

RPCANet: Deep Unfolding RPCA Based Infrared Small Target Detection

Fengyi Wu^{1,*} Tianfang Zhang^{1,*} Lei Li² Yian Huang¹ Zhenming Peng^{1,†}

¹University of Electronic Science and Technology of China, Chengdu, China

²University of Copenhagen, Denmark

fengyiwu@std.uestc.edu.cn; sparkcarleton@gmail.com; lilei@di.ku.dk

huangyian@std.uestc.edu.cn; zmpeng@uestc.edu.cn

Abstract

Deep learning (DL) networks have achieved remarkable performance in infrared small target detection (ISTD). However, these structures exhibit a deficiency in interpretability and are widely regarded as black boxes, as they disregard domain knowledge in ISTD. To alleviate this issue, this work proposes an interpretable deep network for detecting infrared dim targets, dubbed RPCANet. Specifically, our approach formulates the ISTD task as sparse target extraction, low-rank background estimation, and image reconstruction in a relaxed Robust Principle Component Analysis (RPCA) model. By unfolding the iterative optimization updating steps into a deep-learning framework, time-consuming and complex matrix calculations are replaced by theory-guided neural networks. RPCANet detects targets with clear interpretability and preserves the intrinsic image feature, instead of directly transforming the detection task into a matrix decomposition problem. Extensive experiments substantiate the effectiveness of our deep unfolding framework and demonstrate its trustworthy results, surpassing baseline methods in both qualitative and quantitative evaluations. Our source code is available at <https://github.com/fengyiwu98/RPCANet>.

1. Introduction

Research on infrared small target detection (ISTD) plays a crucial role in both civilian and military applications, such as maritime rescue [23], early-warning systems [6], and reconnaissance activities [51]. However, targets usually appear as point targets in just a few pixels due to the inherent extensive imaging distance in infrared detection, leading to limited shape and texture information. Further, background clutter and imaging system noise interfere with distinguishing infrared targets from complex noise, making ISTD a

*These authors contributed equally to this work.

†Corresponding author.

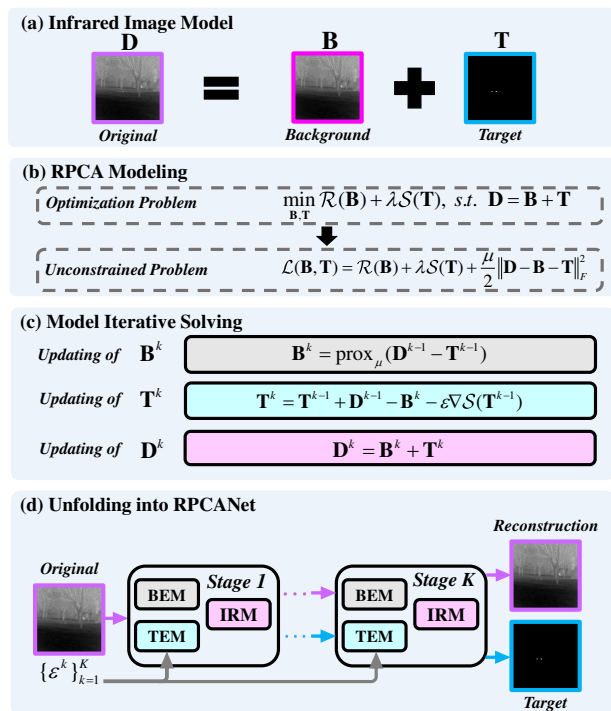


Figure 1. Overview of the suggested RPCANet. The optimization steps of the ISTD model are unfolded into a deep framework.

challenging academic topic.

In recent years, ISTD techniques have been advanced, falling into model-driven and data-driven categories. The model-driven framework includes filter-based [2, 9, 14], human vision system (HVS)-based [5, 17, 36], and optimization-based methods [6, 10, 35]. These methods provide theoretical and physical foundations and interpretable detection results, but their efficacy often hinges on finely tuned parameters, which may not suffice when the environment changes.

Data-driven methods for detecting small infrared targets have been enabled by the evolution of Artificial Neural Networks (ANNs). However, general deep-learning-based object detection techniques [11, 20, 25] have shown limited

performance on ISTD due to the significant difference in shape and texture information between visible light and infrared. Therefore, recent studies have formulated ISTD as a segmentation problem [7, 8, 19, 47], which can better capture the discriminative and salient features of small targets. However, common segmentation ISTD methods [7, 8] that adopt ImageNet [27] backbones may limit the infrared data representation [38]. Moreover, the multiple downsampling operations in mainstream networks can cause target loss due to the small target size. The reliability of these networks is questionable as they are designed without ISTD domain knowledge.

Taking into account the benefits and drawbacks of both model and data-driven ISTD techniques, a pertinent question arises: *Can we devise a balanced ISTD method that integrates the strengths of data-driven and model-driven methods to yield more reliable ISTD results?* Deep unfolding network (DUN), also known as algorithm unrolling, is an emerging technique that bridges the gap between iterative algorithms and neural networks. This technique has attracted a great deal of attention across multiple fields [26, 29, 33, 42]. DUN constructs a network by unrolling the iterative solving algorithm of an existing model at the iteration level. Its hyperparameters are subsequently updated in a network-based fashion. By establishing systematic and precise connections between iterative algorithms and neural networks, DUNs exhibit considerable potential for building more interpretable networks [1].

Interestingly, optimization-based ISTD schemes, as an extended application of image completion, involve optimization steps such as model formulation (as shown in Fig. 1 (a), low-rank background and sparse target) and iterative solving methods [10, 37], which align with the requirement of DUNs. However, complex matrix operations hinder scholars from delving into DUN-based ISTD. Some studies in medical image restoration [16] and video separation [4] have modeled images using robust principle component analysis (RPCA). But directly learning parameters in singular value thresholding (SVT) and soft thresholding (ST) overlooks the inherent correlation of images and complicates the empirical selection of regularization parameters.

Thus, to design an ISTD network that addresses the above issues and balances efficiency and interpretability, we develop a deep architecture called Robust Principle Component Analysis Network (RPCANet), as shown in Fig. 1 (d). Rather than applying ST on deep features, this framework uses neural layers to approximate the sparsity constraint function within a target extraction module (TEM). In addition, we design a background extraction module (BEM) to mimic the proximal function with convolution layers, thus approximating the background and eliminating steps such as SVT. An image reconstruction module (IRM) is also introduced to iteratively combine the target and background

outputs. In summary, our work contributes to the field of ISTD by:

- RPCANet is a learnable deep network architecture derived from a relaxed RPCA model in Fig. 1 (b) and (c), combining the accuracy of data-driven networks with the interpretability of model-driven ISTD approaches.
- Our TEM and BEM approximate the target and background for segmentation. Nonlinear proximal mapping problems for targets are handled effectively, while complex matrix calculations in background estimation are replaced with neural layers. Moreover, the proposed IRM combines the background and targets, aiming to reconstruct the image.
- Experimental results on multiple datasets demonstrate the effectiveness of RPCANet compared to state-of-the-art baselines. Visualization of learned characteristics provides comprehensive and reliable results.

2. Related Work

2.1. Optimization-based ISTD

Numerous ISTD methods have been developed over the past decades. Optimization-based methods are dominant in the model-driven category, compared to traditional filter-based [2, 9, 14] and HVS-based [5, 17, 36] schemes. The infrared patch image (IPI) model by Gao *et al.* [10] introduces the method of low-rank and sparse decomposition within the stable RPCA model [52]. This matrix-based category has been enriched by works in single [44, 49] and multi [35] subspaces. Based on the hypothesis that tensors leverage correlation information better than matrices [12], Dai *et al.* proposed a reweighted infrared patch tensor (RIPT) model for ISTD [6]. Various strategies [18, 37, 45] based on tensors have been applied to both single and multi-frame infrared sequences. The optimization-based methods model targets and backgrounds in a mathematically interpretable manner within RPCA. However, they have limitations in robustness, parameter tuning, and efficiency [46, 48]. Therefore, our goal is to develop a model that preserves interpretability while improving robustness and efficiency.

2.2. Deep Learning-based ISTD

In contrast to model-driven ones, convolutional neural networks (CNNs) are data-driven to learn the non-linear mappings between the original images and masks and are flexible to complicated scenarios. Wang *et al.* proposed a conditional GAN (MDvsFA-cGAN) [34] to reduce false alarms and missed detections in ISTD. Dai *et al.* presented an asymmetric contextual modulation (ACM) method [7] to enhance feature fusion. Zhang *et al.* [47] extracted contextual information of the target in the deep layer, based

on ACM. Li *et al.* [19] designed the DNANet with an enhanced receptive field to prevent small targets from disappearing. Wu *et al.* [38] integrated residual u-blocks into the network, which preserves feature resolution. Zhang *et al.* [46] merged an edge block in a U-Net framework to enhance edge features, considering target shapes vary. Ying *et al.* [40] updated labels with single-point supervision, based on effective models [7, 8, 19]. However, most data-driven methods tailor the segmentation network, which is a black box, instead of incorporating classic ISTD algorithms with field knowledge. Only a few studies have tried this approach [8, 15]. Therefore, we aim to find a solution that balances interpretability and effectiveness.

2.3. Deep Unfolding Networks

Deep unfolding networks (DUNs) solve iterative optimization problems via neural networks and have been widely used in image processing. Gregor and LeCun [13] first proposed the Learned ISTA (LISTA) model, which learns the parameters of the iterative shrinkage thresholding algorithm (ISTA). Zhang *et al.* [41, 42] improved it by introducing ISTA-Nets, which incorporate neural layers into the update steps. Yang *et al.* [39] developed ADMM-CSNet, which is based on unfolding the alternating direction multiplier method (ADMM) for MRI-oriented image reconstruction. Borgerding *et al.* [3] applied the approximate message passing algorithm (AMP) to learn sparse linear inverse problems, and Zhang *et al.* [50] extended it to compressive sensing. Other works also use DUNs for optimization models, such as low-rank representation [48] and RPCA [4, 28]. These DUNs frameworks blend the strengths of model and data-driven methods, achieving high and robust performance. Given the solid knowledge of optimization and deep learning-based methods, a progression towards DUN-based ISTD is a logical next step.

3. Deep Unfolding RPCA Network

3.1. Problem Formulation

For an infrared image \mathbf{D} , the physical model separates it into low-rank background \mathbf{B} and sparse target \mathbf{T} as:

$$\mathbf{D} = \mathbf{B} + \mathbf{T} , \quad (1)$$

where $\mathbf{D}, \mathbf{B}, \mathbf{T} \in \mathbb{R}^{m \times n}$. In a general RPCA [52] manner, we aim to recover the low-rank background \mathbf{B} that could pair the given \mathbf{D} by constraining sparse target \mathbf{T} and usually transform the detection problem into:

$$\min_{\mathbf{B}, \mathbf{T}} \text{rank}(\mathbf{B}) + \lambda \|\mathbf{T}\|_0 \quad \text{s.t. } \mathbf{D} = \mathbf{B} + \mathbf{T} , \quad (2)$$

where λ indicates a positive trade-off parameter, and $\|\cdot\|_0$ denotes the l_0 -norm as the number of nonzero entries.

However, solving (2) is NP-hard since the rank function and l_0 -norm are both non-convex and discontinuous. Considering this, models like IPI [10] individually replace them

with the nuclear norm ($\|\cdot\|_*$, the sum of singular values in a matrix) and l_1 -norm ($\|\cdot\|_1$) through principal component pursuit (PCP) and reformulate (2) as:

$$\min_{\mathbf{B}, \mathbf{T}} \|\mathbf{B}\|_* + \lambda \|\mathbf{T}\|_1 \quad \text{s.t. } \mathbf{D} = \mathbf{B} + \mathbf{T} . \quad (3)$$

In complex infrared scenarios, the background and target may vary in complexity and sparsity, and a single norm or rank function may not capture the practical constraints [49]. Thus, we use $\mathcal{R}(\mathbf{B})$ and $\mathcal{S}(\mathbf{T})$ to constrain the prior knowledge of the background and target images, respectively:

$$\min_{\mathbf{B}, \mathbf{T}} \mathcal{R}(\mathbf{B}) + \lambda \mathcal{S}(\mathbf{T}) \quad \text{s.t. } \mathbf{D} = \mathbf{B} + \mathbf{T} . \quad (4)$$

Moreover, to simplify the complexity of updating variables due to the augmented Lagrange multipliers [6], we adopt a simpler and more intuitive l_2 -norm to transform the constrained problem into an unconstrained one [30] as:

$$\mathcal{L}(\mathbf{B}, \mathbf{T}) = \mathcal{R}(\mathbf{B}) + \lambda \mathcal{S}(\mathbf{T}) + \frac{\mu}{2} \|\mathbf{D} - \mathbf{B} - \mathbf{T}\|_F^2 , \quad (5)$$

where μ is a penalty coefficient, and $\|\cdot\|_F$ indicates the Frobenius norm (F-norm). For a matrix \mathbf{X} , its F-norm equals $\sqrt{\sum_{i=1}^m \sum_{j=1}^n |\mathbf{X}_{ij}|^2}$. Based on (5), we can optimize the background and target individually in an iterative scheme.

3.2. Model Iterative Solving

Updating \mathbf{B}^* : To update the background, we draw the sub-problem as:

$$\mathbf{B}^* = \arg \min_{\mathbf{B}} \mathcal{R}(\mathbf{B}) + \frac{\mu}{2} \|\mathbf{B} + \mathbf{T} - \mathbf{D}\|_F^2 . \quad (6)$$

As (3) illustrates, former optimization-based ISTD approaches usually set $\mathcal{R}(\mathbf{B})$ as $\|\mathbf{B}\|_*$, then degenerates (6) into the sum of the nuclear norm and l_2 norm, and this problem has an analytical solution as:

$$\mathbf{B}^* = \mathcal{D}_\mu(\mathbf{D} - \mathbf{T}) , \quad (7)$$

where $\mathcal{D}_\mu(\cdot)$ denotes the SVT operator [21] with the threshold of μ . However, solving (7) involves SVD and functions on its singular values. In DUNs, we emulate this using neural networks, which means we must perform SVD on each neural tensor in each forward propagation. This poses challenges for time consumption and accuracy. [4] proposes an initialization method based on the best rank-r approximation SVD, but still faces precision issues. [48] introduces a technique with different rank constraints for the submatrices, which improves interpretability but depends on manual tuning and ignores the intrinsic image properties when simulating matrix computation in neural layers.

Thus, instead of adopting the nuclear norm and solving it with complex SVDs, we degrade it to a constrain function $\mathcal{R}(\mathbf{B})$ in this study, and introduce a proximal operator $\text{prox}_\mu(\cdot)$ to approximate the closed-form solution for the background, which is formulated as:

$$\mathbf{B}^* = \text{prox}_\mu(\mathbf{D} - \mathbf{T}) . \quad (8)$$

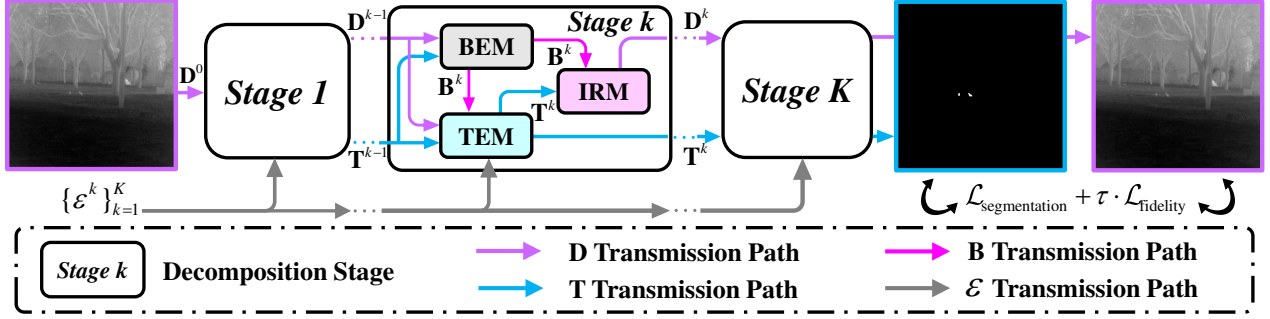


Figure 2. Overall structure of RPCANet. The network is composed of K stages, each following the unfolding steps detailed in Fig. 3.

We use convolutional layers to approximate proximal functions and solve the optimization problem. This method eliminates the need for complex matrix operations and leverages the nonlinear capabilities of neural networks to extract deep features from images in a data-driven way. And we will illustrate the detailed construction in Section 3.3.

Updating \mathbf{T}^* : Similar with (6), the sub-problem of optimizing the target is written as:

$$\mathbf{T}^* = \arg \min_{\mathbf{T}} \lambda \mathcal{S}(\mathbf{T}) + \frac{\mu}{2} \|\mathbf{T} + \mathbf{B} - \mathbf{D}\|_F^2. \quad (9)$$

As discussed in Section 3.1, common optimization approaches impose an l_1 norm constraint on the sparse target image. However, this poses the challenge of mapping the soft thresholding into the neural network [42]. Moreover, sparse constraints often vary depending on the detection scenarios change [49]. Thus, we aim to derive a simpler and more intuitive representation for the closed-form solution of (9). To handle the above issues, we consider conducting Taylor expansion of $\mathcal{S}(\mathbf{T})$. For a function $f(t)$, with its Lipschitz continuous gradient function $\nabla f(t)$, then $f(t)$ can be Taylor approximated at the a fix point t_0 by:

$$\hat{f}(t, t_0) \leftarrow \frac{L}{2} \left\| t - t_0 + \frac{1}{L} f(t_0) \right\|^2 + C, \quad (10)$$

where L is a constant, $C = -\frac{1}{2L} \|\nabla f(t_0)\|^2 + f(t_0)$ (please refer to the supplementary material for detail deductions). Based on this, we approximate $\mathcal{S}(\mathbf{T})$ at the last \mathbf{T}^{k-1} as:

$$\hat{\mathcal{S}}(\mathbf{T}, \mathbf{T}^{k-1}) \leftarrow \frac{L_S}{2} \left\| \mathbf{T} - \mathbf{T}^{k-1} + \frac{1}{L_S} \nabla \mathcal{S}(\mathbf{T}^{k-1}) \right\|_2^2 + C_s, \quad (11)$$

where L_s is the Lipschitz constant of $\mathcal{S}(\mathbf{T})$ and $C_s = -\frac{1}{2L_s} \|\nabla \mathcal{S}(\mathbf{T}^{k-1})\|_2^2 + \mathcal{S}(\mathbf{T}^{k-1})$ represents as a constant. And the updating formulate of the target matrix can be substituted as:

$$\begin{aligned} \mathbf{T}^* &= \arg \min_{\mathbf{T}} \lambda \hat{\mathcal{S}}(\mathbf{T}, \mathbf{T}^{k-1}) + \frac{\mu}{2} \|\mathbf{T} + \mathbf{B} - \mathbf{D}\|_F^2 \\ &= \arg \min_{\mathbf{T}} \frac{L_S}{2} \left\| \mathbf{T} - \mathbf{T}^{k-1} + \frac{1}{L_S} \nabla \mathcal{S}(\mathbf{T}^{k-1}) \right\|_2^2 \\ &\quad + \frac{\mu}{2} \|\mathbf{T} + \mathbf{B} - \mathbf{D}\|_F^2, \end{aligned} \quad (12)$$

which simply involves the sum of two l_2 norms, rather than a traditional l_1 norm constraint optimization problem, and therefore does not require the use of conventional algorithms or simulating soft thresholding [42]. By taking the derivative of the equation and equating it to zero, a closed-form solution for updating \mathbf{T} at the k -th step can be derived:

$$\begin{aligned} \mathbf{T}^k &= \frac{\lambda L_S}{\lambda L_S + \mu} \mathbf{T}^{k-1} + \frac{\mu}{\lambda L_S + \mu} (\mathbf{D}^{k-1} - \mathbf{B}^k) \\ &\quad - \frac{\lambda}{\lambda L_S + \mu} \nabla \mathcal{S}(\mathbf{T}^{k-1}), \end{aligned} \quad (13)$$

where all three coefficients are constant values. By assigning each of them a learnable vector, the final equation for updating the target matrix can be reformulated:

$$\mathbf{T}^k = \gamma \mathbf{T}^{k-1} + (1-\gamma)(\mathbf{D}^{k-1} - \mathbf{B}^k) - \varepsilon \nabla \mathcal{S}(\mathbf{T}^{k-1}), \quad (14)$$

where $\gamma = \frac{\lambda L_S}{\lambda L_S + \mu}$, $\varepsilon = \frac{\lambda}{\lambda L_S + \mu}$. We learn the function $\nabla \mathcal{S}$ end-to-end without complex matrix operations such as soft thresholding, satisfying the Lipschitz continuity assumption. And the updating equation for reconstructed \mathbf{D}^k is:

$$\mathbf{D}^k = \mathbf{B}^k + \mathbf{T}^k. \quad (15)$$

3.3. RPCANet Framework

This section describes the overall architecture of the network and module design of RPCANet, based on the optimization equations in Section 3.2. As shown in Fig. 2, the input to the network is an infrared image $\mathbf{X} \in \mathbb{R}^{H \times W}$ with targets, where H and W are the image height and width, and the update parameters are initialized as $\mathbf{D}^0 = \mathbf{X}$ and $\mathbf{T}^0 = 0$. These parameters are then passed through K decomposition stages, each corresponding to an iterative matrix low-rank sparse decomposition process, to simulate the update operation of multiple iterations in model-driven approaches.

In detail, the updated parameters \mathbf{D}^{k-1} and \mathbf{T}^{k-1} are fed into the k -th decomposition stage, where $k \in \{1, \dots, K\}$. The background \mathbf{B}^k , target \mathbf{T}^k , and reconstructed result \mathbf{D}^k for the current stage are estimated by BEM, TEM, and IRM, respectively. Typically, \mathbf{B}^k represents the latent variable of the current decomposition stage and is not involved in the parameter transfer between stages.

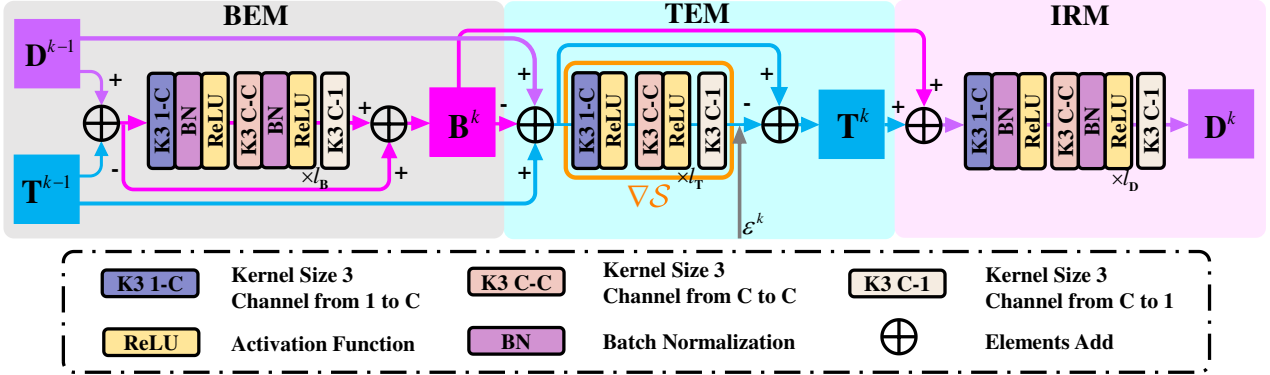


Figure 3. Detail network structure of the single stage from Fig. 2 in RPCANet: background extraction module (BEM), target extraction module (TEM), and image reconstruction module (IRM).

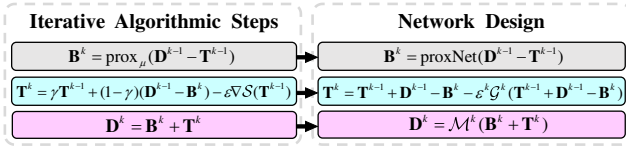


Figure 4. Relationship between the steps of the optimization algorithm and the modules of the deep unfolding network.

Background Estimation Module (BEM): As shown in Fig. 3, we adopt BEM to estimate the background. In (8), proximal operator $\text{prox}_{\mu}(\cdot)$ is to-be-decided. Here, inspired by former DUN-based works [29, 32, 33, 50] and in the spirits of flexibility [41], we adopt a residual structure $\text{proxNet}(\cdot)$ to simulate this operator. As (16) formulates:

$$\begin{aligned} \mathbf{B}^k &= \text{proxNet}(\mathbf{D}^{k-1} - \mathbf{T}^{k-1}) \\ &= \mathbf{D}^{k-1} - \mathbf{T}^{k-1} + \mathcal{F}^k(\mathbf{D}^{k-1} - \mathbf{T}^{k-1}), \end{aligned} \quad (16)$$

where $\mathcal{F}^k(\cdot)$ indicates the 3×3 convolution group in the structure as Fig. 3 shows. Here, $\mathcal{F}^k(\cdot)$ consists of l_B middle layers [$\text{Conv} + \text{BN} + \text{ReLU}$], and two convolution layers: feature extraction layer [$\text{Conv} + \text{BN} + \text{ReLU}$] and image reconstruction layer [Conv], respectively. Here, BN stands for batch normalization and ReLU for rectified linear unit [22]. All of them are in the stride of 1 and the padding of 1, and we set $C = 32$ and $l_B = 3$ in this study.

Target Extraction Module (TEM): This module takes the updated background \mathbf{B}^k , target \mathbf{T}^{k-1} , and reconstruction result \mathbf{D}^{k-1} as inputs, as shown in Fig. 3. Moreover, we set $\gamma = 0.5$ to evenly treat the three inputs and rewrite (14) as:

$$\mathbf{T}^k = \mathbf{T}^{k-1} + \mathbf{D}^{k-1} - \mathbf{B}^k - \epsilon \nabla S(\mathbf{T}^{k-1}). \quad (17)$$

We assign the parameter learning task to ϵ as ϵ^k , which is a learnable scalar independent of each reconstruction stage and does not share parameters. As to the Lipschitz continuous gradient function ∇S , [31] finds that a single-layer CNN consisting of a convolution layer and a ReLU activation function is Lipschitz continuous, which also holds for multiple stacked layers. Therefore, to avoid complex network design, we adopt simple convolution layers and

ReLU to simulate the function ∇S as shown in the orange box in Fig. 3, in accordance with the assumption in Section 3.2. We also introduce the difference of last \mathbf{D}^{k-1} and updated \mathbf{B}^k to enhance successive target feature, and the update function of \mathbf{T}^k is written as:

$$\mathbf{T}^k = \mathbf{T}^{k-1} + \mathbf{D}^{k-1} - \mathbf{B}^k - \epsilon^k \mathcal{G}^k(\mathbf{T}^{k-1} + \mathbf{D}^{k-1} - \mathbf{B}^k). \quad (18)$$

Specifically, $\mathcal{G}^k(\cdot)$ comprises an initial convolution layer, l_T of middle layers, and a reconstruction layer. Since the Lipschitz continuous property of the statistical operation BN is not clear [31], we omit the BN from the convolution block.

Image Reconstruction Module (IRM): To align with the RPCA process that maps the target and background into a restored image, we devise an IRM that converts the decomposition task into an image reconstruction task with a neural network $\mathcal{M}^k(\cdot)$, as presented in Fig. 3:

$$\mathbf{D}^k = \mathcal{M}^k(\mathbf{B}^k + \mathbf{T}^k). \quad (19)$$

Instead of applying residual blocks or other complex networks in the reconstruction module [32, 33], we employ a simple and decent CNN architecture [43] in learning image features and mapping the decomposed background and target effectively. Similar to $\mathcal{F}^k(\cdot)$, $\mathcal{M}^k(\cdot)$ has three types of convolution layer with l_D middle layer, where $l_D = 3$.

The corresponding updating operations in optimization-based and our DUN-based frameworks are shown in Fig. 4. To sum up, we propose an end-to-end training framework for the ISTD task named RPCANet. Its trainable parameters set Θ includes the convolutional network parameters in BEM, TEM, and IRM in each decomposition stage, and the independent trainable scalar ϵ in TEM, which collects as $\Theta = \{\Theta_{\text{BEM}}^k, \Theta_{\text{TEM}}^k, \Theta_{\text{IRM}}^k, \epsilon^k\}_{k=1}^K$, where K is the number of overall decomposition stages.

4. Network Training

4.1. Training Loss

Since we separate an ISTD task into the target segmentation and infrared image reconstruction, thus the loss func-

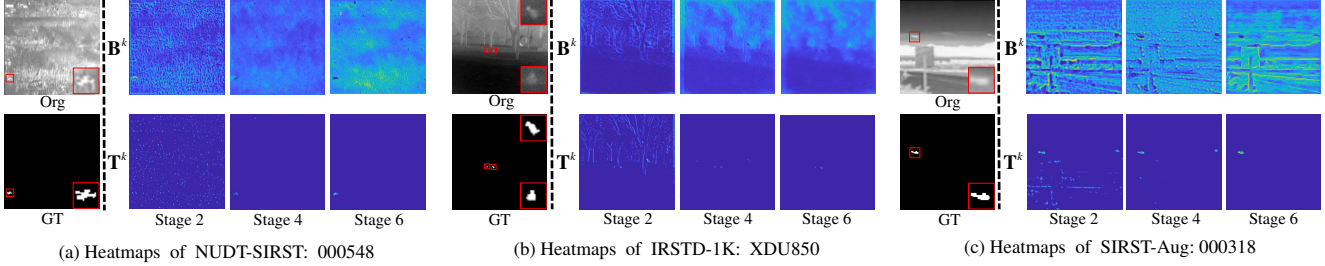


Figure 5. Heatmaps of variables \mathbf{B}^k and \mathbf{T}^k in different decomposition stages of RPCANet on three representative scenes when $K = 6$.

tion consists of two components: $\mathcal{L}_{\text{segmentation}}$ and $\mathcal{L}_{\text{fidelity}}$. The former measures the target segmentation performance using SoftIoU [24], while the latter measures the infrared image reconstruction performance using the least squares error between the reconstructed image and the original image. The loss function is defined as follows:

$$\begin{aligned} \mathcal{L}_{\text{total}} &= \mathcal{L}_{\text{segmentation}} + \tau \cdot \mathcal{L}_{\text{fidelity}} \\ &= 1 - \frac{1}{N_t} \sum_{i=1}^{N_t} \frac{TP}{FP+TP+FN} + \frac{\tau}{N_t N} \sum_{i=1}^{N_t} \|\mathbf{D}^K - \mathbf{D}\|_F^2, \end{aligned} \quad (20)$$

where N_t and N are the total training number and total pixels per image. τ is the regularization parameter and set to 0.01 in our experiment.

4.2. Implementation Details

We conduct experiments on three publicly available datasets: SIRST-Aug [47], NUDT-SIRST [19], and IRSTD-1K [46], with their split strategies, all images are normalized into 256×256 . Besides, evaluation metrics are two target-level: probability of detection (P_d) and false alarm rate (F_a); two pixel-level: mean intersection over union ($mIoU$) and F-measure (F_1); and the receiver operating characteristic (ROC) curve and the area under ROC (AUC).

We train our model for 400 epochs in the PyTorch framework on each dataset using an Nvidia GeForce 3090 GPU. We use an Adam optimizer with an initial learning rate of 10^{-4} in the poly policy, where the initial learning rate is multiplied by $(1 - \frac{iter}{total_iter})^{0.9}$ and a batch size of 8.

5. Experiment

To confirm the fundamental process of the suggested network, we verify the model by visualizing mid-layers and conduct ablation studies. Next, we describe experiments on open-source datasets for performance evaluation.

5.1. Model Verification

Fig. 5 presents the heatmaps of the intermediate update variable \mathbf{B}^k and \mathbf{T}^k , where $k \in \{2, 4, 6\}$. At lower stages, the network prioritizes low-level edge texture information in the background, leading to insufficient background reconstruction. As the decomposition deepens, the network

Table 1. Effect of stage number K on the detection performance in $mIoU$ (%), F_1 (%), P_d (%), and F_a (10^{-5}) on SIRST-Aug [47].

Stages (K)	Params	$mIoU \uparrow$	$F_1 \uparrow$	$P_d \uparrow$	$F_a \downarrow$
1	0.113M	60.10	75.08	98.21	62.20
2	0.227M	69.26	81.84	98.49	46.12
3	0.340M	69.82	82.23	98.07	42.63
4	0.453M	69.50	82.01	98.07	40.36
5	0.567M	70.24	82.52	96.01	36.02
6	0.680M	72.54	84.08	98.21	34.14
7	0.793M	70.98	83.02	96.15	35.85

Table 2. Effect of layer number l_T on the detection performance in $mIoU$ (%), F_1 (%), P_d (%), and F_a (10^{-5}) on SIRST-Aug [47].

TEM	Params	$mIoU \uparrow$	$F_1 \uparrow$	$P_d \uparrow$	$F_a \downarrow$
$l_T = 1$	0.402M	67.07	80.29	99.17	40.53
$l_T = 3$	0.513M	70.63	82.79	97.80	41.36
$l_T = 9$	0.846M	67.99	80.95	94.36	32.75
$l_T = 12$	1.013M	69.98	82.34	97.25	38.46
$l_T = 6$	0.680M	72.54	84.08	98.21	34.14

Table 3. Studies on different proximal networks and w/o IRM on the detection performance in $mIoU$ (%), F_1 (%), P_d (%), and F_a (10^{-5}) on SIRST-Aug [47].

BEM	IRM	Params	$mIoU \uparrow$	$F_1 \uparrow$	$P_d \uparrow$	$F_a \downarrow$
RB	✓	0.624M	66.33	79.76	98.07	61.39
CNN	✓	0.679M	71.43	83.33	97.77	32.45
Ours	✗	0.507M	67.07	80.29	96.56	36.89
Ours	✓	0.680M	72.54	84.08	98.21	34.14

acquires more comprehensive information, demonstrating its ability to learn from formula-guided training.

Likewise, \mathbf{T}^k progressively learns the target’s location from disorganized high-frequency information, resulting in a sparse matrix as the segmented output. Fig. 5 (c) shows that false alarms in the initial stages are eliminated under mask supervision, showing the data-driven way corrects the detection result. These successive heatmaps verify the effectiveness of RPCANet in learning infrared variables according to training data while preserving the interpretability of RPCA model that facilitates the ISTD task.

5.2. Ablation Studies

Effects of Parameter K and l_T : Table 1 compares

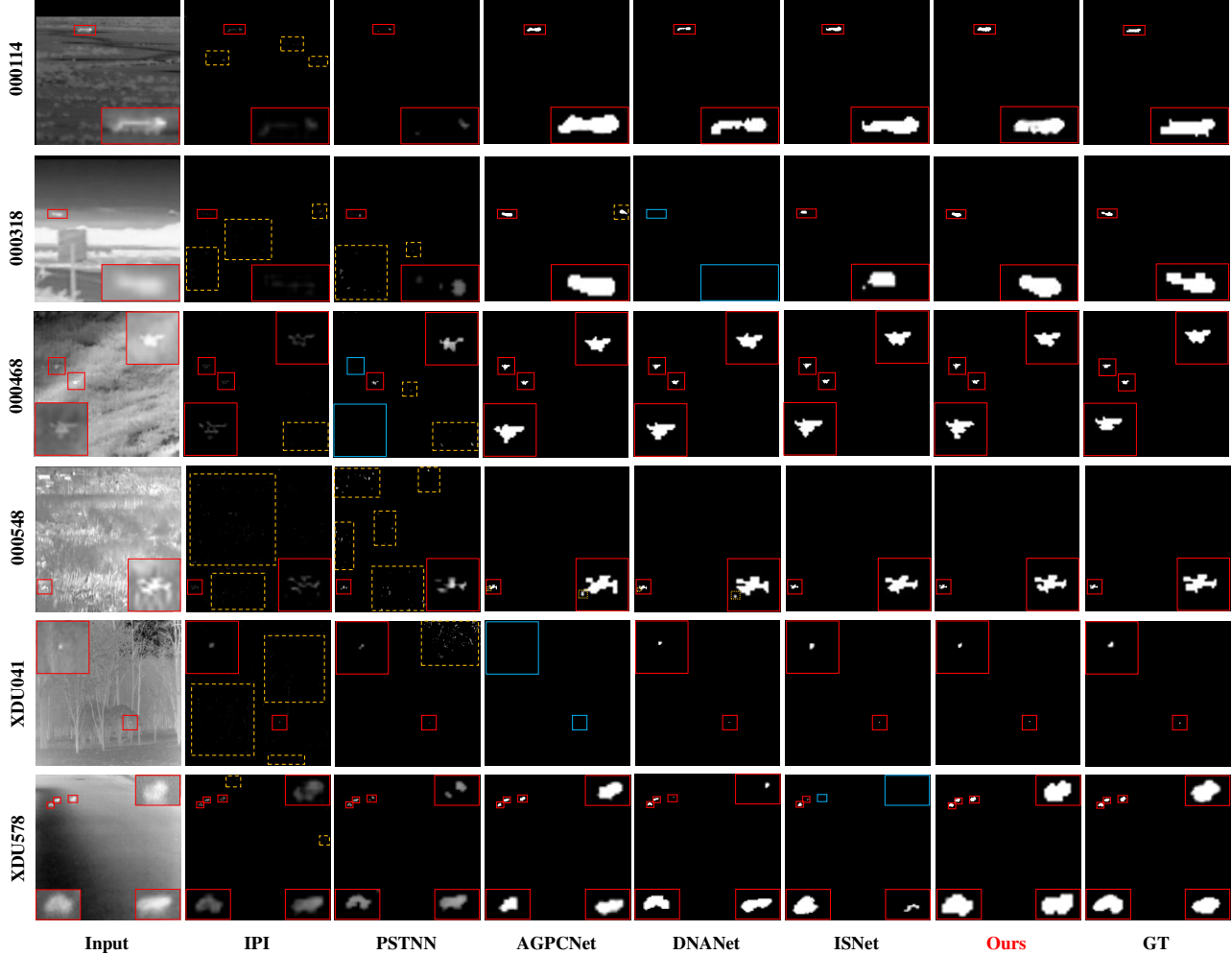


Figure 6. Representative visual results from various ISTD methods. Accurately detected targets, missed targets, and false alarms are each highlighted by boxes in red, blue, and yellow.

Table 4. Studies on $\nabla\mathcal{S}$ simulation network in detection performance $mIoU$ (%), F_1 (%) on three different datasets.

Config	NUDT-SIRST [19]		IRSTD-1K [46]		SIRST-Aug [47]	
	mIoU \uparrow	F1 \uparrow	mIoU \uparrow	F1 \uparrow	mIoU \uparrow	F1 \uparrow
\mathbf{T}^{k-1}	88.54	93.92	60.98	75.76	70.54	82.72
Ours	89.31	94.35	63.21	77.45	72.54	84.08

the effects of different reconstruction stage counts. Our method only reaches significant detection performance in two stages, validating the essential attributes of the suggested RPCANet. Additionally, we see that $K = 7$ has a worse ISTD performance than $K = 6$, which makes sense given that a higher K could hinder gradient propagation, based on this, we set K to 6. Similarly, as shown in Table 2, the number of the convolutional layer $l_{\mathbf{T}}$ faces the same manner, the performance of the network can be improved as $l_{\mathbf{T}}$ raises, but the gain effect of too many reconstruction stages on the network performance is limited, thus, we take $l_{\mathbf{T}} = 6$ in all our experiments.

Studies of proxNet(\cdot) and IRM: The emulation of proximal operators acts an essential role in background estimation [33, 41], we investigate two proxNet(\cdot) constructions (residual block (RB) [32] and plain CNN) for proximal operators. Table 3 demonstrates that RB can yield some effects and a small parameter count, our network in Section 3.3 achieves better results with a minimal parameter increase. We also examine the influence of the IRM. As depicted in the third row of Table 3, RPCANet, guided by the reconstruction module, exhibits substantial improvements in four metrics, proving the module’s effectiveness.

Studies of $\nabla\mathcal{S}$ Simulation Network: We investigate the impact of incorporating \mathbf{D}^{k-1} and \mathbf{B}^k subtraction within $\mathcal{G}^k(\cdot)$ for target feature enhancement. Table 4 compares two configurations. While using \mathbf{T}^{k-1} alone yields satisfactory results, incorporating additional information in (18) significantly improves $mIoU$ and F_1 . We can conclude that appropriate modifications to the network structure, guided by the model’s prior, can enhance detection performance.

Table 5. $mIoU$ (%), F_1 (%), P_d (%), F_a (10^{-5}), and runtime values of different methods performance on NUDT-SIRST [19], IRSTD-1K [46], and SIRST-Aug [47]. The second column records the parameter statistics for data-driven algorithms.

Methods	Params	NUDT-SIRST [19]				IRSTD-1K [46]				SIRST-Aug [47]				Time (s) CPU/GPU
		$mIoU$ \uparrow	F_1 \uparrow	P_d \uparrow	F_a \downarrow	$mIoU$ \uparrow	F_1 \uparrow	P_d \uparrow	F_a \downarrow	$mIoU$ \uparrow	F_1 \uparrow	P_d \uparrow	F_a \downarrow	
Tophat [2]	-	22.23	36.37	96.19	76.27	5.35	10.15	68.73	86.50	16.70	28.62	95.05	11.77	0.0111/-
MPCM [36]	-	9.26	16.95	70.58	32.74	14.87	25.89	68.73	6.51	19.76	33.00	93.40	3.14	0.0624/-
IPI [10]	-	34.62	51.43	92.38	7.54	18.64	31.42	78.01	11.10	21.93	35.87	80.06	1.62	3.0972/-
PSTNN [45]	-	25.46	40.58	78.52	7.95	16.38	28.15	69.07	7.65	13.83	24.30	59.97	1.56	0.2249/-
ACM [7]	0.398M	69.00	81.66	95.98	13.34	61.56	76.20	92.93	8.88	70.49	82.62	96.70	35.29	-/0.0072
ALCNet [8]	0.427M	71.48	83.37	96.30	11.45	58.23	73.61	92.92	10.31	66.21	79.67	97.80	37.40	-/0.0070
ISNet [46]	0.967M	87.51	93.34	97.35	3.37	55.29	71.21	94.61	14.19	70.51	82.71	97.66	31.57	-/0.0132
AGPCNet [47]	12.360M	85.40	92.13	98.10	4.72	61.00	75.76	89.35	5.34	72.16	83.83	99.03	35.56	-/0.0205
DNANet [19]	4.697M	83.94	91.27	98.52	6.21	60.51	75.40	91.07	5.43	69.58	82.06	96.14	27.26	-/0.0250
UIUNet [38]	50.540M	88.71	94.01	91.43	1.89	63.06	77.35	93.60	6.57	71.80	83.59	98.35	28.29	-/0.0261
Ours	0.680M	89.31	94.35	97.14	2.87	63.21	77.45	88.31	4.39	72.54	84.08	98.21	34.14	-/0.0096

5.3. Comparison to State-of-the-art Methods

For model-driven methods, we select: filter-based Tophat [2], HVS-based MPCM [36], matrix optimization-based IPI [10], and tensor optimization-based PSTNN [45]. For data-driven algorithms, we conduct experiments on ACM [7], ALCNet [8], ISNet [46], AGPCNet [47], DNANet [19], and UIUNet [38].

Qualitative Results: Fig. 6 depicts visual results obtained from various algorithms applied to three datasets. RPCANet effectively generalizes complex scenarios, producing outputs with accurate target shapes and low false alarm rates. Compared to model-driven algorithms, our network excels at suppressing false alarms while preserving the sparse shape of infrared targets. Although some DL-based models can effectively reduce false alarms, they may struggle in small and complex situations, leading to missed detections. Algorithms like ISNet largely maintain the target shape; however, they occasionally encounter difficulties in multi-target scenarios. The visualization results clearly illustrate that our network combines the advantages of optimization-based and DL-based methods.

Quantitative Results: To showcase the effectiveness of our network, we compare its detection performance with the SOTA baseline. Table 5 demonstrates that RPCANet outperforms most model-driven and data-driven ISTD frameworks in four indicators using fewer parameters. Model-driven methods like IPI excel in target-level metrics but lack pixel-level accuracy. Data-driven networks improve $mIoU$ and F_1 scores while preserving the target’s shape. However, DL-based algorithms like AGPCNet and UIUNet suffer from overfitting and require more parameters. In contrast, RPCANet combines model-driven priors with accurate object extraction and segmentation guided by data, achieving superior performance with fewer parameters. ROC curves in Fig. 7 highlight that our model rapidly reaches the upper-left corner and exhibits competitive performance in terms of AUC: RPCANet (0.9857), ACM (0.9830), and UIUNet (0.9477). Moreover, the last column in Table 5 shows that our framework obtains great computational efficiency.

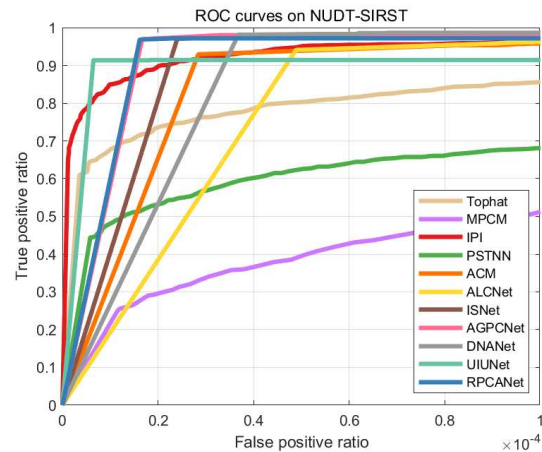


Figure 7. ROC curves of different methods on NUDT-SIRST [19].

6. Conclusion

In this paper, we propose an interpretable framework for infrared small target detection based on deep unfolding networks. We model the ISTD task as a relaxed RPCA problem and solve the optimization steps via network emulations, including the proximal network and sparsity-constrained neural layers. Our scheme produces trustworthy visualizations and outstanding detection results in extensive experiments on various public datasets. The architecture of RPCANet effectively guides neural layers to learn low-rank backgrounds and sparse targets, facilitating the detection tasks in an almost “white box” manner. We hope our findings cloud inspire researchers to explore more interpretable solutions for ISTD problems in the future.

Acknowledgments

This work was supported by the Natural Science Foundation of Sichuan Province of China (No.2022NSFSC40574) and National Natural Science Foundation of China (No.61775030, No.61571096).

References

- [1] Botao An, Shibin Wang, Fuhua Qin, Zhibin Zhao, Ruqiang Yan, and Xuefeng Chen. Adversarial algorithm unrolling network for interpretable mechanical anomaly detection. *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, 2023. **2**
- [2] Xiangzhi Bai and Fugen Zhou. Analysis of new top-hat transformation and the application for infrared dim small target detection. *Pattern Recognition (PR)*, 43(6):2145–2156, 2010. **1, 2, 8**
- [3] Mark Borgerding, Philip Schniter, and Sundeep Rangan. Amp-inspired deep networks for sparse linear inverse problems. *IEEE Transactions on Signal Processing (TSP)*, 65(16):4293–4308, 2017. **3**
- [4] HanQin Cai, Jialin Liu, and Wotao Yin. Learned robust pca: A scalable deep unfolding approach for high-dimensional outlier detection. *Advances in Neural Information Processing Systems (NIPS)*, 34:16977–16989, 2021. **2, 3**
- [5] CL Philip Chen, Hong Li, Yantao Wei, Tian Xia, and Yuan Yan Tang. A local contrast method for small infrared target detection. *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, 52(1):574–581, 2013. **1, 2**
- [6] Yimian Dai and Yiquan Wu. Reweighted infrared patch-tensor model with both nonlocal and local priors for single-frame small target detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS)*, 10(8):3752–3767, 2017. **1, 2, 3**
- [7] Yimian Dai, Yiquan Wu, Fei Zhou, and Kobus Barnard. Asymmetric contextual modulation for infrared small target detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 950–959, 2021. **2, 3, 8**
- [8] Yimian Dai, Yiquan Wu, Fei Zhou, and Kobus Barnard. Attentional local contrast networks for infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, 59(11):9813–9824, 2021. **2, 3, 8**
- [9] Suyog D Deshpande, Meng Hwa Er, Ronda Venkateswarlu, and Philip Chan. Max-mean and max-median filters for detection of small targets. In *Signal and Data Processing of Small Targets 1999*, volume 3809, pages 74–83. SPIE, 1999. **1, 2**
- [10] Chenqiang Gao, Deyu Meng, Yi Yang, Yongtao Wang, Xiaofang Zhou, and Alexander G Hauptmann. Infrared patch-image model for small target detection in a single image. *IEEE Transactions on Image Processing (TIP)*, 22(12):4996–5009, 2013. **1, 2, 3, 8**
- [11] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1440–1448, 2015. **1**
- [12] Donald Goldfarb and Zhiwei Qin. Robust low-rank tensor recovery: Models and algorithms. *SIAM Journal on Matrix Analysis and Applications (SIMAX)*, 35(1):225–253, 2014. **2**
- [13] Karol Gregor and Yann LeCun. Learning fast approximations of sparse coding. In *Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML)*, pages 399–406, 2010. **3**
- [14] Mohiy M Hadhoud and David W Thomas. The two-dimensional adaptive lms (tdlms) algorithm. *IEEE Transactions on Circuits and Systems (TCAS)*, 35(5):485–494, 1988. **1, 2**
- [15] Qingyu Hou, Zhipeng Wang, Fanjiao Tan, Ye Zhao, Hao-liang Zheng, and Wei Zhang. Ristdnet: Robust infrared small target detection network. *IEEE Geoscience and Remote Sensing Letters (GRSL)*, 19:1–5, 2021. **3**
- [16] Wenqi Huang, Ziwen Ke, Zhuo-Xu Cui, Jing Cheng, Zhi-lang Qiu, Sen Jia, Leslie Ying, Yanjie Zhu, and Dong Liang. Deep low-rank plus sparse network for dynamic mr imaging. *Medical Image Analysis (MIA)*, 73:102190, 2021. **2**
- [17] Sungho Kim, Yukyung Yang, Joohyoung Lee, and Yongchan Park. Small target detection utilizing robust methods of the human visual system for first. *Journal of Infrared, Millimeter, and Terahertz Waves*, 30:994–1011, 2009. **1, 2**
- [18] Xuan Kong, Chunping Yang, Siying Cao, Chaochai Li, and Zhenming Peng. Infrared small target detection via nonconvex tensor fibered rank approximation. *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, 60:1–21, 2021. **2**
- [19] Boyang Li, Chao Xiao, Longguang Wang, Yingqian Wang, Zaiping Lin, Miao Li, Wei An, and Yulan Guo. Dense nested attention network for infrared small target detection. *IEEE Transactions on Image Processing (TIP)*, 32:1745–1758, 2022. **2, 3, 6, 7, 8**
- [20] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European Conference on Computer Vision (ECCV)*, pages 21–37. Springer, 2016. **1**
- [21] Shiqian Ma, Donald Goldfarb, and Lifeng Chen. Fixed point and bregman iterative methods for matrix rank minimization. *Mathematical Programming*, 128(1-2):321–353, 2011. **3**
- [22] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML)*, pages 807–814, 2010. **5**
- [23] Dilip K Prasad, Deepu Rajan, Lily Rachmawati, Eshan Rajabally, and Chai Quek. Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey. *IEEE Transactions on Intelligent Transportation Systems (TITS)*, 18(8):1993–2016, 2017. **1**
- [24] Md Atiqur Rahman and Yang Wang. Optimizing intersection-over-union in deep neural networks for image segmentation. In *International Symposium on Visual Computing (ISVC)*, pages 234–244. Springer, 2016. **6**
- [25] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, 2016. **1**
- [26] Chao Ren, Xiaohai He, Chuncheng Wang, and Zhibo Zhao. Adaptive consistency prior based deep network for image denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8596–8606, 2021. **2**

- [27] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision (IJCV)*, 115:211–252, 2015. [2](#)
- [28] Oren Solomon, Regev Cohen, Yi Zhang, Yi Yang, Qiong He, Jianwen Luo, Ruud JG van Sloun, and Yonina C Eldar. Deep unfolded robust pca with application to clutter suppression in ultrasound. *IEEE Transactions on Medical Imaging (TMI)*, 39(4):1051–1063, 2019. [3](#)
- [29] Marc Tomás-Cruz, Alessandro Sebastianelli, Bartomeu Coll, Joan Duran, and Jamila Mifdal. Deep unfolding for hyper-sharpening using a high-frequency injection module. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2105–2114, 2023. [2](#), [5](#)
- [30] Paul Tseng. On accelerated proximal gradient methods for convex-concave optimization. *SIAM Journal on Optimization (SIOPT)*, 2(3), 2008. [3](#)
- [31] Aladin Virmaux and Kevin Scaman. Lipschitz regularity of deep neural networks: analysis and efficient estimation. *Advances in Neural Information Processing Systems (NIPS)*, 31, 2018. [5](#)
- [32] Hong Wang, Yuexiang Li, Haimiao Zhang, Deyu Meng, and Yefeng Zheng. Indudonet+: A deep unfolding dual domain network for metal artifact reduction in ct images. *Medical Image Analysis (MIA)*, 85:102729, 2023. [5](#), [7](#)
- [33] Hong Wang, Qi Xie, Qian Zhao, and Deyu Meng. A model-driven deep neural network for single image rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3103–3112, 2020. [2](#), [5](#), [7](#)
- [34] Huan Wang, Luping Zhou, and Lei Wang. Miss detection vs. false alarm: Adversarial learning for small object segmentation in infrared images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8509–8518, 2019. [2](#)
- [35] Xiaoyang Wang, Zhenming Peng, Dehui Kong, and Yanmin He. Infrared dim and small target detection based on stable multisubspace learning in heterogeneous scene. *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, 55(10):5481–5493, 2017. [1](#), [2](#)
- [36] Yantao Wei, Xinge You, and Hong Li. Multiscale patch-based contrast measure for small infrared target detection. *Pattern Recognition (PR)*, 58:216–226, 2016. [1](#), [2](#), [8](#)
- [37] Fengyi Wu, Hang Yu, Anran Liu, Junhai Luo, and Zhenming Peng. Infrared small target detection using spatio-temporal 4d tensor train and ring unfolding. *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, 2023. [2](#)
- [38] Xin Wu, Danfeng Hong, and Jocelyn Chanussot. Uiu-net: U-net in u-net for infrared small object detection. *IEEE Transactions on Image Processing (TIP)*, 32:364–376, 2023. [2](#), [3](#), [8](#)
- [39] Yan Yang, Jian Sun, Huibin Li, and Zongben Xu. Admmcsnet: A deep learning approach for image compressive sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 42(3):521–538, 2018. [3](#)
- [40] Xinyi Ying, Li Liu, Yingqian Wang, Ruoqing Li, Nuo Chen, Zaiping Lin, Weidong Sheng, and Shilin Zhou. Mapping degeneration meets label evolution: Learning infrared small target detection with single point supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15528–15538, 2023. [3](#)
- [41] Di You, Jingfen Xie, and Jian Zhang. Ista-net++: Flexible deep unfolding network for compressive sensing. In *2021 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2021. [3](#), [5](#), [7](#)
- [42] Jian Zhang and Bernard Ghanem. Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1828–1837, 2018. [2](#), [3](#), [4](#)
- [43] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018. [5](#)
- [44] Landan Zhang, Lingbing Peng, Tianfang Zhang, Siying Cao, and Zhenming Peng. Infrared small target detection via non-convex rank approximation minimization joint l 2, 1 norm. *Remote Sensing*, 10(11):1821, 2018. [2](#)
- [45] Landan Zhang and Zhenming Peng. Infrared small target detection based on partial sum of the tensor nuclear norm. *Remote Sensing*, 11(4):382, 2019. [2](#), [8](#)
- [46] Mingjin Zhang, Rui Zhang, Yuxiang Yang, Haichen Bai, Jing Zhang, and Jie Guo. Isnet: Shape matters for infrared small target detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 877–886, 2022. [2](#), [3](#), [6](#), [7](#), [8](#)
- [47] Tianfang Zhang, Lei Li, Siying Cao, Tian Pu, and Zhenming Peng. Attention-guided pyramid context networks for detecting infrared small target under complex background. *IEEE Transactions on Aerospace and Electronic Systems (TAES)*, 2023. [2](#), [6](#), [7](#), [8](#)
- [48] Tianfang Zhang, Lei Li, Christian Igel, Stefan Oehmcke, Fabian Gieseke, and Zhenming Peng. Lr-csnet: Low-rank deep unfolding network for image compressive sensing. In *2022 IEEE 8th International Conference on Computer and Communications (ICCC)*, pages 1951–1957. IEEE, 2022. [2](#), [3](#)
- [49] Tianfang Zhang, Zhenming Peng, Hao Wu, Yanmin He, Chaohai Li, and Chunping Yang. Infrared small target detection via self-regularized weighted sparse model. *Neurocomputing*, 420:124–148, 2021. [2](#), [3](#), [4](#)
- [50] Zhonghao Zhang, Yipeng Liu, Jiani Liu, Fei Wen, and Ce Zhu. Amp-net: Denoising-based deep unfolding for compressive image sensing. *IEEE Transactions on Image Processing (TIP)*, 30:1487–1500, 2020. [3](#), [5](#)
- [51] Mingjing Zhao, Wei Li, Lu Li, Jin Hu, Pengge Ma, and Ran Tao. Single-frame infrared small-target detection: A survey. *IEEE Geoscience and Remote Sensing Magazine (GRSM)*, 10(2):87–119, 2022. [1](#)
- [52] Zihan Zhou, Xiaodong Li, John Wright, Emmanuel Candès, and Yi Ma. Stable principal component pursuit. In *2010 IEEE International Symposium on Information Theory (ISIT)*, pages 1518–1522. IEEE, 2010. [2](#), [3](#)