., Chen, Y., Wang, Y., Lv, B., Yang, J., & Zou, J. (2026). DA-DISTA-Net: Degradation-Aware Unfolding Network for Closely-Spaced Infrared Small Target Unmixing. *Chinese Journal of Information Fusion*, 3(4), 251–261.

© 2026 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

Optimization-based solvers such as the iterative shrinkage-thresholding algorithm (ISTA) [1] and its accelerated variant FISTA [2] handle this recovery through ℓ_1-regularized sparse reconstruction, but they still converge slowly and need careful manual tuning of hyperparameters. Deep unfolding methods, such as ISTA-Net [3] and DISTA-Net [4], unroll the ISTA iterations into a trainable network and thus combine the interpretability of the model with the efficiency of learning. DISTA-Net adds dynamic convolution weights and dynamic soft-thresholding, and currently gives the best unmixing accuracy on the CSIST-100K benchmark [4].

These methods share one assumption that is rarely questioned: the PSF parameter σ is fixed and known at test time. In real systems, the PSF changes with atmospheric turbulence, imaging distance, and focal-plane temperature. To measure the effect of this assumption, we build a test set in which the blur standard deviation σ is drawn uniformly from $[0.2, 0.9]$ and compare it with the fixed-degradation benchmark ($\sigma = 0.5$). The CSO-mAP [4] of ISTA falls from 7.46% to 4.71% (a relative drop of 36.9%), and that of DISTA-Net falls from 46.82% to 42.51% (a relative drop of 9.2%). Methods that rely on a fixed PSF are therefore hard to deploy in scenes where the PSF is unknown and changes over time.

In our view, robust CSIST unmixing under an unknown PSF is an information fusion problem. Besides the spatial layout of the targets, the low-resolution observation carries implicit information about how it was degraded, and the reconstruction should use both to adapt to the actual PSF. In image restoration, conditioning the reconstruction on degradation information, either given explicitly as in USRNet [5] or estimated from the measurement as in DAUHST [6], has been shown to improve robustness to changing degradations. To our knowledge, this idea has not yet been used for CSIST unmixing.

Following this line, we propose DA-DISTA-Net (Degradation-Aware Dynamic ISTA Unfolding Network). A lightweight Global Degradation Estimator (GDE) extracts a compact degradation embedding $\mathbf{d}$ from the low-resolution observation $\mathbf{y}$ before the unfolding iterations. The embedding is then fed to every stage through a Degradation-Aware Step Modulator (DASM) and a Degradation-Aware Threshold Modulator (DATM), so that the network can handle unseen degradation levels without explicit PSF labels.

The main contributions of this paper are as follows:

- **Degradation-aware unfolding architecture.** We design DA-DISTA-Net, in which a degradation embedding estimated from the observation itself is fused into the sparse recovery at every unfolding stage to calibrate both the step size and the threshold.
- **Random-degradation dataset.** We re-render the CSIST-100K images from their original annotations with $\sigma \sim U(0.2, 0.9)$, using an exact pixel-integration imaging model. The dataset allows different methods to be compared fairly under unknown degradation.
- **Accuracy and interpretability.** On the new dataset, DA-DISTA-Net obtains an mAP of 43.29%, 0.78 percentage points above DISTA-Net, with only 0.07 M extra parameters (3.2%). A t-SNE visualization suggests that the learned embeddings vary systematically with PSF blur strength.

2 Related Work

2.1 Closely-Spaced Infrared Small Target Unmixing

CSIST unmixing differs from conventional infrared small target detection [7, 8]. The main difficulty is not to separate targets from a cluttered background, but to split overlapping PSF responses into individual sub-pixel targets. Model-based approaches formulate this as ℓ_1-regularized sparse recovery solved by ISTA [1, 2], where the observation matrix $\mathbf{\Phi}$ is built analytically from a Gaussian PSF model. These methods are physically interpretable, but they converge slowly and are sensitive to the regularization parameter λ, which has to be tuned for each scene. Learned approximate message passing (LAMP) [9] and LISTA [10] bring data-driven learning into iterative solvers, but they are not designed for the structured observation model of CSIST. DISTA-Net [4] is the first end-to-end deep unfolding method for this task; it uses dynamic convolution and adaptive thresholding based on large selective kernel (LSK) attention [11] in each ISTA iteration. However, its observation matrix $\mathbf{\Phi}$ is built for a single PSF, which makes the network sensitive to PSF changes.

2.2 Deep Unfolding for Sparse Recovery

Deep unfolding was introduced by Gregor and LeCun [10] with LISTA, which turns a fixed number of sparse-coding iterations into network layers with learnable parameters. ISTA-Net [3] applies this idea to image compressive sensing (CS): it replaces the

hand-crafted sparsifying transform with a learned convolutional transform, so that the network can be read as truncated proximal gradient descent with learned transforms. ISTA-Net++ [12] predicts the step size ρ and the noise level σ of each iteration with a condition module whose input is the CS ratio γ, and shows that conditioning per-layer hyperparameters on a global descriptor improves performance across sampling ratios. Chen et al. [13] proved linear convergence of unfolded ISTA under a weight-coupling structure, and ALISTA [14] further showed that the weight matrices can be computed analytically, so that only step sizes and thresholds need to be learned. Hybrid ISTA [15] inserts free-form deep networks into unfolded ISTA while keeping convergence guarantees. USRNet [5] unfolds half-quadratic splitting (HQS) for single image super-resolution, and its hyper-parameter module predicts the per-iteration coefficients α_k and β_k from the blur kernel and the noise level; this conditioning clearly outperforms fixed parameters. In all of these methods, however, the degradation is either fixed or given at inference time. In CSIST unmixing, the PSF parameters are unknown and differ from one observation to another.

2.3 Degradation-Aware Methods

Real images seldom follow one known degradation model, and many restoration methods therefore estimate or embed degradation information. DASR [16] learns degradation representations for blind super-resolution without supervision, encoding different blur and noise conditions into compact latent vectors without degradation labels. DAUHST [6] brings this idea to hyperspectral compressive imaging: it estimates degradation parameters from the compressed measurement and the physical mask and injects them into every unfolding iteration to control the hyperparameters that depend on the ill-posedness. KDSR [17] trains a teacher that extracts an implicit degradation representation from paired high- and low-resolution images and distills it into a student that only sees the low-resolution input. These studies indicate that a degradation embedding estimated from the corrupted observation, rather than a ground-truth label, can effectively control the restoration. This has not been explored for CSIST unmixing, although PSF variation is a serious problem there. In this work, a lightweight GDE infers a PSF embedding from the low-resolution observation, and step and threshold modulators conditioned on this embedding are used in every unfolding stage.

3 Methodology

3.1 Problem Formulation

CSIST unmixing is modeled as a linear sparse recovery problem. Let $\mathbf{y} \in \mathbb{R}^M$ be the low-resolution observation of a local infrared patch that contains several closely spaced targets. The imaging process is

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^N$ ($N = c^2M$) is the target distribution on a sub-pixel grid refined c times in each direction, $\mathbf{\Phi} \in \mathbb{R}^{M\times N}$ is the observation matrix obtained by integrating a Gaussian PSF over the photosensitive area of each pixel, and $\mathbf{n}$ is additive Gaussian noise. Since $N = c^2M > M$ (e.g., $N = 9M$ for $c = 3$), the system is underdetermined. Using the sparsity of point targets, the reconstruction is formulated as the ℓ_1-regularized problem

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{\Phi}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{\Psi}\mathbf{x}\|_1, \tag{2}$$

where $\frac{1}{2}\|\mathbf{\Phi}\mathbf{x} - \mathbf{y}\|_2^2$ is the data fidelity term, $\lambda\|\mathbf{\Psi}\mathbf{x}\|_1$ is the regularization term that enforces sparsity in the transform domain $\mathbf{\Psi}$, and λ is the regularization parameter.

ISTA solves Eq. (2) by alternating a gradient descent step and a proximal mapping:

$$\mathbf{r}^{(k)} = \mathbf{x}^{(k-1)} - \rho\,\mathbf{\Phi}^\top\left(\mathbf{\Phi}\mathbf{x}^{(k-1)} - \mathbf{y}\right), \tag{3}$$

$$\mathbf{x}^{(k)} = \arg\min_{\mathbf{x}} \frac{1}{2}\left\|\mathbf{x} - \mathbf{r}^{(k)}\right\|_2^2 + \lambda\|\mathbf{\Psi}\mathbf{x}\|_1, \tag{4}$$

where k is the iteration index and ρ is the step size. Eq. (4) is a special case of the proximal mapping

$$\mathrm{prox}_{\lambda\phi}(\mathbf{r}) = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x} - \mathbf{r}\|_2^2 + \lambda\phi(\mathbf{x}) \tag{5}$$

with $\phi(\mathbf{x}) = \|\mathbf{\Psi}\mathbf{x}\|_1$.

ISTA needs manual parameter tuning and converges slowly. ISTA-Net [3] replaces the transform $\mathbf{\Psi}$ with a learnable network $\mathcal{F}(\cdot)$:

$$\mathbf{x}^{(k)} = \arg\min_{\mathbf{x}} \frac{1}{2}\left\|\mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{r}^{(k)})\right\|_2^2 + \theta\|\mathcal{F}(\mathbf{x})\|_1. \tag{6}$$

With the ℓ_1 regularizer, this problem has the well-known closed-form solution

$$\mathcal{F}(\mathbf{x}^{(k)}) = \mathrm{soft}\left(\mathcal{F}(\mathbf{r}^{(k)}), \theta\right), \tag{7}$$

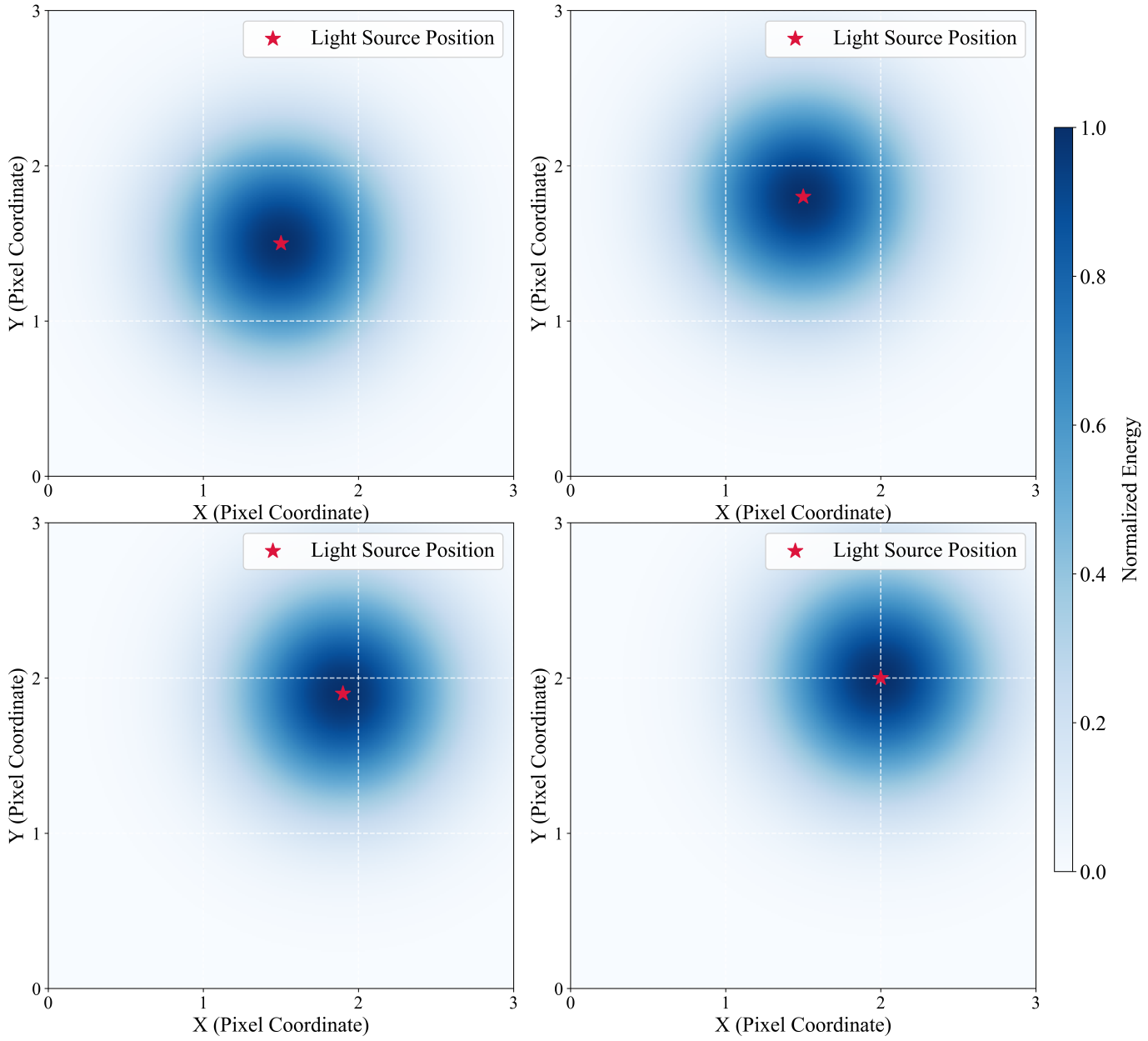


Figure 1. Pixel-integrated response of a continuous Gaussian PSF for different sub-pixel target positions.

and, with a learned left inverse $\tilde{\mathcal{F}}(\cdot)$ of $\mathcal{F}(\cdot)$,

$$\mathbf{x}^{(k)} = \tilde{\mathcal{F}}\Big(\mathrm{soft}\Big(\mathcal{F}(\mathbf{r}^{(k)}), \theta\Big)\Big). \tag{8}$$

DISTA-Net [4] further makes the convolutions and the thresholds dynamic, so that the transform weights and the soft-thresholding level depend on the input. Both methods, however, are formulated for a single fixed PSF and do not take into account inputs with different degradations; their accuracy therefore decreases when the actual degradation changes.

3.2 Random-Degradation Dataset Construction

Each image of CSIST-100K is re-rendered from its original annotation with a randomly chosen PSF. For each sample, σ (in low-resolution pixels) is drawn independently from $U(0.2, 0.9)$, and the low-resolution observation is synthesized with an exact pixel-integration imaging model:

$$y_{(i,j)} = \sum_{p} a_p \iint_{\Omega_{i,j}} G(\xi - u_p,\, \eta - v_p; \sigma)\, \mathrm{d}\xi\, \mathrm{d}\eta, \tag{9}$$

where (u_p, v_p) and a_p are the sub-pixel coordinates and the radiance of the p-th target, $\Omega_{i,j}$ is the photosensitive area of pixel (i, j), and $G(\cdot; \sigma)$ is a 2D Gaussian PSF. The integral is needed because, when a target is not at the center of a pixel, the energy collected by each pixel differs clearly from a Gaussian sampled at pixel centers, as shown in Figure 1. Zero-mean additive Gaussian

noise with a standard deviation of 5 gray levels (8-bit scale) is added to each observation after PSF formation; the values are then clipped and quantized to $[0, 255]$ and stored as PNG images. The same noise model is used for the training, validation, and test sets. The dataset contains 100,000 samples, split into 80,000 for training, 10,000 for validation, and 10,000 for testing.

Baseline experiments confirm the limitation of the fixed-degradation assumption: DISTA-Net trained under the fixed-σ assumption drops from 46.82% to 42.51% mAP when evaluated on the random-σ test set. This drop motivates the degradation-aware design described next.

3.3 DA-DISTA-Net

The proposed DA-DISTA-Net places a lightweight GDE in front of all unfolding iterations to deal with PSF variation. The degradation embedding $\mathbf{d}$ produced by the GDE is passed to two modules that are present in every unfolding stage: DASM and DATM. The overall architecture is shown in Figure 2.

3.3.1 Global Degradation Estimator (GDE)

Before unfolding starts, the low-resolution observation $\mathbf{y}$ (in its 2D form) is passed through a lightweight encoder:

$$\begin{aligned}\mathbf{d} &= \mathrm{GDE}(\mathbf{y}) \\ &= \mathrm{FC}(\mathrm{GAP}(\mathrm{Conv}_2(\mathrm{ReLU}(\mathrm{Conv}_1(\mathbf{y}))))) ,\end{aligned} \tag{10}$$

where Conv_1 and Conv_2 are 3×3 convolutional layers, ReLU is the rectified linear unit, GAP is global average pooling, and FC is a fully connected (linear) layer that projects the pooled features to a C_{deg}-dimensional degradation embedding. The GDE is run only once before the K-stage loop, and $\mathbf{d}$ is reused in all stages. No ground-truth σ is needed for training or inference: $\mathbf{d}$ is learned only through the reconstruction loss back-propagated from the final output.

3.3.2 Degradation-Aware Step Modulator (DASM)

In classical ISTA, the step size ρ is one scalar shared by all inputs and iterations. This is not ideal when the PSF blur varies. A strongly blurred observation leads to a worse-conditioned inverse problem, for which a large ρ may overshoot or diverge, whereas a sharp PSF allows a larger ρ and faster convergence. DASM therefore predicts a scalar correction from the degradation embedding:

$$\begin{aligned}\delta^{(k)} &= \mathrm{DASM}^{(k)}(\mathbf{d}) \\ &= \tanh(\mathrm{FC}_2(\mathrm{ReLU}(\mathrm{FC}_1(\mathbf{d})))) .\end{aligned} \tag{11}$$

The step size of the k-th stage is

$$\rho^{(k)} = \rho_0 \left(1 + \alpha\, \delta^{(k)}\right), \tag{12}$$

where ρ_0 is the base step size and $\alpha = 0.1$ is a fixed scaling factor. Since $\delta^{(k)} \in (-1, 1)$, $\rho^{(k)}$ stays within $\pm 10\%$ of ρ_0, so DASM only applies a small residual correction to the nominal step. The gradient update of the k-th stage becomes

$$\mathbf{r}^{(k)} = \mathbf{x}^{(k-1)} - \rho^{(k)} \mathbf{\Phi}^{\top} \left(\mathbf{\Phi}\mathbf{x}^{(k-1)} - \mathbf{y}\right). \tag{13}$$

3.3.3 Degradation-Aware Threshold Modulator (DATM)

In DISTA-Net, the soft-thresholding level is produced by an LSK attention module applied to the sparse feature map $\mathbf{z} = \mathcal{F}(\mathbf{r}^{(k)})$. This channel-wise threshold follows local feature activations but does not know the global PSF blur. Under strong blur (large σ), the response of a weak target spreads over neighboring pixels and its per-channel activation becomes smaller; a threshold suited to a sharp PSF may then remove these weak responses. DATM addresses this with a channel-wise multiplicative scale computed from the degradation embedding:

$$\begin{aligned}\mathbf{s}^{(k)} &= \mathrm{DATM}^{(k)}(\mathbf{d}) \\ &= \mathrm{Sigmoid}(\mathrm{FC}_2(\mathrm{ReLU}(\mathrm{FC}_1(\mathbf{d})))) .\end{aligned} \tag{14}$$

We use channel-wise rather than spatial modulation because the PSF parameter is shared by the whole observation patch, so spatially uniform coefficients agree with the degradation model. This also keeps the module small: each stage produces a 64-dimensional scale vector instead of a full spatial map.

The modulated threshold and the resulting proximal step are

$$\mathbf{\Theta}^{(k)} = \mathrm{LSK}(\mathbf{z}) \odot \mathbf{s}^{(k)}, \tag{15}$$

$$\mathbf{x}^{(k)} = \tilde{\mathcal{F}}\Big(\mathrm{sign}(\mathbf{z}) \odot \mathrm{ReLU}\Big(|\mathbf{z}| - \mathbf{\Theta}^{(k)}\Big)\Big), \tag{16}$$

where $\odot$ denotes element-wise multiplication (with $\mathbf{s}^{(k)}$ broadcast over spatial positions). Because $\mathbf{s}^{(k)} \in (0, 1)^{64}$, DATM can only lower the LSK threshold. The intended behavior is that, for large σ, $\mathbf{d}$ encodes strong blur and DATM outputs values of $\mathbf{s}^{(k)}$ well below one, which lowers the effective threshold and keeps weak target responses; for small σ (sharp PSF), $\mathbf{s}^{(k)}$ approaches one and the LSK threshold is left almost unchanged.

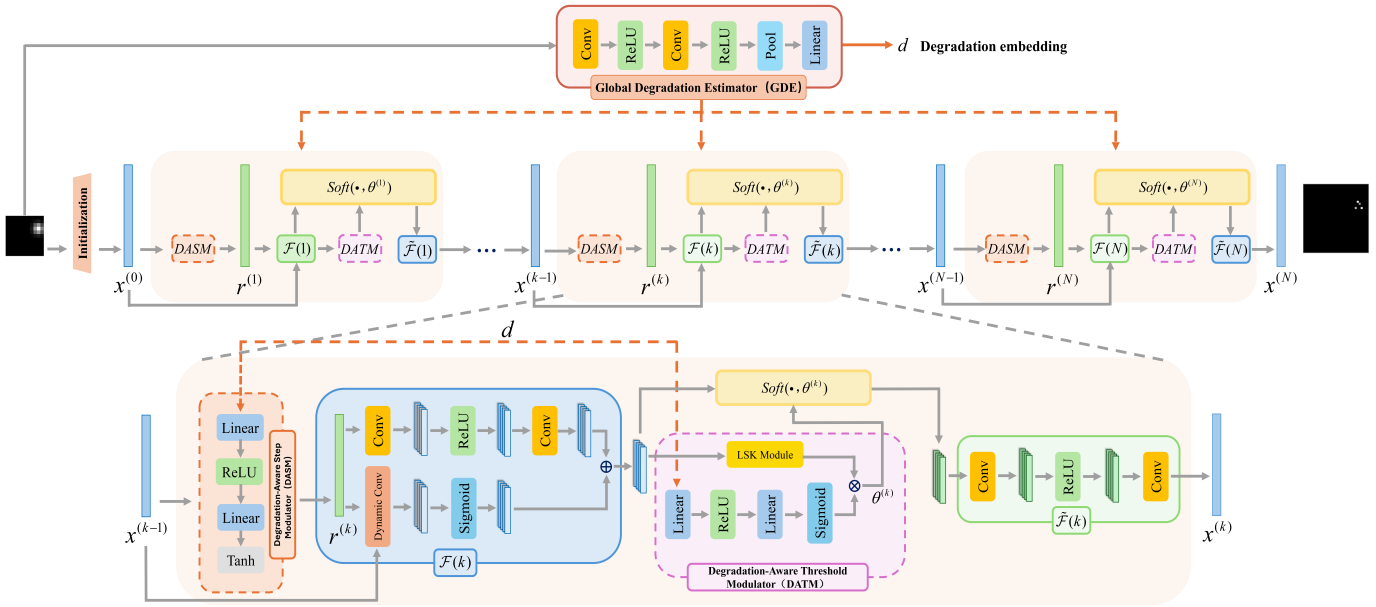


Figure 2. Architecture of the proposed DA-DISTA-Net. The GDE encodes the low-resolution observation into a degradation embedding $\mathbf{d}$ that is shared by all K stages. In stage k, DASM modulates the gradient step size to produce $\mathbf{r}^{(k)}$, and DATM rescales the soft-thresholding level to produce the sparse estimate $\mathbf{x}^{(k)}$.

3.4 Training Objective

The training loss consists of a pixel-level reconstruction term $\mathcal{L}_{\text{discrepancy}}$ and a symmetry term $\mathcal{L}_{\text{constraint}}$ that encourages the learned transforms $\mathcal{F}$ and $\tilde{\mathcal{F}}$ to be approximately inverse to each other:

$$\mathcal{L} = \mathcal{L}_{\text{discrepancy}} + \gamma\, \mathcal{L}_{\text{constraint}}, \tag{17}$$

$$\mathcal{L}_{\text{discrepancy}} = \frac{1}{N_b N} \sum_{i=1}^{N_b} \left\| \mathbf{x}_i^{(K)} - \mathbf{x}_i \right\|_2^2, \tag{18}$$

$$\mathcal{L}_{\text{constraint}} = \frac{1}{N_b N} \sum_{i=1}^{N_b} \sum_{k=1}^{K} \left\| \tilde{\mathcal{F}}\left(\mathcal{F}(\mathbf{x}_i^{(k)})\right) - \mathbf{x}_i^{(k)} \right\|_2^2, \tag{19}$$

where N_b is the number of samples per training batch, $\mathbf{x}_i$ is the ground truth of the i-th sample, and $\mathbf{x}_i^{(k)}$ is its estimate at stage k. $\mathcal{L}_{\text{constraint}}$ enforces $\tilde{\mathcal{F}} \circ \mathcal{F} \approx \mathcal{I}$ and keeps the sparse latent space from collapsing. The weight is set to $\gamma = 0.01$. The whole network is trained end-to-end from scratch on the random-degradation training set; training details are given in Section 4.

4 Experiments

4.1 Experimental Setup

All experiments use the random-degradation dataset of Section 3.2, with the sub-pixel magnification factor fixed at $c = 3$. Performance is measured with the CSO-mAP protocol designed for CSIST tasks [4]. Bounding-box metrics are unreliable when targets are closer than 1.9σ because their Airy spots interfere; CSO-mAP instead defines true and false positives through strict matching of sub-pixel position and intensity. We report five average precision (AP) values, AP-05 to AP-25, at distance thresholds $\tau \in \{0.05, 0.10, 0.15, 0.20, 0.25\}$ pixels. CSO-mAP is the mean of these AP values and therefore reflects performance under different localization accuracy requirements.

For DA-DISTA-Net, we set the number of unfolding stages to $K = 6$, the embedding dimension to $C_{\text{deg}} = 16$, and the dynamic branch fusion weight to $\lambda_{\text{weight}} = 0.7$. The observation matrix $\mathbf{\Phi}$ and the initialization matrix $\mathbf{Q}$ are precomputed with $\sigma = 0.5$ and kept fixed during training. The network is trained end-to-end with the Adam optimizer [18]; the initial learning rate is 1×10^{-4} and is halved every 80 epochs. We use a batch size of 64 and train for 250 epochs on a single NVIDIA GPU. Three groups of baselines are compared: (i) the classical optimization method ISTA; (ii) three super-resolution (SR) networks adapted to CSIST, namely RCAN [19], RDN [20], and SRFBN [21]; and (iii) four deep unfolding methods, namely ISTA-Net and ISTA-Net+ [3], FISTA-Net [22], and DISTA-Net [4]. All learning-based baselines are retrained from scratch with the same training schedule.

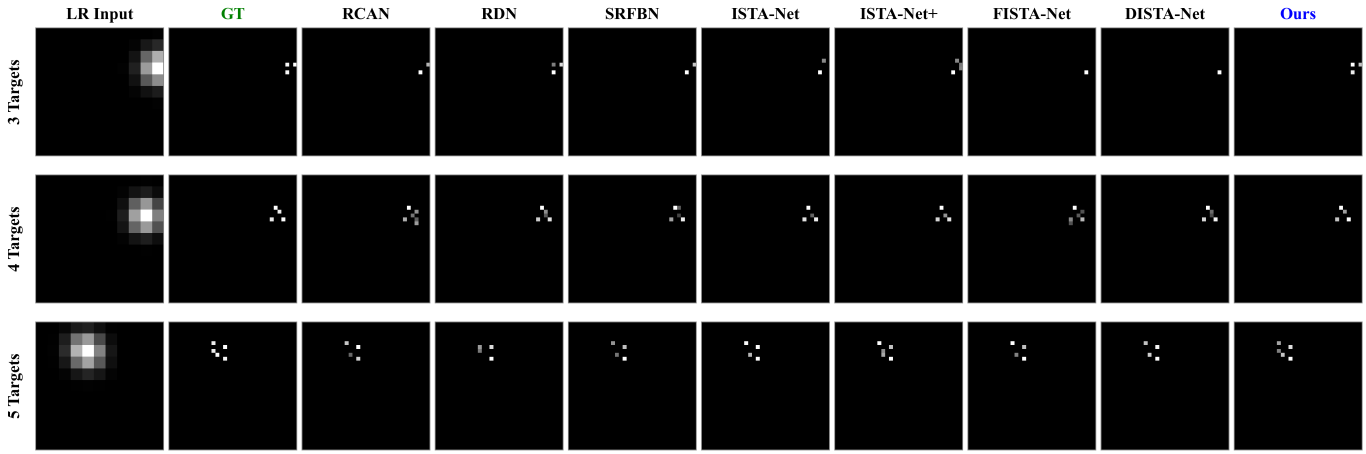


Figure 3. Visual comparison on scenes with three, four, and five closely-spaced infrared small targets.

Table 1. Comparison with state-of-the-art methods on the random-degradation dataset ($c = 3$). #P: number of parameters. All AP and mAP values are in %.

Category	Method	#P	mAP	AP-05	AP-10	AP-15	AP-20	AP-25
Traditional optimization	ISTA	-	4.71	0.03	0.44	2.51	7.12	13.47
Super-resolution network	RCAN	1.079 M	42.10	<u>0.37</u>	**7.47**	<u>36.77</u>	75.41	90.48
	RDN	22.306 M	41.30	0.31	6.54	34.93	74.35	90.38
	SRFBN	0.373 M	40.22	0.34	6.99	34.87	72.13	86.79
Deep unfolding network	ISTA-Net	0.171 M	41.53	**0.40**	<u>7.45</u>	36.35	74.37	89.08
	ISTA-Net+	0.337 M	41.85	0.34	6.64	35.91	75.66	90.72
	FISTA-Net	0.074 M	40.22	0.35	6.55	33.63	72.28	88.31
	DISTA-Net	2.179 M	<u>42.51</u>	0.33	6.70	36.68	<u>77.34</u>	<u>91.52</u>
	DA-DISTA-Net (ours)	2.249 M	**43.29**	**0.40**	7.06	**37.88**	**78.52**	**92.58**

4.2 Comparison with State-of-the-Art Methods

Table 1 lists the quantitative results on the random-degradation test set. The best and second-best results are shown in bold and underlined, respectively.

DA-DISTA-Net obtains the highest mAP, 43.29%, which is 0.78 percentage points above DISTA-Net. The improvement is larger for loose tolerances: AP-15, AP-20, and AP-25 increase by 1.20, 1.18, and 1.06 points, respectively. At the strict AP-10 threshold, RCAN (7.47) and ISTA-Net (7.45) remain ahead of our method (7.06). The results suggest that degradation-aware inference mainly improves coarse localization, which agrees with the expectation that calibrated step sizes limit the build-up of initialization errors over the unfolding iterations.

SR networks versus deep unfolding networks. Deep unfolding networks use the physical iteration as a prior, and the lightweight ones (ISTA-Net, ISTA-Net+, and FISTA-Net) are more parameter-efficient than the SR networks. RCAN reaches a competitive mAP of 42.10% but needs more parameters than these three unfolding networks, while ISTA-Net+ achieves a similar 41.85% with less than one third of the parameters of RCAN. Larger capacity alone does not help: RDN, with 22.306 M parameters, reaches only 41.30% mAP, below RCAN and below ISTA-Net, ISTA-Net+, and DISTA-Net. This suggests that, at a given parameter budget, the structural prior of deep unfolding is more useful than additional capacity.

Parameter efficiency. DA-DISTA-Net adds only 0.07 M parameters to DISTA-Net (2.249 M vs. 2.179 M, i.e., 3.2%), so the improvement can be attributed to degradation-conditioned inference rather than to a larger model.

Visual comparison. Figure 3 shows scenes with three to five closely spaced targets. The compared methods often give a wrong number of targets

or misplaced responses, whereas DA-DISTA-Net recovers the number and positions of the targets more accurately.

4.3 Ablation Study

To measure the contribution of each module, we perform a cross-domain ablation on both the fixed-σ ($\sigma = 0.5$) and the random-σ ($\sigma \sim U(0.2, 0.9)$) test sets. Four configurations are compared: (A) the DISTA-Net baseline; (B) baseline + GDE + DASM; (C) baseline + GDE + DATM; and (D) the full DA-DISTA-Net.

As Table 2 shows, GDE with DASM (Config B) improves the random-σ mAP by 0.52 points, and GDE with DATM (Config C) by 0.34 points. The two modules act on different parts of the iteration: DASM corrects the gradient step at the signal level and reduces the drift caused by a mismatched PSF, while DATM works at the feature level and keeps weak target responses from being removed by an overly high threshold under strong blur. The full model (Config D) gives the largest gain, 0.78 points. This gain is smaller than the sum of the individual gains (0.86 points), so the two modules are complementary but partly overlapping, and both are needed to reach the best result.

On the fixed-σ test set, the degradation-aware modules change mAP only slightly (-0.04 to -0.21 points). Because all inputs share the same PSF there, we expect the GDE to produce nearly identical embeddings and the modulators to behave almost as identity mappings. Hence, the proposed modules cause no noticeable negative transfer on single-degradation inputs.

4.4 Degradation Estimator Analysis

To check whether the GDE learns a physically meaningful representation rather than a shortcut unrelated to the task, we extract the embeddings $\mathbf{d}$ of 5,000 randomly selected test images, project them to 2D with t-SNE [23], and color each point by its ground-truth σ.

As shown in Figure 4, the embeddings form several clusters, and within each cluster σ changes gradually from small to large values. Since t-SNE preserves local neighborhoods but does not give a quantitative measure of embedding quality, this result is only qualitative evidence. Still, the ordering within clusters agrees with the GDE learning a representation that varies systematically with PSF blur strength; an embedding obtained by memorization would not be expected to show such ordering with respect to σ.

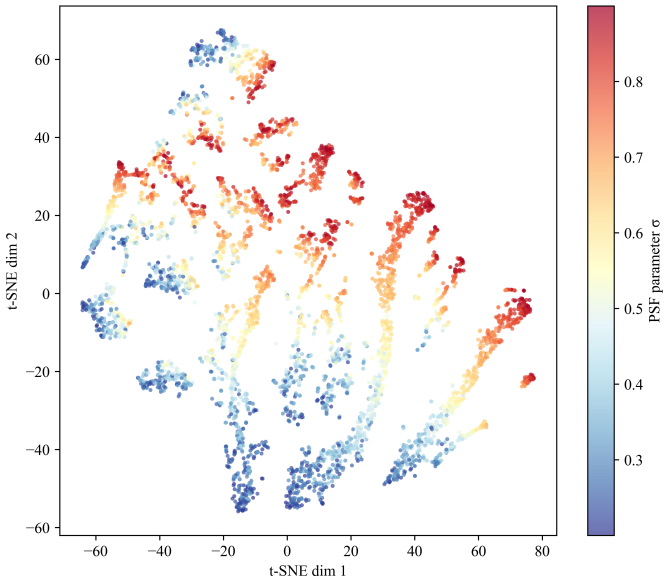


Figure 4. t-SNE visualization of the degradation embedding $\mathbf{d}$, colored by the ground-truth PSF parameter σ.

A natural question is whether the embedding mainly reflects blur strength or is confounded by scene factors such as the number and layout of targets, which vary independently of σ. The presence of several clusters in Figure 4 shows that some scene-dependent factor is likely encoded as well. Two observations, however, suggest that this factor has limited influence on the modulation. First, the GDE applies global average pooling right after two shallow convolutional layers, which removes fine spatial details such as target positions and keeps coarse statistics that are closely related to blur-induced energy spreading. Second, in the ablation study (Section 4.3), the degradation-aware modules change mAP only slightly on the fixed-σ test set (-0.04 to -0.21 points), although the number and layout of targets vary freely in that set. If the embedding were dominated by scene content rather than by σ, the step and threshold corrections would introduce inconsistent, sample-dependent perturbations under a fixed PSF, which is not observed.

5 Conclusion

This paper studies CSIST unmixing when the PSF degradation is unknown and proposes DA-DISTA-Net, a degradation-aware deep unfolding network. A GDE extracts an implicit degradation representation from the low-resolution observation, and the stage-wise DASM and DATM use it to calibrate the step size and the threshold of each unfolding stage. On the new random-degradation dataset, DA-DISTA-Net reaches an mAP of 43.29%, 0.78 percentage points above the

Table 2. Ablation study of the individual components of DA-DISTA-Net (mAP, %). Δ: change on the random-σ test set relative to Config A, in percentage points.

Config	GDE	DASM	DATM	Fixed-σ	Random-σ	Δ
A	✗	✗	✗	46.82	42.51	-
B	✓	✓	✗	46.78	43.03	+0.52
C	✓	✗	✓	46.61	42.85	+0.34
D	✓	✓	✓	46.74	43.29	+0.78

strongest baseline, with 3.2% more parameters. The t-SNE analysis suggests that the learned embeddings capture blur strength without explicit supervision. These results show that adding degradation awareness to deep unfolding is an effective way to improve CSIST unmixing under non-stationary PSFs. Future work will separate blur strength from scene content in the learned embedding, extend the framework to non-Gaussian and mixed degradations, and reduce the computational cost for real-time onboard processing.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported in part by the Distinguished Young Scholar Foundation of Hunan Province, China, under Grant 2024JJ2063 and in part by the National Natural Science Foundation of China under Grant 62401590.

Conflicts of Interest

Jungang Yang served as an Guest Editor of the *Chinese Journal of Information Fusion* at the time of manuscript submission. To ensure the integrity of the peer-review process, Jungang Yang was not involved in the editorial handling, peer review, or decision-making process for this manuscript, which was handled independently by a regular editor. The remaining authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Daubechies, I., Defrise, M., & De Mol, C. (2004). An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics, 57*(11), 1413-1457. [CrossRef]

[2] Beck, A., & Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences, 2*(1), 183-202. [CrossRef]

[3] Zhang, J., & Ghanem, B. (2018, June). ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1828-1837). IEEE. [CrossRef]

[4] Han, S., Yang, S., Li, Y., Zhang, X., Li, X., Yang, J., ... & Dai, Y. (2025, October). DISTA-Net: Dynamic closely-spaced infrared small target unmixing. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 14655-14664). IEEE. [CrossRef]

[5] Zhang, K., Van Gool, L., & Timofte, R. (2020). Deep unfolding network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3217-3226). [CrossRef]

[6] Cai, Y., Lin, J., Wang, H., Yuan, X., Ding, H., Zhang, Y., ... & Gool, L. V. (2022). Degradation-aware unfolding half-shuffle transformer for spectral compressive imaging. *Advances in Neural Information Processing Systems, 35*, 37749-37761. [CrossRef]

[7] Li, B., Xiao, C., Wang, L., Wang, Y., Lin, Z., Li, M., ... & Guo, Y. (2023). Dense nested attention network for infrared small target detection. *IEEE Transactions on Image Processing, 32*, 1745-1758. [CrossRef]

[8] Yuan, S., Qin, H., Yan, X., Akhtar, N., & Mian, A. (2024). SCTransNet: Spatial-channel cross transformer network for infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing, 62*, 1-15. [CrossRef]

[9] Borgerding, M., Schniter, P., & Rangan, S. (2017). AMP-inspired deep networks for sparse linear inverse problems. *IEEE Transactions on Signal Processing, 65*(16), 4293-4308. [CrossRef]

[10] Gregor, K., & LeCun, Y. (2010, June). Learning fast

approximations of sparse coding. In *Proceedings of the 27th international conference on international conference on machine learning* (pp. 399-406). https://mlanthology.org/icml/2010/gregor2010icml-learning

[11] Li, Y., Hou, Q., Zheng, Z., Cheng, M. M., Yang, J., & Li, X. (2023, October). Large selective kernel network for remote sensing object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE. [CrossRef]

[12] You, D., Xie, J., & Zhang, J. (2021). ISTA-Net++: Flexible deep unfolding network for compressive sensing. In *Proceedings of the 2021 IEEE International Conference on Multimedia and Expo* (pp. 1-6). [CrossRef]

[13] Chen, X., Liu, J., Wang, Z., & Yin, W. (2018). Theoretical linear convergence of unfolded ISTA and its practical weights and thresholds. *Advances in Neural Information Processing Systems, 31*. https://proceedings.neurips.cc/paper/2018/hash/cf8c9be2a4508a24ae92c9d3d379131d-Abstract.html

[14] Liu, J., Chen, X., Wang, Z., & Yin, W. (2019). ALISTA: Analytic weights are as good as learned weights in LISTA. In *International Conference on Learning Representations*. https://mlanthology.org/iclr/2019/liu2019iclr-alista

[15] Zheng, Z., Dai, W., Xue, D., Li, C., Zou, J., & Xiong, H. (2023). Hybrid ISTA: Unfolding ISTA with convergence guarantees using free-form deep neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45*(3), 3226-3244. [CrossRef]

[16] Wang, L., Wang, Y., Dong, X., Xu, Q., Yang, J., An, W., & Guo, Y. (2021, June). Unsupervised degradation representation learning for blind super-resolution. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 10581–10590). IEEE. [CrossRef]

[17] Xia, B., Zhang, Y., Wang, Y., Tian, Y., Yang, W., Timofte, R., & Van Gool, L. (2022). Knowledge distillation based degradation estimation for blind super-resolution. *arXiv preprint arXiv:2211.16928*. [CrossRef]

[18] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. [CrossRef]

[19] Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018). Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European Conference on Computer Vision* (pp. 286-301). [CrossRef]

[20] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018, June). Residual dense network for image super-resolution. In 2018 IEEE/CVF conference on computer vision and pattern recognition (pp. 2472-2481). IEEE. [CrossRef]

[21] Li, Z., Yang, J., Liu, Z., Yang, X., Jeon, G., & Wu, W. (2019, June). Feedback network for image super-resolution. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3867–3876). IEEE. [CrossRef]

[22] Xiang, J., Dong, Y., & Yang, Y. (2021). FISTA-Net: Learning a fast iterative shrinkage thresholding network for inverse problems in imaging. *IEEE Transactions on Medical Imaging, 40*(5), 1329-1339. [CrossRef]

[23] Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research, 9*, 2579–2605. https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf?fbcl



Wenjun Yang received the B.E. degree in underwater acoustics engineering from Harbin Engineering University (HEU), Harbin, China, in 2023. She is currently pursuing the M.S. degree with the College of Electronic Science and Technology, National University of Defense Technology (NUDT), Changsha, China. Her research interests include infrared image super-resolution and infrared dense target super-resolution. (Email: ywj@nudt.edu.cn)



Yaxiu Chen received the B.E. degree and is currently pursuing the master's degree at the National University of Defense Technology (NUDT), Changsha, China. Her research interests include remote sensing image super-resolution. (Email: chenyaxiu21@nudt.edu.cn)



Yang Wang received the M.E. degree in electronic and information engineering from the University of Chinese Academy of Sciences (UCAS), Beijing, China, and the Institute of Optics and Electronics (IOE), Chinese Academy of Sciences, Chengdu, China, in 2024. He is currently pursuing the Ph.D. degree with the College of Electronic Science and Technology, National University of Defense Technology (NUDT), Changsha, China. His research interests include infrared image super-resolution and infrared dense target super-resolution. (Email: wangyang0329@nudt.edu.cn)



Bo Lv received the B.S. degree from the National University of Defense Technology (NUDT), Changsha, China. He is currently pursuing the M.S. degree with the College of Electronic Science and Technology, NUDT. His research interests include infrared image processing and super-resolution. (Email: lvbo_20@nudt.edu.cn)



Jungang Yang received the B.E. and Ph.D. degrees from the National University of Defense Technology (NUDT), Changsha, China, in 2007 and 2013, respectively. From 2011 to 2012, he was a visiting Ph.D. student with the University of Edinburgh, Edinburgh, U.K. He is currently a Full Professor with the College of Electronic Science and Technology, NUDT. His research interests include computational imaging, image processing, compressive sensing, and sparse representation. He received the New Scholar Award of the Chinese Ministry of Education in 2012, and the Youth Innovation Award and the Youth Outstanding Talent award of NUDT in 2016. (Email: yangjungang@nudt.edu.cn)



Jiangwei Zou was born in Ji'an, China, in 1981. He received the bachelor's degree in information engineering and the master's and Ph.D. degrees in electronic science and technology from the National University of Defense Technology, Changsha, China, in 2003, 2004, and 2016, respectively. He is currently a Senior Engineer with the National University of Defense Technology. His research interests include image processing and information fusion. (Email: zoujiangwei@nudt.edu.cn)