A Deep Unfolded Prior-Aided RPCA Network For Cloud Removal

A Deep Unfolded Prior-Aided RPCA Network For Cloud Removal

Shoaib Imran, Muhammad Tahir *Senior Member, IEEE*, Zubair Khalid *Senior Member, IEEE*, and Momin Uppal *Senior Member, IEEE*

***Abstract*—Clouds, together with their shadows, usually occlude ground-cover features in optical remote sensing images. This hinders the utilization of these images for a range of applications such as earth observation, land-cover classification and urban planning. In this work, we propose a deep unfolded and prior-aided robust principal component analysis (DUPA-RPCA) network for removing clouds and recovering ground-cover information in multi-temporal satellite images. We model these cloud-contaminated images as a sum of low rank and sparse elements and then unfold an iterative RPCA algorithm that has been designed for reweighted ℓ_1 minimization. As a result, the activation function in DUPA-RPCA adapts for every input at each layer of the network. Our experimental results on both Landsat and Sentinel images indicate that our method gives better accuracy and efficiency when compared with existing state of the art methods.**

***Index Terms*—Cloud removal, shadow removal, deep unfolding, inverse problems, RPCA**

I. Introduction

Remote sensing using optical imagery is crucial for important applications including, but not limited to, urban informatics, crop mapping, building footprint extraction, and land cover classification and change detection. Given that the average global cloud cover in an year is estimated to be about 68% [1], about one-third of optical satellite images are affected by some form of cloud-cover [2]. The standard method for obtaining cloud-free images, e.g. on Google Earth Engine, is to first remove the clouds in individual images of the same area taken over a period of time, usually an year, and then reconstruct the missing region by taking an average of these cloud-free images. However, the average image obtained does not reflect the land-cover changes beneath the clouds on individual dates, which limits its utilization for the applications mentioned above. Consequently, efficient and accurate recovery of ground-cover information using fewer past images is of significant importance in remote sensing.

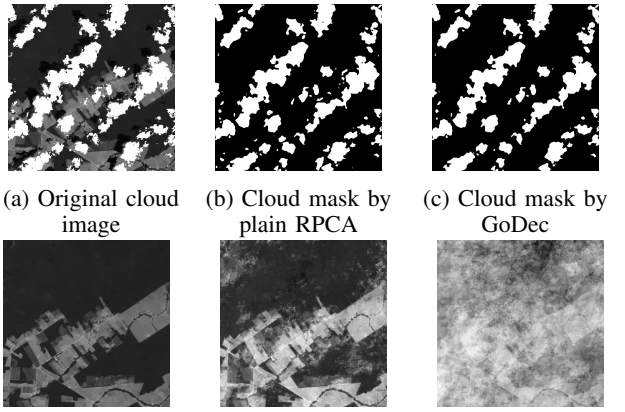
Another important consideration is the impact of cloud thickness on image contamination. Thin clouds only partially occlude ground-cover such that part of ground-cover information can still be seen, and this problem is fairly easy to formulate and solve. On the other hand, reconstructing ground information in the presence of thick clouds is challenging as they completely block the ground cover beneath it. In this letter, we focus on removing thick clouds and their shadows from a time-series of optical images contaminated by clouds.

Current cloud removal methods can be classified into spatial, spectral and temporal methods. Spatial methods involve detecting and removing clouds in individual images and consist of two steps: cloud removal and missing region restoration. The most popular method in this category, is the Fmask algorithm [3]. However, it uses information in bands outside of the visible spectrum to remove clouds. There are few other notable spatial methods in the literature. As an example, the method proposed by Cheng et al. [4], first estimates the land-cover pattern of the known regions and then reconstructs the missing regions by combining a series of pixels from these known regions. On the other hand, Zheng et al. [5] use a generative adversarial network (GAN) for both removal of thick clouds and image restoration. However, spatial methods face difficulty in the reconstruction of large missing regions with reasonable precision.

Spectral methods, on the other hand, use multi-spectral sensing data for cloud removal. For instance, Zhu et al. [2] used Synthetic Aperture Radar (SAR) optical images for reconstructing the missing region. However, these methods struggle to reconstruct missing region information underneath thick clouds as spectral bands information becomes scarce amidst thick clouds. In addition, the multispectral images are typically available at a lower resolution than those available in the visible spectrum.

More pertinent to our work, temporal methods reconstruct missing information beneath the clouds using multi-temporal remote sensing images of the same area. For instance, Tseng et al. [6] exchange cloud pixels with cloud free pixels from previous images of the same area and eliminate the seam artifacts around the boundaries of clouds through multiscale Wavelet fusion together with color matching. The current state-of-the-art multi-temporal method [7] uses a group-sparse robust principle component analysis (RPCA) for initial detection of clouds and shadow regions, which become the balance values for the cloud-contaminated and cloud-free pixels in the discriminative RPCA. However, these methods are sensitive to large clouds and land-cover changes and are not computationally efficient.

To overcome the aforementioned disadvantages, we propose DUPA-RPCA, an RPCA deep network architecture that results from unfolding an iterative algorithm for solving a re-weighted ℓ_1 minimization. It gives better accuracy and efficiency for cloud removal and ground-cover reconstruction in remote sensing image sequences heavily contaminated by clouds while using only the bands in the visible spectrum and outperforming state-of-the-art approaches.



(a) Original cloud image (b) Cloud mask by plain RPCA (c) Cloud mask by GoDec

(d) Ground truth of cloud-free image (e) Recovered image by plain RPCA (f) Recovered image by GoDec

Fig. 1: RPCA cloud separation results

II. Background on RPCA

A. Optimization-Based Methods for RPCA

RPCA has become a popular choice in a range of applications involving outliers detection. With the data arranged as the columns of a matrix, RPCA separates the data matrix D into a low-rank matrix L and a sparse matrix S by solving [8]

$$\min_{L,S} \quad \|L\|_* + \lambda\|S\|_1 \quad \text{s.t. } D = L + S, \tag{1}$$

where $||.||_1$ is the ℓ_1 norm, $||.||_*$ is the nuclear norm and λ is a tuning regularizer. In the context of the cloud removal problem, the multi-temporal ground-cover images can be modeled by a low-rank representation whereas the transient clouds and their shadows can be modelled as sparse outliers.

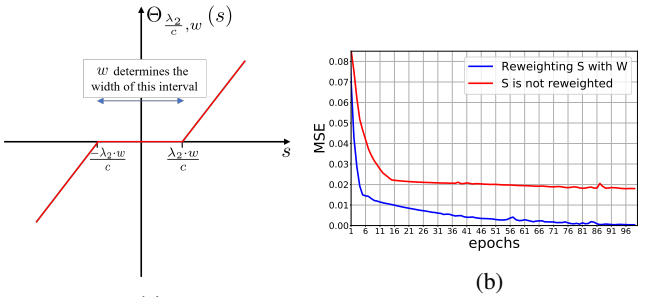
Fig. 1 shows the visual cloud separation results of RPCA and of its computationally efficient variant, GoDec [9]. We observe that both the plain RPCA algorithm and its variant identify the clouds reasonably well but they miss out on the cloud shadows. More importantly, even though the plain RPCA reconstructs the background image, the results are far from desirable. Consequently, our goal is to efficiently and accurately recover the ground-cover information while correctly identifying the clouds and their accompanying shadows.

III. PA-RPCA

A. The Proposed Prior-Aided Model

Consider a temporally varying sequence of remote sensing images, d_i for time instances $= 1, 2, \cdots, q$, contaminated by clouds and each of size $m \times n$ having their low-rank and sparse components denoted as l_i and s_i, respectively. We arrange the image sequence d_i into a matrix $D \in \mathbb{R}^{p\times q}$ where $p = mn$. Similarly, for image sequences l_i and s_i, we obtain their corresponding matrices $L \in \mathbb{R}^{p\times q}$ and $S \in \mathbb{R}^{p\times q}$. Inspired by [10], we consider two measurement matrices H_1 and H_2 of appropriate dimensions, which relate the observations in D with the unknown quantities (here $H_1 = H_2 = I$) such that the acquired D matrix is composed as

$$D = H_1 L + H_2 S + N. \tag{2}$$



(a) (b)

Fig. 2: (a) w (element of matrix W) leads to a distinct proximal operator for the corresponding s (element of S) (b) average mean square error (MSE) plot with and without W

We further observe that both the plain RPCA and GoDec give good enough masks of clouds. Moreover, on our dataset of cloud contaminated image sequences, composed of 20 images of the same area in which each image is of 400 x 400 pixels, we found that the plain RPCA takes 9.709 seconds per image on average as compared to its computationally efficient variant, GoDec, which takes just 0.0326 seconds. We therefore propose reweighting the elements of S with matrix $W = \sigma(\rho, \hat{W})$, where $\sigma(.)$ is the sigmoid function with a gain of ρ and $\hat{W}$ is the land-cover mask, obtained by inverting the clouds mask from GoDec. We formulate the proposed prior-aided RPCA (PA-RPCA) in a Lagrangian form as

$$\min_{L,S} \quad \frac{1}{2}\|D-(H_1L + H_2S)\|_F^2 + \lambda_1\|L\|_* + \lambda_2\|W \circ S\|_1, \tag{3}$$

where $||.||_F$ is the Frobenius norm, λ_1 and λ_2 are the tuning regularizers and "$\circ$" denotes the Hadamard product. PA-RPCA directly leverages the output from GoDec as the prior matrix W leads to a distinct proximal operator for each entry of the input, making PA-RPCA more adaptive to input data as shown in Fig. 2. Furthermore, ρ controls the extent to which $\hat{W}$ changes the the length of the thresholding interval.

B. GoDec RPCA: Getting the Land-Cover Mask

We use GoDec method to obtain the land-cover mask $\hat{W}$. GoDec is a non-convex variant of RPCA which is formulated as [9]

$$\begin{aligned} \min_{L,S} \|D - (L + S)\|_F^2 \\ \text{s. t. } \text{rank(L)} \leq \text{r}, \\ \text{card}(S) \leq k \end{aligned} \tag{4}$$

where card(.) denotes the cardinality of a set, L denotes the low-rank matrix and S denotes the sparse component. GoDec puts a direct constraint on the rank of L and on $\|S\|_0$ while RPCA instead minimizes their convex hulls $\|L\|_*$ and $\|S\|_1$. In particular, GoDec alternately allocates the sparse approximation of $D - L$ and rank r approximation of $D - S$ to S and L, respectively. Furthermore, bilateral random projection based low-rank estimation is used to significantly accelerate the rank-r approximation of $D - S$ under power scheme

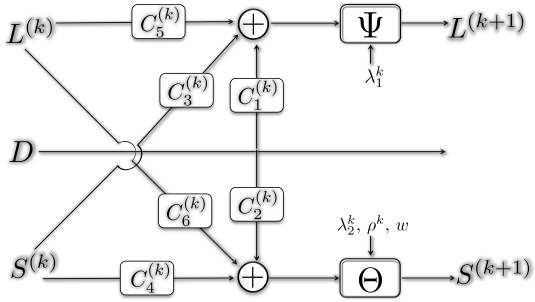


Fig. 3: Unfolded DUPA-RPCA architecture

modification. Compared with the nuclear norm minimization methods, GoDec is significantly more efficient because it does not require time consuming SVD for D. At last, we set an empirical threshold value to attain the mask from the sparse component S of GoDec RPCA.

C. Solving PA-RPCA

ISTA method [11] is used to solve the minimization problem in (3). The general iterative step of ISTA is

$$Y^{k+1} = \text{prox}\left(Y^k - \frac{1}{c}\nabla(Y^k)\right), \tag{5}$$

where Y can be either L or S, c denotes the Lipschitz constant, prox denotes the proximal operator [12] and $\nabla(.)$ denotes the gradient of the quadratic part of (3). The low-rank component L is computed using singular value thresholding $\Psi\{.\}$ [13], and the sparse component is computed using soft-thresholding $\Theta\{.\}$ [11]. Algorithm 1 summarizes the iterative steps for solving (3).

Algorithm 1: The Proposed Algorithm

Input: D, w *(element of matrix W)*, H_1, H_2, λ_1,λ_2 and maximum number of iterations K

Output: L, S

Initialisation: $S = L = 0$ and $k = 1$

1: **for** $k = 1$ to K **do**
2: $F_1^k = \left(I - \frac{1}{c}H_1^T H_1\right) L^k - H_1^T H_2 S^k + H_1^T D$
3: $F_2^k = \left(I - \frac{1}{c}H_2^T H_2\right) S^k - H_2^T H_1 L^k + H_2^T D$
4: $L^{k+1} = \Psi_{\lambda_1/c}(F_1^k)$
5: $S^{k+1} = \Theta_{\lambda_2/c,w}(F_2^k)$
6: **end for**
7: **return** L^K, S^K

IV. The DUPA-RPCA

A. The DUPA-RPCA Network Architecture

The percentage of cloud cover can vary from no cloud at all to full cloud coverage such that an increased number of iterations would be needed to get a good separation of L and S. This drives the need for an algorithm of fixed complexity that can converge in fewer iterations while giving accurate reconstruction results. Hence, following the principles of deep unfolding [10], we unroll the iterations of Algorithm 1, by substituting all the matrices that are a function of H_1 and H_2 with convolutional layers having kernels $C_1^k....C_6^k$, all of which are learned per layer during the training phase. PA-RPCA is unfolded into a K_o-layered ($K_o << K$) DUPA-RPCA network whose architecture (only k^{th} layer) is shown in Fig. 3. The k^{th} iteration in PA-RPCA corresponds to the k^{th} layer in DUPA-RPCA. At each layer, L and S is updated as

$$L^{k+1} = \Psi_{\lambda_1^k}\left\{C_5^k * L^k + C_3^k * S^k + C_1^k * D\right\}, \tag{6}$$

$$S^{k+1} = \Theta_{\lambda_2^k,\rho^k,\hat{w}}\left\{C_6^k * L^k + C_4^k * S^k + C_2^k * D\right\}, \tag{7}$$

where $*$ denotes the convolution operator. In Fig. 3, λ_1^k, λ_2^k and ρ^k are the tuning parameters that are learned per layer during the training phase together with the kernels $C_1^k....C_6^k$.

In our case, DUPA-RPCA is composed of 10 layers ($K_o = 10$). For our experiments, we choose kernels of size (5, 5, 1) with strides (1, 1, 1), padding (2, 2, 0) and a bias for the first three layers and kernels of size (3, 3, 1) with strides (1, 1, 1), padding (1, 1, 0) and a bias for the remaining seven layers. Finally, the loss function is chosen as

$$\mathcal{L}(\Upsilon) = \frac{1}{2N}\sum_{i=1}^{N}\left\|L_i - \overline{L}_i\right\|_F^2 + \frac{1}{2N}\sum_{i=1}^{N}\left\|S_i - \overline{S}_i\right\|_F^2 \tag{8}$$

where N is the total number of training sequences in the dataset, $\left\{\overline{L}_i, \overline{S}_i\right\}_{i=1}^N$ are the low-rank and the sparse components predicted by the network respectively, $\{L_i, S_i\}_{i=1}^N$ are the ground-truths of the land-cover and clouds and their shadows respectively, and $\Upsilon = \left\{C_1^k, \cdots, C_6^k, \lambda_1^k, \lambda_2^k\, \rho^k\right\}_{k=1}^{K_o}$ are the parameters to be learnt by the network.

B. Experiments

To demonstrate the performance of DUPA-RPCA, we choose Sentinel-2 Surface Reflectance and Landsat-8 Surface Reflectance images downloaded from Google Earth Engine with the image size kept at 400 × 400 pixels. We first make 5 cloud-free images in which each image is constructed by taking mean of images of the same area over an year. The time-period, over which we take the mean, does not overlap. Random convex combinations of these 5 images are then used to generate 15 more images. These 20 images represent multi-temporal (images of the same area taken at different times) cloud-free images. All these images are vectorized and then stacked as vectors, with the i^{th} image vector denoted by $\hat{l}^i \in \mathbb{R}^{160000}$, to construct the matrix $\hat{L} \in \mathbb{R}^{160000\times 20}$, having a rank of 5. We further add zero mean and a low variance truncated (between 0 and +1) Gaussian noise to every column of $\hat{L}$ to construct the low-rank matrix L.

Moreover, to create the sparse matrix $S \in \mathbb{R}^{160000\times 20}$, cloud-cover is cropped from cloud-contaminated images. Superimposing these cropped clouds on the low-rank ground-cover matrix L gives the data matrix D, having the same dimensions as L and S. We create 450 such sequences of cloud-contaminated remote sensing images D and their corresponding L and S matrices, all normalized to unit range. Out of the 450 sequences, 400 and 50 sequences are used

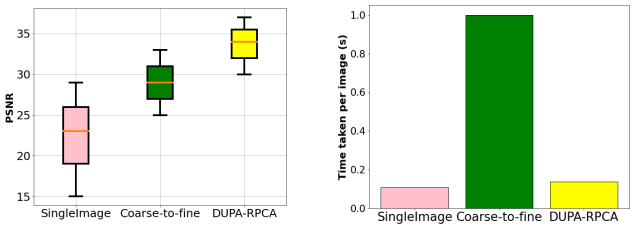


Fig. 4: Performance of DUPA-RPCA

for training and testing respectively. We have also made the dataset available online [14].

DUPA-RPCA is trained for 100 epochs using Adam optimizer with the learning rate of 2×10^{-5}, and the batch size is kept at 16. We further trained DUPA-RPCA on an experimental dataset of 70 image sequences in which the duration between two consecutive images in the L matrix varied from one cycle period[1] to several cycle periods. S and the corresponding D matrices were generated in the same way as before.

C. Results and Analysis

In this section, we compare DUPA-RPCA with the spatial method of Single Image [5] and the multi-temporal RPCA based method of Coarse-to-fine [7]. Fig. 4 shows the performance of these algorithms based on peak signal-to-noise ratio (PSNR) together with the time taken by these algorithms. These results are obtained on 15 image sequences consisting of both Landsat-8 and Sentinel-2 images and prepared in a similar manner as the additional 70 images on which we trained our network. The cloud-cover ratio in these images ranged from 30% to 70% which is more consistent with the actual temporally varying cloud contaminated images. DUPA-RPCA has achieved the highest median PSNR while taking significantly less time than the other multi-temporal method and just marginally more as compared to the spatial method of Single Image.

Single Image, being a spatial method, involves removing clouds and restoring ground cover information in individual images by first detecting the cloud using U-Net and then reconstructing the missing region using GAN. However, the generator is prone to filling the cloud regions with unrealistic images if the cloud cover ratio is high, which is often the case. The Coarse-to-fine method, on the other hand, processes a batch of images. It also uses a two-stage process by first detecting clouds with group-sparse RPCA and then using a discriminative RPCA to further facilitate cloud removal and more importantly restore the missing region. However, the PSNR values of reconstructed regions drop significantly when all of the images in the sequence have a high cloud cover ratio as is often the case when the data matrix is composed of consecutive images with a time interval of not more than one cycle between them. This problem could be solved by incorporating more contextual information.

[1]Landsat-8 and Sentinel-2 revisit a place after every 16 days and 5 days respectively.

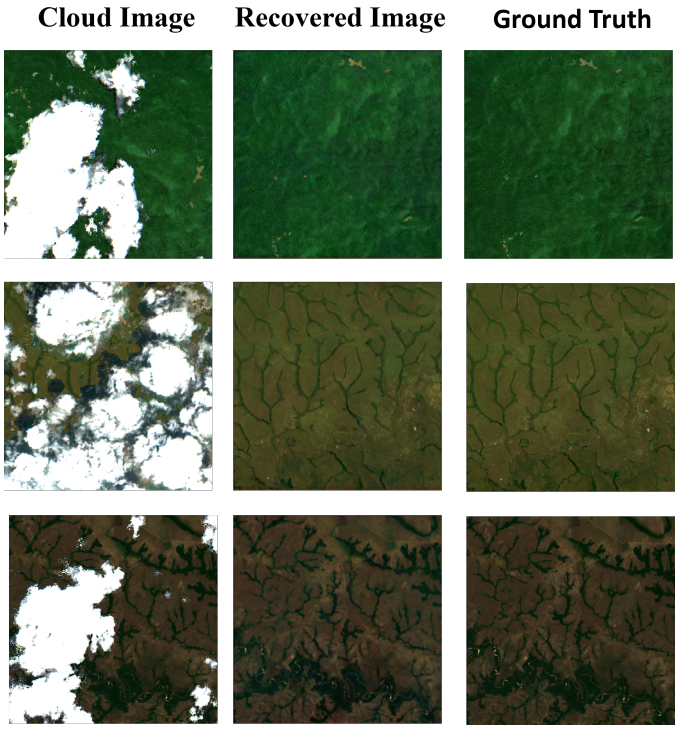


Fig. 5: Recovered cloud-free images by DUPA-RPCA

In our method, this contextual information is incorporated by way of an end-to-end trained model based deep neural network, DUPA-RPCA, which learns all of its parameters from the data itself. Fig. 5 shows the visual reconstruction results by DUPA-RPCA. Reweighting the sparse component with W leads to a distinct proximal operator for each entry of the input such that DUPA-RPCA achieves much better accuracy with the reweighting scheme, as shown in Fig. 2b. Furthermore, at each layer, the S matrix from previous layer is one of the inputs for computing L such that a better estimation of S results in an improved estimation of L.

V. Conclusion

In this paper, we proposed a deep unfolded and prior-aided RPCA network for two-fold functionality: removing clouds and recovering ground-cover information in remote sensing images. Our design resulted in a distinct proximal operator at each iteration/layer for every input making this model-aided network highly adaptive such that the results from DUPA-RPCA have attained much better accuracy and efficiency than the existing state-of-the-art algorithms. The proposed DUPA-RPCA utilizes the high temporal correlation between the images which can also be modified to incorporate the spatial correlation between the images in future work.

References

[1] Claudia Stubenrauch, William B. Rossow, Stefan Kinne, and Steve Ackerman. Gewex cloud assessment: A review. In *Radiation Processes in the Atmosphere and Ocean, IRS 2012 - Proceedings of the International Radiation Symposium (IRC/IAMAS)*, AIP Conference Proceedings, pages 404–407, 2013.

[2] Cheng Zhu, Zhiqin Zhao, Xiaozhang Zhu, Zaiping Nie, and Qing Huo Liu. Cloud removal for optical images using sar structure data. In *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, pages 1872–1875, 2016.

[3] Zhe Zhu, Shixiong Wang, and Curtis E. Woodcock. Improvement and expansion of the fmask algorithm: cloud, cloud shadow, and snow detection for landsats 4–7, 8, and sentinel 2 images. *Remote Sensing of Environment*, 159:269–277, 2015.

[4] Qing Cheng, Huanfeng Shen, Liangpei Zhang, and Zhenghong Peng. Missing information reconstruction for single remote sensing images using structure-preserving global optimization. *IEEE Signal Processing Letters*, 24(8):1163–1167, 2017.

[5] Jiahao Zheng, Xiao-Yang Liu, and Xiaodong Wang. Single image cloud removal using u-net and generative adversarial networks. *IEEE Transactions on Geoscience and Remote Sensing*, 59(8):6371–6385, 2021.

[6] Din-Chang Tseng, Hsiao-Ting Tseng, and Chun-Liang Chien. Automatic cloud removal from multi-temporal spot images. *Applied Mathematics and Computation*, 205(2):584–600, 2008. Special Issue on Advanced Intelligent Computing Theory and Methodology in Applied Mathematics and Computation.

[7] Yongjun Zhang, Fei Wen, Zhi Gao, and Xiao Ling. A coarse-to-fine framework for cloud removal in remote sensing image sequence. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8):5963–5974, 2019.

[8] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *J. ACM*, 58(3), Jun 2011.

[9] Tianyi Zhou and Dacheng Tao. Godec: Randomized low-rank & sparse matrix decomposition in noisy case. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11, page 33–40, Madison, WI, USA, 2011. Omnipress.

[10] Oren Solomon, Regev Cohen, Yi Zhang, Yi Yang, Qiong He, Jianwen Luo, Ruud J. G. van Sloun, and Yonina C. Eldar. Deep unfolded robust pca with application to clutter suppression in ultrasound. *IEEE Transactions on Medical Imaging*, 39(4):1051–1063, 2020.

[11] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[12] Zhao Tan, Yonina C. Eldar, Amir Beck, and Arye Nehorai. Smoothing and decomposition for analysis sparse recovery. *IEEE Transactions on Signal Processing*, 62(7):1762–1774, 2014.

[13] Jian-Feng Cai, Emmanuel J. Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

[14] The dataset available online. https://github.com/convexoptimist/dupa-rpca-cloud-removal.