



# Anomaly Detection in Hyperspectral Images via Regularization by Denoising

Mauro Luiz Brandão Junior , Victor Carneiro Lima , Thomás Antônio Portugal Pereira Teixeira, Eduardo Rodrigues de Lima, and Renato da Rocha Lopes

**Abstract**—In a recent work, Fu et al. (2021) proposed an anomaly detector (AD) for hyperspectral images called DeCNN-AD, which decomposes the image in a low rank representation of the background and the anomalies. DeCNN-AD is regularized by an implicit plug and play prior (PnP), providing state-of-the-art anomaly detection performance in hyperspectral images. In this article, we propose a different AD using the Regularization by Denoising (RED) framework. The regularizer that emerges in RED algorithms is advantageous in the sense that it can be minimized by many solvers, unlike PnP, which has to be rederived for each specific solver like the alternating direction method of multipliers or the proximal gradient method. The proposed detector was compared to DeCNN-AD in several experiments, and was shown to present a more stable behavior better convergence properties, to be less sensitive to the choice of parameters, while leading to very similar performance in terms of optimal AUC of receiver operating characteristic curves.

**Index Terms**—Anomaly detection, hyperspectral images, regularization by denoising.

## I. INTRODUCTION

**H**YPERSPECTRAL anomaly detectors (ADs) aim to find pixels that stand out from the cluttered background [1]. This is an important task in the field of hyperspectral images, as the extra dimension of such data often allows one to distinguish between different materials and phenomena. A relevant application of such automatic ADs is predictive maintenance in electric power substations, in which equipment downtime must be reduced to a minimum.

Classically, it is assumed that anomalies occur rarely when compared to background pixels. Based on this assumption, the Reed–Xiaoli (RX) method and its variants [2], [3], [4], [5] are derived as hypothesis tests on a given statistical background

Manuscript received 28 May 2022; revised 23 July 2022 and 26 August 2022; accepted 11 September 2022. Date of publication 23 September 2022; date of current version 30 September 2022. This work was supported in part by the R&D Project PD-07130-0062/2020 “Predictive Failure Analysis by Artificial Intelligence,” funded by Transmissora Aliança de Energia Elétrica SA (TAESA) with resources from ANEEL R&D Program, and in part by the National Council for Scientific and Technological Development, under Grant 305480/2018-9. (Corresponding author: Mauro Luiz Brandão Junior.)

Mauro Luiz Brandão Junior, Victor Carneiro Lima, Thomás Antônio Portugal Pereira Teixeira, and Renato da Rocha Lopes are with the Communications Department, Faculty of Electrical and Computer Engineering, University of Campinas, Campinas, SP 13083-970, Brazil (e-mail: m228001@dac.unicamp.br; v157460@dac.unicamp.br; thomasportugal5@gmail.com; rlopes@unica mp.br).

Eduardo Rodrigues de Lima is with the Eldorado Institute, University of Campinas, Campinas, SP 13083-898, Brazil (e-mail: eduardo.lima@eldorado.org.br).

Digital Object Identifier 10.1109/JSTARS.2022.3209101

model. Other classical strategies include the collaborative representation detector (CRD) [6], which exploits the fact that background pixels can be approximately represented using its spatial neighbors, and support vector data description (SVDD) detectors [7], [8], which attempt to model the support of the background distribution.

Over the years, several ADs based on pattern recognition using different statistical priors have been proposed [9], [10], [11] with increasingly superior anomaly detection performance. Recently, Fu et al. [12] proposed an AD that decomposes the image as the sum of a low-rank background and an anomaly image. Their method achieves state-of-the-art performance by using the plug and play prior (PnP) framework [13] to regularize the decomposition. Since it uses the deep convolutional neural network (DeCNN) denoiser [14], [15], the method is called the DeCNN-AD.

Regularization by denoising (RED) [16] is another regularization framework that has shown superior performance in tasks such as image super-resolution and image deblurring [16]. As with PnP, RED leverages powerful denoising algorithms for regularization. In contrast with PnP, RED can use several optimization solvers, and, for some denoisers, it has an explicit cost function with provable convergence results [17]. Also, experimental evidence often points to more stable results when using RED when compared to PnP, in the sense that they have better convergence properties and are less sensitive to the choice of parameters for a wide range of denoisers [16].

In this article, we propose a RED algorithm for anomaly detection in hyperspectral images, as an alternative to the state-of-the-art DeCNN-AD [12]. In Section II, we quickly review the anomaly detection task, as well as some well-known detectors. In Section III, we present the proposed detector. Simulation results comparing our proposal to DeCNN in real hyperspectral scenes are presented in Section IV, which show the improved convergence and robustness of RED. Section V concludes this article.

## II. PREVIOUS WORK

AD for hyperspectral images attempt to differentiate background pixels from anomalous ones in a given hyperspectral scene without any prior knowledge of targets or background statistics [1].

Let  $\mathbf{P} \in \mathbb{R}^{N_f \times N_x \times N_y}$  be a matrix representing the hyperspectral data cube, where  $N_f$  is the number of bands, and  $N_x$  and  $N_y$

correspond to the size of the image in each band. These images have been stacked into a vector to form each row of  $\mathbf{P}$ . We are interested in classifying each column of  $\mathbf{P}$ , called a pixel, as either a background pixel or an anomaly.

A large class of ADs [9], [10], [12], [18] follow a two-step structure. The first step, decomposition, is usually designed to separate the image into background and anomaly pixels

$$\mathbf{P} = \mathbf{B} + \mathbf{A} \quad (1)$$

where  $\mathbf{B} \in \mathbb{R}^{N_f \times N_x N_y}$  is the background and  $\mathbf{A} \in \mathbb{R}^{N_f \times N_x N_y}$  is the anomaly. Robust principal components analysis (RPCA) [19] is a classical example of such decomposition, which imposes a low-rank constraint on the background matrix  $\mathbf{B}$  and a sparsity constraint on the anomaly  $\mathbf{A}$ . In the second step, a detection rule is applied to the anomaly matrix  $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{N_x N_y})$ . In most cases, the  $i$ th pixel is classified as an anomaly if  $\|\mathbf{a}_i\|$  is larger than a predetermined threshold. In other words, based on the anomaly component  $\mathbf{A}$ , we generate so-called detection map, which is a vector  $\mathbf{t} \in \mathbb{R}^{N_x N_y}$  whose  $i$ th element is given by

$$t_i = \|\mathbf{a}_i\|. \quad (2)$$

If the decomposition step was successful, after normalizing  $\mathbf{t}$  to the interval  $[0,1]$ , the value of  $t_i$  can be interpreted as the probability that the pixel at index  $i$  is anomalous. A threshold value for  $\mathbf{t}$  can be found manually or heuristically [20]. Pixels with values of  $t_i$  larger than this threshold are declared anomalous.

More sophisticated detectors rely on a low-rank representation of the background under a given dictionary,  $\mathbf{B} = \mathbf{D}\mathbf{S}$ , where the dictionary  $\mathbf{D} \in \mathbb{R}^{N_f \times m}$  should be able to generalize background pixels [21]. Examples of such detectors include the low rank and sparse representation (LRASR) [9], low-rank and collaborative representation (LRCRD) [10], graph and total variation regularized low rank representation (GTVLRR) [18], spectral unmixing and low-rank decomposition (SULRD) [22], [23]. Essentially, the decomposition steps in these detectors seek to minimize different penalty functions in the background representation  $\mathbf{S} \in \mathbb{R}^{m \times N_x N_y}$ , while requiring that the anomaly component be sparse. A detailed review of these detectors can be found in [12].

More recently, Fu et al. [12] proposed the DeCNN-AD AD, which substitutes the penalty functions of [9], [10], [11] with a plug and play prior (PnP) [13]. More precisely, DeCNN-AD aims to solve the following problem

$$\min_{\mathbf{S}, \mathbf{A}} \frac{1}{2} \|\mathbf{P} - \mathbf{D}\mathbf{S} - \mathbf{A}\|_F^2 + \lambda \|\mathbf{A}\|_{2,1} + \beta \phi_{\text{PnP}}(\mathbf{S}) \quad (3)$$

where  $\beta, \lambda > 0$  are regularization parameters, the first term ensures that  $\mathbf{P} \approx \mathbf{D}\mathbf{S} + \mathbf{A}$ ,  $\phi_{\text{PnP}}$  is the implicit PnP prior, and the minimization of  $\|\mathbf{A}\|_{2,1} = \sum_{i=1}^{N_x N_y} \|\mathbf{a}_i\|^2$  ensures that  $\mathbf{A}$  has few columns with nonnegligible norm. As is usual for PnP algorithms, the cost function (3) is minimized with the alternating direction method of multipliers (ADMM) [24], where the minimization step with respect to the regularization penalty  $\phi_{\text{PnP}}$  is replaced by a denoising operation. Fu et al. [12], used the convolutional network FFDNet image denoiser [15], which is sequentially applied to each (normalized) hyperspectral band.

---

### Algorithm 1: DeCNN-AD Decomposition.

---

**Input:**  $\mathbf{P}, \mathbf{D}, \beta, \lambda, \mu_0, \mu_{\max}, \rho, n$

**Output:**  $\mathbf{S}, \mathbf{A}$

---

```

1: for  $k = 0$  to  $n$  do
2:    $\mathbf{S}_{k+1} \leftarrow (\mathbf{D}^T \mathbf{D} + \mu_k \mathbf{I})^{-1} (\mathbf{D}^T (\mathbf{P} - \mathbf{A}_k) + \mu_k \mathbf{Z}_k + \mathbf{V}_k)$ 
3:    $\mathbf{A}_{k+1} \leftarrow \text{prox}_{\lambda \|\cdot\|_{2,1}}(\mathbf{P} - \mathbf{D}\mathbf{S}_{k+1})$ 
4:    $\mathbf{Z}_{k+1} \leftarrow f_{\beta/\mu_k}(\mathbf{S}_{k+1} - \mathbf{V}_k/\mu_k)$ 
5:    $\mathbf{V}_{k+1} \leftarrow \mu_k (\mathbf{Z}_{k+1} - \mathbf{S}_{k+1})$ 
6:    $\mu_{k+1} \leftarrow \min\{\rho\mu_k, \mu_{\max}\}$ 
7: end for
8: return  $\mathbf{S}_k, \mathbf{A}_k$ 

```

---

The complete ADMM solution of the DeCNN-AD decomposition (3) is given by Algorithm 1.

After the decomposition step, DeCNN-AD classifies each pixel of the observed image as an anomalous pixel or a background pixel, based on the norm  $t_i$  of the anomaly component of the corresponding pixel, as discussed earlier. More details on the implementation of DeCNN-AD were presented in [12].

### III. REDAD

The proposed detector, called the Regularization by Denoising Anomaly Detector (REDAD), replaces the implicit PnP prior by the explicit RED prior of [16], leading to the following optimization problem

$$\min_{\mathbf{S}, \mathbf{A}} \frac{1}{2} \|\mathbf{P} - \mathbf{D}\mathbf{S} - \mathbf{A}\|_F^2 + \lambda \|\mathbf{A}\|_{2,1} + \frac{\beta}{2} \text{tr}[\mathbf{S}^T (\mathbf{S} - f_\sigma(\mathbf{S}))] \quad (4)$$

where  $\beta, \lambda > 0$  are regularization parameters and  $f_\sigma$  is a given denoiser assuming noise level  $\sigma$ . The regularizer

$$\phi_{\text{RED}}(\mathbf{S}) = \frac{1}{2} \text{tr}[\mathbf{S}^T (\mathbf{S} - f_\sigma(\mathbf{S}))] \quad (5)$$

contrasts with the implicit one from DeCNN-AD. The main hypothesis of the original RED framework can be rewritten for a matrix denoiser  $f : \mathbb{R}^{M \times N} \rightarrow \mathbb{R}^{M \times N}$  as follows:

- C1) (*Local Homogeneity*)  $f(\epsilon \mathbf{S}) = \epsilon f(\mathbf{S})$  for small  $\epsilon \geq 0$  and for all  $\mathbf{S} \in \mathbb{R}^{M \times N}$ .
- C2) (*Strong Passivity*)  $\eta[J(\mathbf{S})] \leq 1$  for all  $\mathbf{S} \in \mathbb{R}^{M \times N}$ , where  $\eta[\cdot]$  is the spectral radius and  $J(\mathbf{S})$  is the jacobian of the flattened denoiser  $f$  applied to the flattened version of  $\mathbf{S}$ .
- C3) (*Row stochasticity*)  $J(\mathbf{S})$  is entry-wise non-negative and the vector  $(1, \dots, 1)^T \in \mathbb{R}^{M \times N}$  is an eigenvector of  $J(\mathbf{S})$  for any  $\mathbf{S} \in \mathbb{R}^{M \times N}$ .

As shown in [16], under the assumptions (C1)–(C3), the explicit regularizer  $\phi_{\text{RED}}(\mathbf{S})$  is differentiable, convex, and its gradient is given by the denoising residual  $\mathbf{S} - f_\sigma(\mathbf{S})$ . There is an ongoing discussion on the exact conditions that should be assumed for these results to be valid [17], [25]. Despite that, in real applications, RED algorithms converge and achieve excellent performance in different inverse problems even when those conditions are partially satisfied [16]. For this reason, henceforth, we will use the approximation  $\nabla \phi_{\text{RED}}(\mathbf{S}) = \mathbf{S} - f_\sigma(\mathbf{S})$ .

**Algorithm 2:** REDAD Decomposition.**Input:**  $P, D, \beta, \lambda, \mu, \sigma, n$ **Output:**  $S, A$ 


---

```

1: for  $k = 0$  to  $n$  do
2:    $S_{k+1} \leftarrow (D^T D + \mu I)^{-1} (D^T (P - A_k) + \mu Z_k + V_k)$ 
3:    $A_{k+1} \leftarrow \text{prox}_{\lambda \|\cdot\|_{2,1}}(P - D S_{k+1})$ 
4:    $Z_{k+1} \leftarrow \frac{1}{\beta + \mu} [\beta f_\sigma(Z_k) + \mu (S_{k+1} - \frac{V_k}{\mu})]$ 
5:    $V_{k+1} \leftarrow \mu (Z_{k+1} - S_{k+1})$ 
6: end for
7: return  $S_k, A_k$ 

```

---

This simple gradient expression allows us to write first-order optimization methods such as steepest descent (SD), fixed point (FP) schemes, ADMM and block coordinate descent (BCD) in order to minimize the cost function in (4) [17]. For the sake of comparison with the DeCNN-AD in Algorithm 1, we shall describe the ADMM solution of the REDAD decomposition in the remainder of this section. A usual when deriving the ADMM algorithm, we first write the augmented lagrangian

$$\frac{1}{2} \|P - DS - A\|_F^2 + \beta \phi(Z) + \lambda \|A\|_{2,1} + \frac{\mu}{2} \left\| Z - S + \frac{V}{\mu} \right\|_F^2 \quad (6)$$

where  $Z$  is a dummy variable such that  $Z = S$ , and  $V$  is the dual variable. Note that (6) is defined for a generic prior  $\phi$ . By setting  $\phi = \phi_{\text{PnP}}$  we have the lagrangian for the original DeCNN-AD, and by setting  $\phi = \phi_{\text{RED}}$  we have the lagrangian for the REDAD. Therefore, the only difference between the two ADMM solutions is the updating of  $Z$ , which is given by

$$Z^+ \in \arg \min_Z \beta \phi(Z) + \frac{\mu}{2} \left\| Z - S + \frac{V}{\mu} \right\|_F^2. \quad (7)$$

DeCNN-AD replaces this step with a direct application of the denoiser  $f_\sigma$ . In contrast, for the REDAD prior, (7) can be solved explicitly using the gradient approximation. For instance, one iteration of the fixed point method to solve (7) is given by

$$Z^+ = \frac{1}{\beta + \mu} \left[ \beta f_\sigma(Z) + \mu \left( S - \frac{V}{\mu} \right) \right] \quad (8)$$

where  $f_\sigma(Z)$  is a denoised estimate of  $Z$ . Using a single step of this fixed point strategy per outer ADMM iteration, the number of denoiser activations per iteration of both REDAD and DeCNN-AD are the same, and the computational cost of both decompositions are roughly the same.

Following the steps of the DeCNN-AD for updating the other variables  $S$ ,  $A$ , and  $V$ , the complete ADMM solution of the REDAD decomposition (4) is summarized in Algorithm 2.

Comparing Algorithms 1 and 2, we see that unlike the DeCNN-AD, the penalty parameter  $\mu$  and the noise level  $\sigma$  are kept constant during all iterations of the REDAD decomposition. For PnP algorithms, using a fixed penalty  $\mu$  does not seize its full potential [16]. In order to get around this, DeCNN-AD starts with a relatively small  $\mu_0$ , and updates it at each iteration with

the rule

$$\mu_{k+1} = \min\{\rho \mu_k, \mu_{\max}\} \quad (9)$$

where  $\rho = 1 + \epsilon$  is a constant parameter close to but greater than 1, and  $\mu_{\max}$  establishes a maximum amount of penalization. Since the noise level assumed in DeCNN-AD depends on the penalty parameter  $\mu_k$ , this updating provides a large amount of denoising in the beginning iterations, followed by a decreasing amount in the remaining ones. This manually crafted constants express an attempt to induce a convergent behavior in PnP algorithms, and they are left out of Algorithm 2, simplifying parameter selection.

#### IV. EXPERIMENTAL RESULTS

In this section, we present experimental results for the proposed detector applied to different hyperspectral scenes. A false color representation of each scene we used is shown in the first row of Fig. 1. The data was extracted from the Airport-Beach-Urban (ABU) dataset, available at [26], which was manually labeled from cropped images of the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and the Reflective Optical System Imaging Spectrometer (ROSIS-03). The ground-truth maps are shown in the second row of Fig. 1. Some of these scenes were also used by Fu et al. [12] in the original DeCNN-AD paper, with slightly different names.

In all experiments, REDAD and DeCNN-AD used the same dictionary construction heuristic proposed by Fu et al. [12] with parameters  $T = 3$ ,  $M = 0.3$  and  $P = 20$ . The number of clusters  $K$  was set differently for each scene. In all cases,  $f$  was chosen to be the FFDNet denoiser [15].

##### A. Detection Performance

In this first experiment, we evaluated both REDAD and DeCNN-AD in all the four scenes shown in Fig. 1. For comparison, we also evaluated seven different state-of-the-art ADs, namely Global Reed-Xiaoli (GRX) [2], Kernel Reed-Xiaoli (KRX) [3], [4], [5], Robust Principal Component Analysis (RPCA) [19], LRASR [9], [11], joint LRASR with potential anomaly dictionaries (LRSR) [27], Low-Rank Embedded Network (LREN) [28] and the spectral-spatial anomaly target detector based on fractional Fourier transform, context-aware Saliency, and Collaborative Representation theory (SS-FSCRD) [6], [29].

Since the ground-truth maps are available, performance was measured in terms of the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) for each of the scenes. The ROC curves are shown in Fig. 3, and the final AUC of all methods are presented in Table I. For each scene, we also plotted the anomaly maps of the four methods with higher AUC. The anomaly maps are shown in Fig. 2, with detected anomalies highlighted in red.

For the KRX, we used the Gaussian radial basis function kernel

$$k(x, y) = \exp\left(\frac{-\|x - y\|_2^2}{c}\right) \quad (10)$$

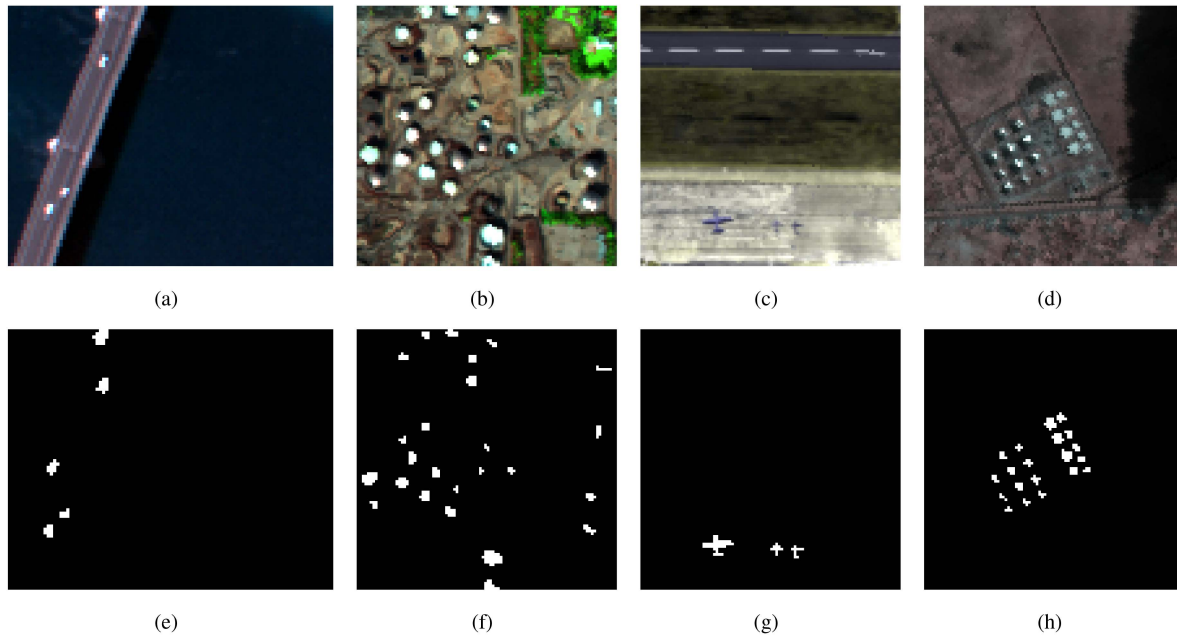


Fig. 1. Datasets used in the experiments. The first row shows false color representations of each hyperspectral scene. The second row shows ground-truth label maps. (a), (e) Pavia Beach. (b), (f) Los Angeles Urban. (c), (g) Gulfport Airport. (d), (h) Texas Coast Urban.

TABLE I  
FINAL AUC SCORES OF DIFFERENT ADS IN FOUR HYPERSPECTRAL SCENES

Detector	Pavia Beach	Los Angeles Urban	Gulfport Airport	Texas Coast Urban
GRX	0.99804	0.98565	0.97782	0.99081
KRX	<i>0.99896</i>	0.99484	0.79992	0.99742
RPCA	0.99779	0.98337	0.96099	0.99154
LRASR	0.99816	<i>0.99564</i>	0.94846	<i>0.99814</i>
LRSR	0.99757	0.98910	0.71201	0.99729
LREN	0.99886	0.97135	0.90194	0.99275
SSFSCRD	0.99495	0.99353	<i>0.99535</i>	0.99591
DeCNN-AD	<i>0.99933</i>	<i>0.99487</i>	<i>0.99530</i>	<b>0.99933</b>
<b>REDAD</b>	<b>0.99953</b>	<b>0.99579</b>	<b>0.99565</b>	<b>0.99933</b>

TABLE II  
REDAD AND DeCNN-AD PARAMETERS IN THE FIRST EXPERIMENT

Detector	Parameter	Pavia Beach	Los Angeles Urban	Gulfport Airport	Texas Coast Urban
DeCNN-AD	$\mu_0$	$10^3$	$10^3$	0.1	$10^2$
	$\rho$	1.1	1.01	1.01	1
	$K$	3	3	7	2
	$\beta$	$10^{-3}$	$10^{-4}$	$10^{-3}$	0.1
	$\lambda$	$10^{-3}$	$10^{-4}$	0.1	0.5
REDAD	$\mu$	$10^4$	$10^3$	1	$10^3$
	$\sigma$	0.1	0.1	0.1	0.1
	$K$	3	3	7	2
	$\beta$	$10^{-3}$	$10^{-4}$	0.7	0.1
	$\lambda$	$10^{-3}$	$10^{-4}$	0.1	0.5

and a dictionary of background pixels clustered with the same heuristic as REDAD and DeCNN-AD with parameters  $K = 3, 1, 9,$  and  $2$  and  $c = 10, 15, 12,$  and  $15$  for the scenes Pavia Beach, Los Angeles Urban, Gulfport Airport and Texas Coast Urban, respectively. The GRX assumes all pixels in each scene are background, which is a reasonable assumption if anomalies

are rare. LRASR also used the same dictionary as DeCNN-AD and REDAD, with the same construction parameters. LRSR and LREN used their own background and anomaly dictionaries as proposed in the original papers. The RPCA parameters were those of the original Candès algorithms [19]. The parameters of DeCNN-AD and REDAD are shown in Table II. The parameters

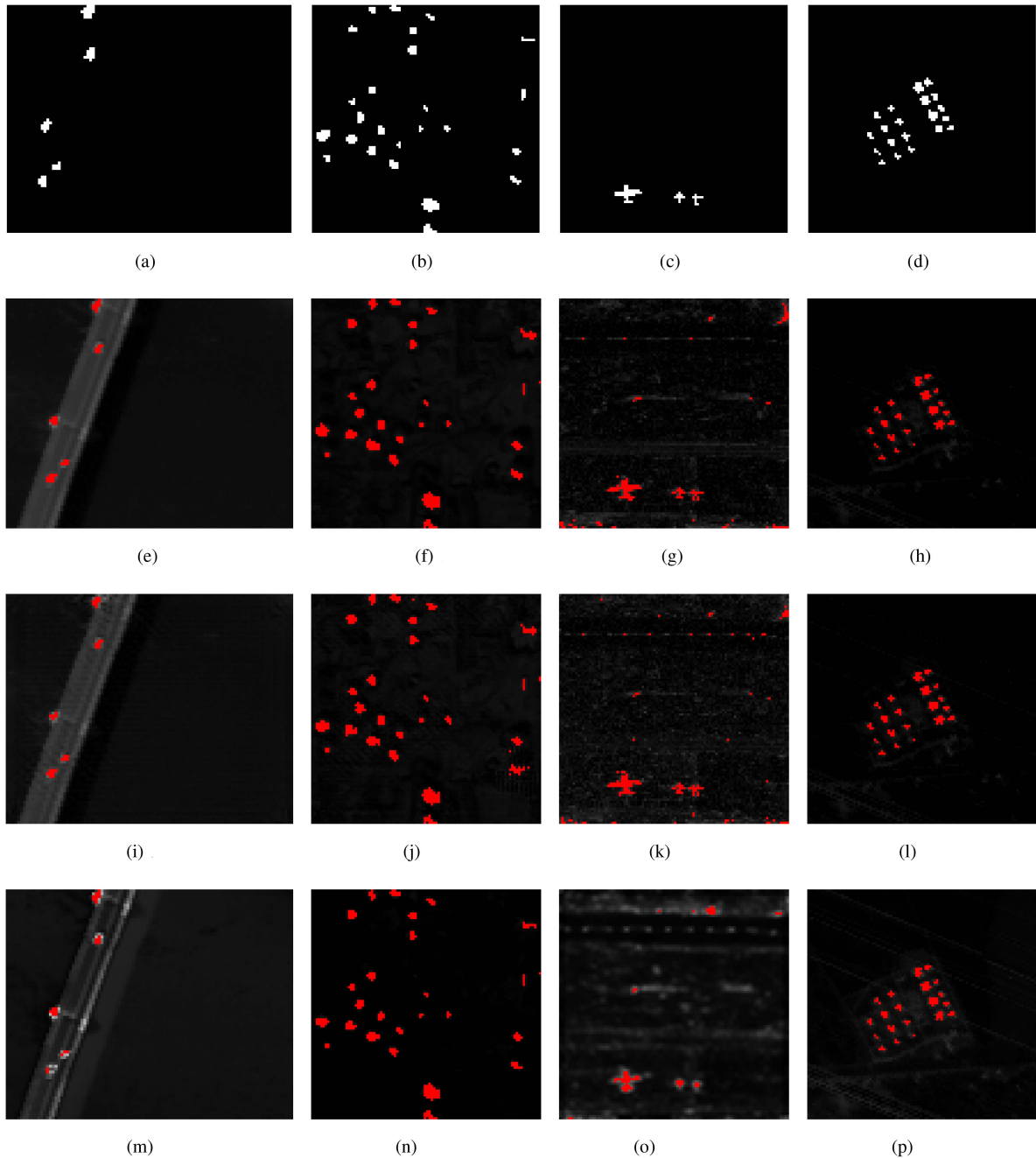


Fig. 2. Anomaly maps with highlighted detections of the three methods with highest AUCs in each scene. The first row shows ground-truth labels. Each column shows the results for different datasets: (a), (e), (i), (m) Pavia Beach; (b), (f), (j), (n) Los Angeles Urban; (c), (g), (k), (o) Gulfport Airport; (d), (h), (l), (p) Texas Coast Urban. (e) REDAD. (f) REDAD. (g) REDAD. (h) REDAD. (i) DeCNN-AD. (j) DeCNN-AD. (k) DeCNN-AD. (l) DeCNN-AD. (m) KRX. (n) LRASR. (o) SSFSCRD. (p) LRASR.

of all methods were set to their default values whenever they were available for the scenes in question. Otherwise, a grid search was performed to find reasonable ones. Whenever applicable, we adopted  $\mu_{\max} = 10^8$  and  $\rho = 1.1$  unless explicitly stated otherwise. Also, the AUC was measured at iteration 200 for all iterative methods.

As shown in Table I, the REDAD detector produced similar, or even better AUC scores than DeCNN-AD and several other state-of-the-art detectors. REDAD produced an AUC close to 1 in all of the hyperspectral scenes tested. More importantly, the

final AUCs presented in Table I corroborate the results provided by [12], which shows that a highly specialized denoiser, such as FFDnet, can be used as a regularizer with outstanding results in the original DeCNN-AD formulation as well as in the RED alternative we are proposing.

In summary, in this section, we compared REDAD with several different state-of-the-art AD methods with different dictionary construction heuristics, regularization engines, and clustering strategies. We observed that REDAD provided similar and possibly better results than the other methods in terms of

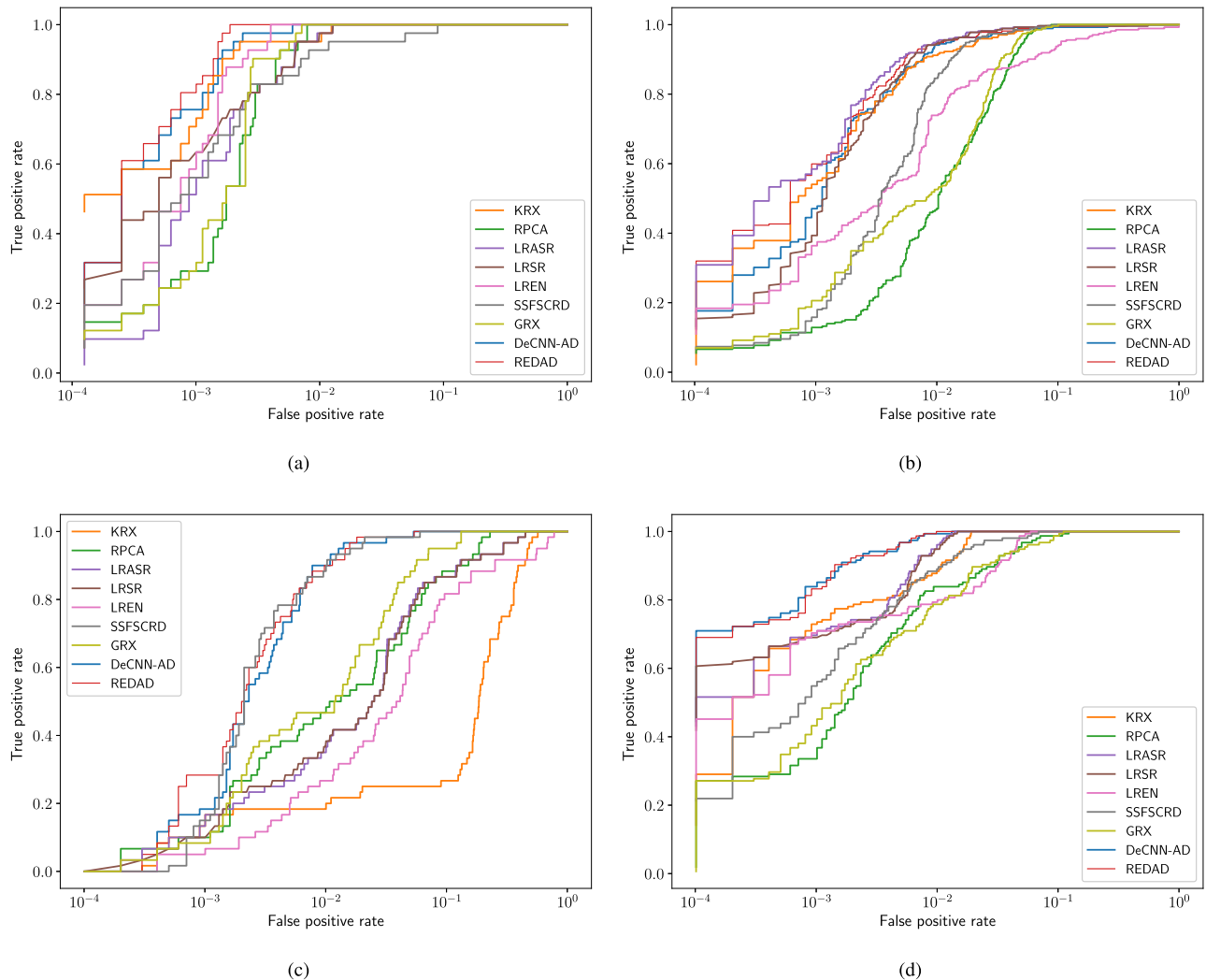


Fig. 3. ROC curves obtained with DeCNN-AD, REDAD, GRX, and KRX in scenes (a) Pavia Center, (b) Texas Coast Urban, and (c) Gulfport. False positive rates are  $\log_{10}$  scaled.

AUC of the ROC curve. Next, we discuss practical aspects of REDAD and DeCNN-AD regarding convergence and the tuning of its parameters.

### B. Robustness to Parameters and Convergence Stability

This second experiment consisted in evaluating REDAD and DeCNN-AD for a wide range of parameters in the Texas Coast Urban scene. This was done in two stages: First, we considered a large set of penalty parameters for both detectors while maintaining regularization parameters  $\beta = 0.1$  and  $\lambda = 0.5$  fixed. Then, we evaluated the detectors for a range of regularization parameters fixing the penalty parameters  $\mu_0 = 100$  and  $\rho = 1.1$  for the DeCNN-AD, and  $\mu = 1000$  and  $\sigma = 0.1$  for the REDAD. The AUC for both detectors, as a function of the iteration, are depicted in Figs. 4 and 5, respectively.

The results for DeCNN-AD, depicted in Figs. 4(a) and 5(a), show that the performance of the detector varies significantly for slight changes in the parameters. The results also indicate that the algorithm frequently diverges after some iterations. On

the other hand, the REDAD curves, depicted in Figs. 4(b) and 5(b), show a more stable behavior with respect to changes in the parameters, and better convergence behavior. Similar curves were obtained for all other scenes, but were omitted here for brevity.

In the previous experiment, we observed a glimpse of one of the main features of RED algorithms: a more stable convergence when compared to PnP. To further clarify, we re-evaluated both detectors in all four hyperspectral scenes. Instead of the AUC, the norm differences

$$\|\mathbf{A}_k - \mathbf{A}_{k-1}\|_F \quad (11)$$

were computed at each iteration. The results are shown in Fig. 6(a) and (b). The parameter choices used here are shown in Table II, except that the number of iterations was set to 1000 instead of 200.

As in the first experiment, both DeCNN-AD and REDAD performed equally well in the first 200 iterations after which, in all the scenes, the detectors reached a norm difference of approximately  $10^{-1}$ . However, after a couple hundred iterations,

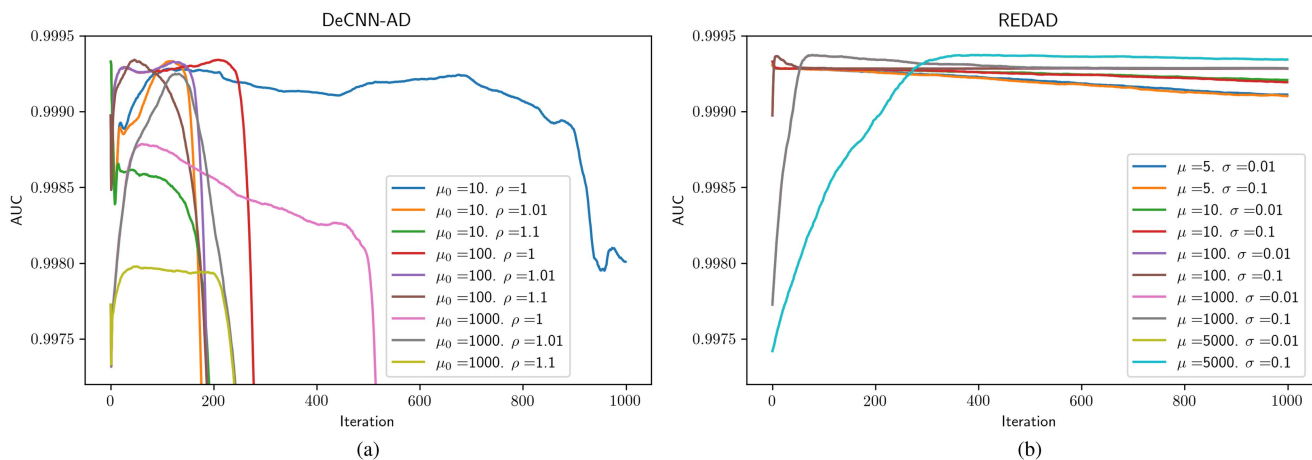


Fig. 4. AUC at each iteration in Texas Coast Urban scene for (a) DeCNN-AD, and (b) REDAD. The curves were obtained with varying penalty parameters, and fixed regularization parameters.

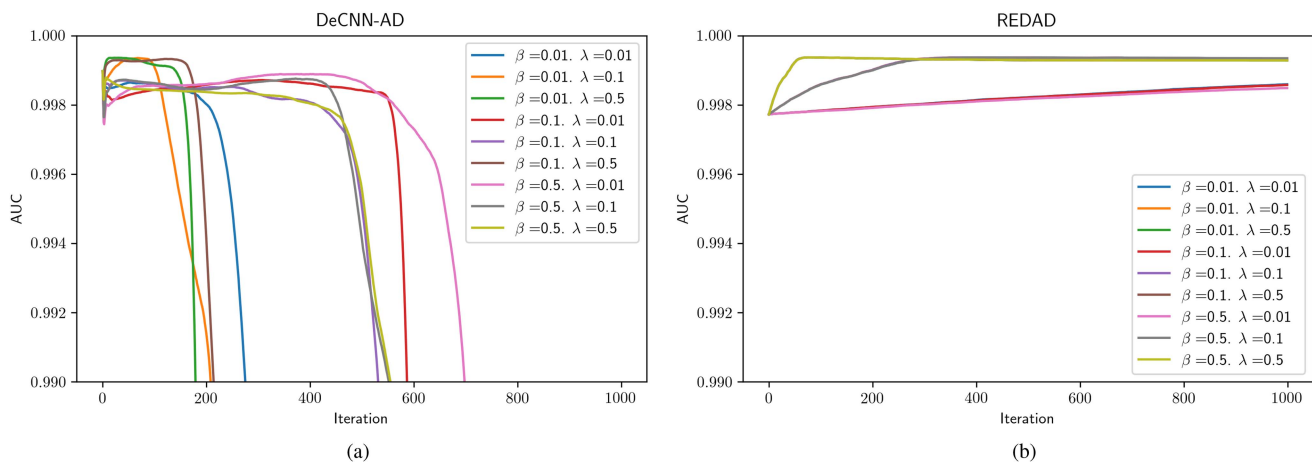


Fig. 5. AUC at each iteration in Texas Coast Urban scene for (a) DeCNN-AD, and (b) REDAD. The curves were obtained with varying regularization parameters, and fixed penalty parameters.

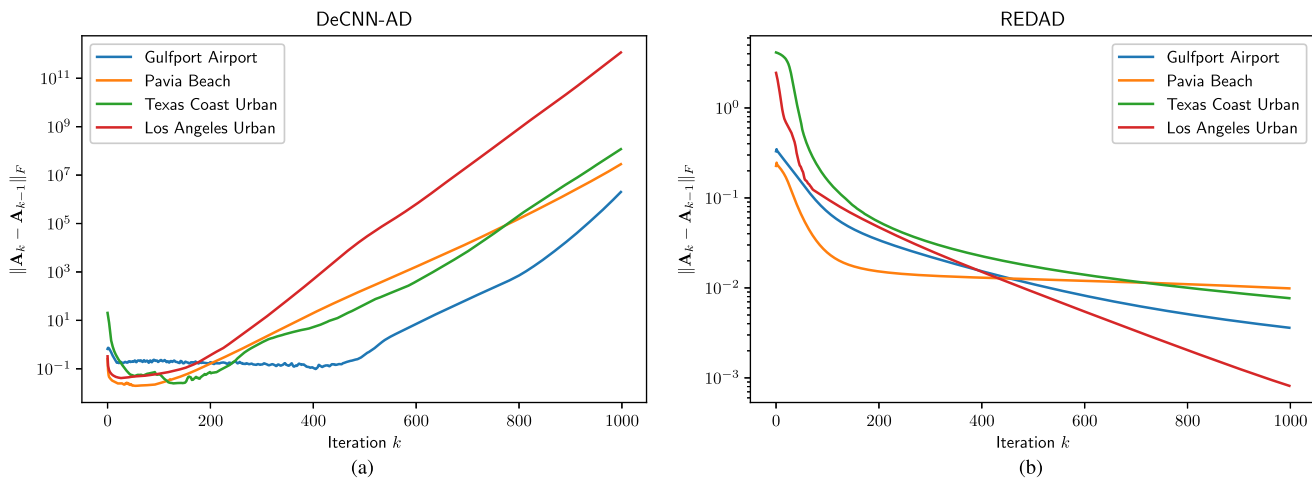


Fig. 6. Norm difference of the anomaly component at each iteration in all hyperspectral scenes for (a) DeCNN-AD, and (b) REDAD.

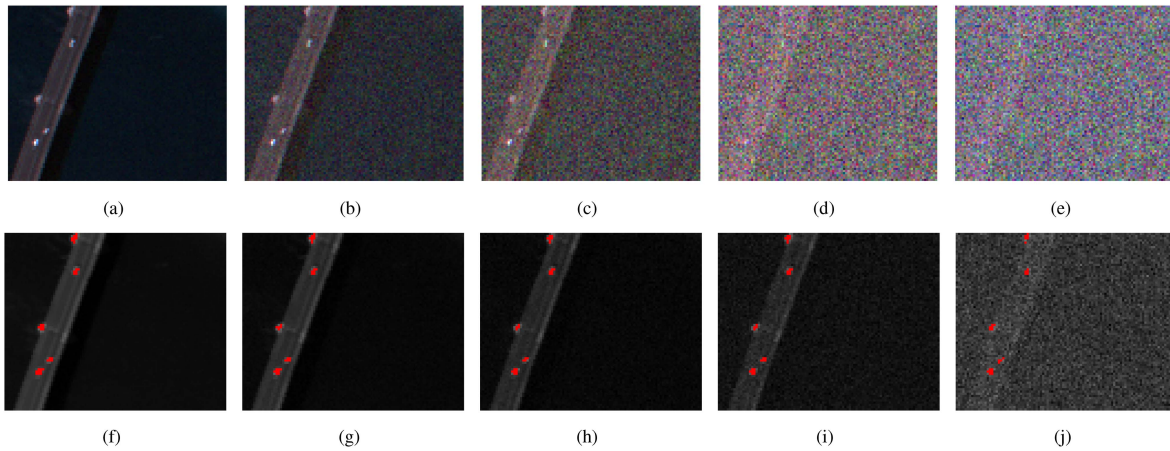


Fig. 7. Pavia Beach scene corrupted by noise. First row shows a false color representation of each scenario. The second row shows the detection maps provided by REDAD in each case, where we highlighted the detected anomalies. (a,f)  $\sigma = 0.01$ . (b,g)  $\sigma = 0.05$ . (c,h)  $\sigma = 0.1$ . (d,i)  $\sigma = 0.2$ . (e,j)  $\sigma = 0.5$ .

the DeCNN-AD diverges, while with REDAD the norm difference seems to get smaller and smaller, further highlighting its superior convergence properties presented by [16]. In fact, the divergence of DeCNN-AD is similar to the convergence behavior reported for other PnP algorithms in [16], which was one of the motivations for their proposal of RED.

### C. Anomaly Detection of Corrupted Data

An important question regarding the use of PnP and RED in anomaly detection is whether those methods perform well in the presence of noise. In this section, we address this topic by repeating the first experiment, but introducing different levels of noise corruption.

For this purpose, the Pavia Beach scene was corrupted with independent additive Gaussian noise, with standard deviation  $\sigma$  ranging from 0.01 to 0.1, resulting in five test scenes. A false color representation of each corrupted test scene is shown in Fig. 7. When the variance increases, it becomes harder to visually distinguish between anomalous pixels and background ones, until a point, as depicted in Fig. 7(d), where the distinction cannot be performed by looking only at the false color representation.

For each noise level, we reran the ADs discussed in the first experiment, using the same set of parameters as before. The detection maps obtained by REDAD are shown in the second row of Fig. 7, where we highlighted the detected anomalies.

The whole experiment was repeated 10 times for each noise level, with new noise samples at each trial. The mean AUCs obtained in these runs are shown in Table III. In the original noiseless Pavia Beach scene, both REDAD and DeCNN-AD had very similar performance, producing an AUC close to 0.999. When noise is added, detection becomes more difficult, and AUC values decrease. From Table III, we see that the REDAD detector was the most robust, and its mean AUC never fell below 0.9691, even in the noisiest case. The mean AUC obtained by DeCNN-AD and LRASR were 0.4870 and 0.7567 at iteration 200. Considering that these two detectors were tied with REDAD in the first experiment, this performance penalty further highlights the stability and robustness of the proposed REDAD.

TABLE III  
MEAN AUC SCORES OF EACH DETECTOR EVALUATED IN THE PAVIA BEACH SCENE UNDER DIFFERENT NOISE LEVELS

$\sigma$	0.01	0.05	0.1	0.3	0.5
GRX	0.9976	0.9911	0.9802	0.92745	0.88648
KRX	0.9991	0.95365	0.9563	0.5212	0.2247
RPCA	0.9956	0.9651	0.9397	0.8733	0.8093
LRASR	0.9981	0.9977	0.9833	0.8053	0.7567
LRSR	0.9981	0.9983	0.9975	0.8662	0.9475
LREN	0.9914	0.9640	0.9639	0.9347	0.8872
SSFSCRD	0.9947	0.9925	0.9894	0.9050	0.8193
DeCNN-AD	0.9992	0.9991	0.9982	0.8968	0.4870
REDAD	<b>0.9995</b>	<b>0.9995</b>	<b>0.9995</b>	<b>0.9978</b>	<b>0.9691</b>

We emphasize that the results in Table III were obtained at iteration 200 of all iterative methods, and that the parameters were the same as in the first experiment. While there may be a set of parameters for which DeCNN-AD achieves similar performance to REDAD, these parameters could be very different from those used previously, and we did not try to fine-tune them. The main point here is that REDAD achieves good performance with no need for returning its parameters, whereas DeCNN-AD suffers a large performance penalty when it is not returned. The performance penalty of DeCNN-AD may be due to the fact that the noise seems to exacerbate the convergence problems of the method. In fact, in the  $\sigma = 0.5$  case, the mean AUC for DeCNN-AD reaches 0.9621 at iteration 100, but degrades to 0.4870 at iteration 200.

The results obtained here reinforce what has already been discussed in the last sections, indicating that REDAD is less sensitive to small perturbations than DeCNN-AD. This analysis extends the comparisons made by Romano et al. [16], pointing to the fact that RED may be a more robust choice than PnP in the problem of hyperspectral anomaly detection.

## V. CONCLUSION

In this article, we presented a novel algorithm for anomaly detection in hyperspectral images. The proposed algorithm

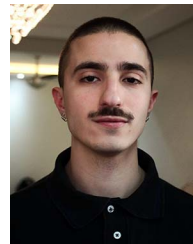


improves on the recently proposed, state-of-the-art DeCNN-AD, by replacing the implicit PnP prior by the explicit RED prior.

The proposed detector was experimentally compared to the DeCNN-AD, leading to similar AUC scores in the ABU dataset. However, the AUC of REDAD was less sensitive to the choice of parameters, and presented a more stable convergence behavior, both desirable features in real applications. These improved properties have also been reported for RED algorithms in other applications like deblurring and super-resolution.

## REFERENCES

- [1] N. M. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, Jan. 2014.
- [2] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.
- [3] H. Kwon and N. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 388–397, Feb. 2005.
- [4] S. Khazai and B. Mojaradi, "A modified kernel-RX algorithm for anomaly detection in hyperspectral images," *Arabian J. Geosci.*, vol. 8, pp. 1487–1495, Mar. 2015.
- [5] J. Zhou, C. Kwan, B. Ayhan, and M. T. Eismann, "A novel cluster kernel RX algorithm for anomaly and change detection using hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6497–6504, Nov. 2016.
- [6] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.
- [7] A. Banerjee, P. Burlina, and R. Meth, "Fast hyperspectral anomaly detection via SVDD," in *Proc. IEEE Int. Conf. Image Process.*, 2007, vol. 4, pp. IV-101–IV-104.
- [8] P. Ma, C. Yao, Y. Li, and J. Ma, "Anomaly detection in hyperspectral image based on SVDD combined with features compression," in *Proc. 5th Int. Conf. Innov. Artif. Intell.*, New York, NY, USA, Association for Computing Machinery, 2021, pp. 103–107.
- [9] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 1990–2000, Apr. 2016.
- [10] H. Su, Z. Wu, A.-X. Zhu, and Q. Du, "Low rank and collaborative representation for hyperspectral anomaly detection via robust dictionary construction," *ISPRS J. Photogrammetry Remote Sens.*, vol. 169, pp. 195–211, 2020.
- [11] Y. Niu and B. Wang, "Hyperspectral anomaly detection based on low-rank representation and learned dictionary," *Remote Sens.*, vol. 8, 2016, Art. no. 289.
- [12] X. Fu, S. Jia, L. Zhuang, M. Xu, J. Zhou, and Q. Li, "Hyperspectral anomaly detection via deep plug-and-play denoising CNN regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9553–9568, Nov. 2021.
- [13] S. V. Venkatakrisnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2013, pp. 945–948.
- [14] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
- [15] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.
- [16] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (RED)," *SIAM J. Imag. Sci.*, vol. 10, pp. 1804–1844, Jan. 2017.
- [17] R. Cohen, M. Elad, and P. Milanfar, "Regularization by denoising via fixed-point projection (RED-PRO)," *SIAM J. Imag. Sci.*, vol. 14, pp. 1374–1406, Jan. 2021.
- [18] T. Cheng and B. Wang, "Graph and total variation regularized low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 391–406, Jan. 2020.
- [19] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM*, vol. 58, no. 3, pp. 1–37, Jun. 2011.
- [20] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *J. Electron. Imag.*, vol. 13, no. 1, pp. 146–165, 2004.
- [21] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
- [22] Y. Qu et al., "Anomaly detection in hyperspectral images through spectral unmixing and low rank decomposition," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 1855–1858.
- [23] Y. Qu et al., "Hyperspectral anomaly detection through spectral unmixing and dictionary-based low-rank decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4391–4405, Aug. 2018.
- [24] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, pp. 1–122, Jan. 2011.
- [25] E. T. Reehorst and P. Schniter, "Regularization by denoising: Clarifications and new interpretations," *IEEE Trans. Comput. Imag.*, vol. 5, no. 1, pp. 52–67, Mar. 2019.
- [26] X. Kang, "Airport-beach-urban dataset," 2020. [Online]. Available: <http://xudongkang.weebly.com/data-sets.html>
- [27] N. Huyan, X. Zhang, H. Zhou, and L. Jiao, "Hyperspectral anomaly detection via background and potential anomaly dictionaries construction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2263–2276, Apr. 2019.
- [28] K. Jiang, W. Xie, J. Lei, T. Jiang, and Y. Li, "LREN: Low-rank embedded network for sample-free hyperspectral anomaly detection," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, pp. 4139–4146.
- [29] C. Zhao, C. Li, S. Feng, N. Su, and W. Li, "A spectral-spatial anomaly target detection method based on fractional fourier transform and saliency weighted collaborative representation for hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5982–5997, 2020.



**Mauro Luiz Brandão Junior** received the B.S. degree in electrical engineering from the State University of Maringá (UEM), Maringá, Brazil, in 2017, and the M.Sc. degree in electrical engineering from the University of Campinas (UNICAMP), Campinas, Brazil, in 2020. He is currently working toward the Ph.D. degree in electrical engineering from the University of Campinas, Campinas, Brazil.

His research interests includes topics on image processing and regularization.



**Victor Carneiro Lima** received the B.Sc. degree in electrical engineering, in 2020, from the University of Campinas (UNICAMP), Campinas, Brazil, where he is currently working toward the M.Sc degree.

Since 2022, he has been a Professor with the State Center for Technological Education "Paula Souza." His research interests includes topics on image processing, signal processing, and embedded systems.



**Thomás Antônio Portugal Pereira Teixeira** received the B.Sc. degree in electrical engineering from the University of Campinas (UNICAMP), Campinas, Brazil, in 2021.

Since 2022, he has been a Data Scientist working with the financial sector. His research interests includes signal processing, optimization, and machine learning.



**Eduardo Rodrigues de Lima** received the bachelor's degree in EE from the University of São Paulo State—UNESP, São Paulo, Brazil, in 1997, the Ph.D. and DEA degrees in telecommunications from the Technical University of València - UPV, Valencia, Spain, in 2016 and 2006, respectively, and Specialization in mobile communication from UPV/Vodafone, in 2004.

He did Vocational Training in general mechanics from SENAI and Electronics Technician. He is R&D Manager of the Exploratory Hardware Design Department at Eldorado Research Institute and Visiting Professor with FEEC/UNICAMP, Campinas, Brazil. He currently, coordinates several R&D projects related do microelectronics, embedded system, wireless communication, smart grid and IoT. He was a visiting Researcher with IMEC and also MCTI/CNPq Fellow in Technological Productivity. His currently main interests are in the implementation (in Silicon, FPGA and Microprocessors) and theoretic aspects of physical layers of wireless and wired communications systems, sensing, IoT and artificial intelligence.



**Renato da Rocha Lopes** received the B.S. and M.Sc. degrees in electrical engineering from the University of Campinas (UNICAMP), Campinas, Brazil, in 1995 and 1997, respectively. He received the M.Sc. degree in mathematics and the Ph.D. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2001 and 2003, respectively.

Since 2003, he has been with the School of Electrical and Computer Engineering, UNICAMP, where he is currently an Associate Professor. His main research interest is in digital signal processing, especially its application to communication systems and to statistic data analysis.