

A Deep-Unfolded Spatiotemporal RPCA Network For L+S Decomposition

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

Abstract—Low-rank and sparse decomposition based methods find their use in many applications involving background modeling such as clutter suppression and object tracking. While Robust Principal Component Analysis (RPCA) has achieved great success in performing this task, it can take hundreds of iterations to converge and its performance decreases in the presence of different phenomena such as occlusion, jitter and fast motion. The recently proposed deep unfolded networks, on the other hand, have demonstrated better accuracy and improved convergence over both their iterative equivalents as well as over other neural network architectures. In this work, we propose a novel deep unfolded spatiotemporal RPCA (DUST-RPCA) network, which explicitly takes advantage of the spatial and temporal continuity in the low-rank component. Our experimental results on the synthetic as well as real world datasets indicate that DUST-RPCA gives better accuracy when compared with the existing state of the art deep unfolded RPCA networks.

Index Terms—Deep unfolding, inverse problems, RPCA, foreground-background separation

I. INTRODUCTION

Robust Principal Component Analysis (RPCA) has become a popular choice in a range of applications including, but not limited to, computer vision [1], speech enhancement [2] and data analysis [3]. In fact, different variants of RPCA exist for various problem settings [4]. RPCA decomposes a data matrix D , the columns of which are constructed from vectorized data frames, into a sparse matrix S , which models the foreground and a low-rank matrix L , which models the background. RPCA is usually formulated on the Principal Component Pursuit (PCP) model. However, the PCP model's performance decreases in the presence of complex occurrences such as background clutter, occlusion and jitter. In such scenarios, a number of methods [5], [6] and [7] have been proposed which exploit the robust correlation between the data frames (columns of the input matrix D). As an example, the authors in [8], observe that these phenomena usually span multiple low-dimensional manifolds in the low-rank representation. Therefore, by incorporating spatial and temporal constraints in the PCP framework, one can control the coherency on these low-dimensional manifolds, resulting in better background-foreground separation accuracy. However, being an optimization method, it can take many iterations to converge with each iteration involving computationally expensive singular value decomposition (SVD) operation on high-dimensional data.

Deep Neural Networks (DNNs), such as ResNet [9], on the other hand, have shown to be much more computationally efficient than optimization based subspace methods while also being robust to occlusion, clutter and fast motion. However,

DNNs often require large training datasets and extreme parameterization [10]. This together with the abstractness of DNNs essentially make them black boxes and non-interpretable.

The recently proposed deep unfolded neural networks, designed by mapping the iterative steps of an optimization algorithm into DNN layers with learnable parameters, aim at solving the interpretability and large training dataset issues of DNNs while avoiding the high computational cost that is associated with traditional iterative optimization. Building upon this approach, the CORONA network [11] solved the low-rank and sparse ($L + S$) decomposition problem using convolution kernels in its unfolding process. As a result, it attained better performance as compared to its iterative equivalent and to other prominent DNNs, such as ResNet. The CORONA network's model, however, does not explicitly take advantage of the high correlation between successive frames of both the background and the foreground. The refRPCA [12] was proposed to solve this problem by incorporating side-information schemes in the shape of (i) a projection matrix P to potentially represent the high correlation between successive frames of the sparse component S , and (ii) a reweighting matrix Q to reweight the thresholds of the proximal operator. Both P and Q matrices are trainable parameters that are learnt by the network, making refRPCA more adaptive such that it outperforms the CORONA network. However, once refRPCA is trained, the reweighting matrix Q remains invariant to the data matrix D , which may vary significantly from one sample to another. Moreover, refRPCA model doesn't exploit the high correlation between the low-rank background frames.

In this paper, we propose a novel deep unfolded spatiotemporal RPCA (DUST-RPCA) network which explicitly takes advantage of the high correlation between successive frames of the background. This is in contrast to the correlation between foreground frames which may vary significantly for various applications. Consequently, we use the masks of the sparse component, given by the computationally efficient GoDec algorithm [13], to reweight the thresholds of the proximal operator [14]. To exploit the high background correlation, we add constraints that ensure continuity in both spatial and temporal dimensions of the background, which as per the model of RPCA is low-rank. The spatial constraint encodes the similarities between the rows of the input matrix D , whereas the temporal constraint encodes the similarities between the columns. Our experiments show that DUST-RPCA achieves better accuracy than both CORONA and refRPCA on synthetic as well as real-world datasets.

II. PROBLEM FORMULATION

A. The Proposed Objective Function

Consider a sequence of data frames, each denoted by d_i for time instances $i = 1, 2, \dots, q$ and each of size $m \times n$, having their low-rank and sparse components denoted as l_i and s_i , respectively. The data sequence d_i is arranged into a matrix D of size $p \times q$, where $p = mn$. Similarly, for sequences l_i and s_i , we obtain their corresponding matrices L and S , both of sizes $p \times q$. The RPCA, in its PCP formulation, separates the data matrix D into its low-rank component L and the sparse component S , as [15]

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

where $\|\cdot\|_1$ is the ℓ_1 norm, $\|\cdot\|_*$ is the nuclear norm and λ is a tuning regularizer. The PCP formulation, however, shows degraded performance amidst background clutter, occlusion, jitter etc. However, these challenges can be addressed by incorporating Laplacian based graph regularization in the PCP framework [8]. Consequently, these structural constraints are used to incorporate the spatial and temporal similitudes by reformulating (1) as

$$\min_{L, S} \|L\|_* + \lambda \|W \circ S\|_1 + \gamma_1 \text{Tr}(L^T A_s L) + \gamma_2 \text{Tr}(L A_t L^T) \text{ s.t. } D = L + S, \quad (2)$$

where $\text{Tr}(\cdot)$ denotes the trace of a matrix, \circ denotes the Hadamard product, A_s and A_t are the spatial and temporal Laplacian matrices respectively, and γ_1 and γ_2 are the regularization parameters which balance the spatial and temporal prior information, respectively. In addition, we also reweight the elements of S , as done in [14], with a prior matrix $W = \Phi(\rho, \hat{W})$, where $\Phi(\cdot)$ is the sigmoid function with a gain of ρ and \hat{W} is the background mask obtained by inverting the foreground mask from the GoDec algorithm. GoDec is significantly more efficient than ADMM [16] as it does not use time-consuming SVD operation for estimating the low-rank component. On average, the GoDec takes about 7% of the time taken by the ADMM algorithm in giving a similar mask.

The Laplacian matrices are computed as $A_g = T_g - W_g$, where $g \in \{s, t\}$, and T_g and W_g denote the corresponding adjacency and degree matrices respectively. The adjacency matrix W_g is computed as

$$W_g(i, j) = \exp\left(\frac{-\|x_i - x_j\|^2}{\tau_g \cdot \kappa_g^2}\right) \quad (3)$$

In case of W_s , x_i and x_j denote the i^{th} and j^{th} rows whereas in case of W_t , x_i and x_j denote the i^{th} and j^{th} columns of the input matrix D respectively. On the other hand, T_g is a diagonal matrix, with its diagonal entries composed of the sum of each row of W_g . The parameter κ_s is the average distance between the rows whereas κ_t is the average distance between the columns of the input matrix D respectively. The parameter τ_g is a constant which is further used to tune the smoothness between the values of W_g . To handle $\|\cdot\|_*$ and $\|\cdot\|_1$ in (2)

separately, we introduce two auxiliary matrices U_1 and U_2 to reformulate (2) as

$$\min_{L, S} \|L\|_* + \lambda \|W \circ S\|_1 + \gamma_1 \text{Tr}(U_1^T A_s U_2) + \gamma_2 \text{Tr}(U_2 A_t U_2^T) \text{ s.t. } D = L + S, L = U_1, L = U_2 \quad (4)$$

B. Subproblems for Spatiotemporal RPCA

The augmented Lagrangian function associated with (4) can be written as

$$\begin{aligned} \mathcal{L}(L, S, U_1, U_2, Y_1, Y_2, Y_3) = & \|L\|_* + \lambda \|W \circ S\|_1 + \\ & \gamma_1 \text{Tr}(U_1^T A_s U_1) + \gamma_2 \text{Tr}(U_2 A_t U_2^T) + \text{Tr}(Y_1^T \cdot (D - L - S)) \\ & + \text{Tr}(Y_2^T \cdot (L - U_1)) + \text{Tr}(Y_3^T \cdot (L - U_2)) + \\ & \frac{\mu}{2} (\|D - L - S\|_F^2 + \|L - U_1\|_F^2 + \|L - U_2\|_F^2) \end{aligned} \quad (5)$$

where $\|\cdot\|_F$ denotes the Frobenius norm, Y_1 , Y_2 and Y_3 are the Lagrange multipliers of size $p \times q$, and μ is a regularization parameter that controls the penalty for violation of linear constraints. We solve (5) using alternating minimization [17], which involves separately minimizing the augmented Lagrangian function. In k^{th} iteration, (5) is minimized with respect to each optimization variable while the remaining variables are kept constant. The sub-problems for updating L , S , U_1 and U_2 can be written as

$$\min_{U_1} \frac{\gamma_1}{\mu} \text{Tr}(U_1^T A_s U_1) + \frac{1}{2} \|L - U_1 + \frac{Y_2}{\mu}\|_F^2, \quad (6)$$

$$\min_{U_2} \frac{\gamma_2}{\mu} \text{Tr}(U_2 A_t U_2^T) + \frac{1}{2} \|L - U_2 + \frac{Y_3}{\mu}\|_F^2, \quad (7)$$

$$\min_L \frac{1}{\mu} \|L\|_* + \frac{1}{2} \|D - L - S + U_1 + U_2 + \frac{(Y_1 + Y_2 + Y_3)}{\mu}\|_F^2, \quad (8)$$

$$\min_S \frac{\lambda}{\mu} \|W \circ S\|_1 + \frac{1}{2} \|D - L - S + \frac{Y_1}{\mu}\|_F^2. \quad (9)$$

The solutions to sub-problems (6) and (7), for updating U_1 and U_2 at iteration k , can be found in closed form as

$$U_1^{k+1} = (2\gamma_1 A_s + \mu^k I)^{-1} (\mu^k L^k + Y_2^k) \quad (10)$$

$$U_2^{k+1} = (\mu^k L^k + Y_3^k) (2\gamma_2 A_t + \mu^k I)^{-1} \quad (11)$$

III. UNFOLDING SPATIOTEMPORAL RPCA

Although there are different ways to unfold an iterative algorithm into a DNN such as described in [11], [18] and [19], we follow the method of [11] in unfolding (8) and (9). Hence, we introduce measurement matrices $\{H_i\}_{i=1}^4$ into (8) and (9) as follows

$$\min_L \frac{1}{\mu} \|L\|_* + \frac{1}{2} \|D - H_1 L - H_2 S + H_3 U_1 + H_4 U_2 + \frac{(Y_1 + Y_2 + Y_3)}{\mu}\|_F^2, \quad (12)$$

$$\min_S \frac{\lambda}{\mu} \|W \circ S\|_1 + \frac{1}{2} \|D - H_1 L - H_2 S + \frac{Y_1}{\mu}\|_F^2. \quad (13)$$

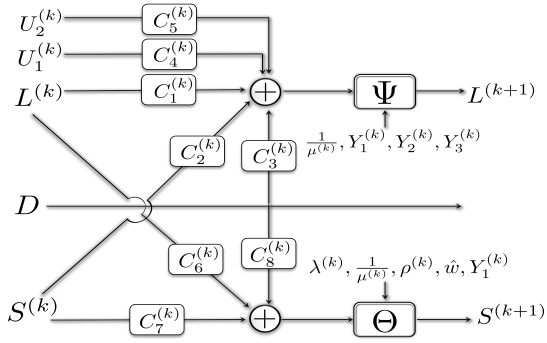


Fig. 1: DUST-RPCA architecture

While $\{H_i\}_{i=1}^4$ can ideally be set with respect to the application at hand, we take them to be identity matrices in our case since we have no prior knowledge of a more informed choice. Moreover, they are further abstracted to enable the architecture of DUST-RPCA by replacing them with learnable kernels in the deep unfolding steps. We solve (12) and (13) using ISTA [20], as

$$B^{k+1} = \text{prox} \left(B^k - \frac{1}{c} \nabla (B^k) \right), \quad (14)$$

where B can be either L or S , c denotes the Lipschitz constant, $\text{prox}(\cdot)$ denotes the proximal operator [21] and $\nabla(\cdot)$ denotes the gradient of the quadratic part with respect to B . The proximal operator for computing the low-rank component is the singular value thresholding operation [22], $\Psi_\alpha(X) = U \text{diag}(\text{Relu}(\sigma_i - \alpha)) V^H$, where X has an SVD given by $X = U \Sigma V^T$, $\text{diag}(y_i)$ represents a diagonal matrix with its i^{th} diagonal entry equal to y_i and σ_i represents the i^{th} singular value of X . The proximal operator for computing the sparse component is the soft-thresholding operation [20], $\Theta_\alpha(x) = \text{sign}(x) \cdot \max(|x| - \alpha, 0)$. The thresholding value α at iteration k for Θ_α is kept as $\frac{\lambda^k}{\mu^k \cdot c} \cdot \Phi(\rho^k, \hat{w})$, where \hat{w} is the element of the matrix \hat{W} . This design leads to a distinct proximal operator for every input at each iteration. Using (14), the iterative steps of L and S , after some algebraic manipulation, turn out to be

$$L^{k+1} = \Psi_{\frac{1}{\mu^k \cdot c}} \left\{ \left(I - \frac{1}{c} H_1^T H_1 \right) L^k + (-H_1^T H_2) S^k + (H_1^T) D + (H_1^T H_3) U_1^k + (H_1^T H_4) U_2^k + (H_1^T) \frac{Y_1^k + Y_2^k + Y_3^k}{\mu^k} \right\}, \quad (15)$$

$$S^{k+1} = \Theta_{\frac{\lambda^k}{\mu^k \cdot c} \cdot \Phi(\rho^k, \hat{w})} \left\{ \left(I - \frac{1}{c} H_2^T H_2 \right) S^k + (-H_2^T H_1) L^k + (H_2^T) D + (H_2^T) \frac{Y_1^k}{\mu^k} \right\}. \quad (16)$$

We unroll (15) and (16) into the multi-layer neural network of DUST-RPCA by replacing the functions of the measurement

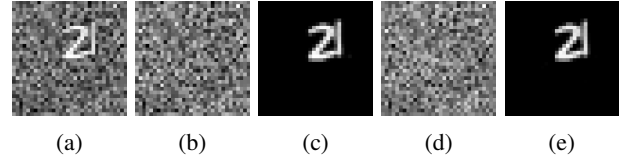


Fig. 2: Separation results of DUST-RPCA: (a) Input image (b) Background groundtruth (c) Foreground groundtruth (d) Background recovered (e) Foreground recovered

matrices with convolution kernels as follows

$$L^{k+1} = \Psi_{\frac{1}{\mu^k}} \left\{ C_1^k * L^k + C_2^k * S^k + C_3^k * D + \right. \quad (17)$$

$$\left. C_4^k * U_1^k + C_5^k * U_2^k + \frac{Y_1^k + Y_2^k + Y_3^k}{\mu^k} \right\}$$

$$S^{k+1} = \Theta_{\frac{\lambda^k}{\mu^k} \cdot \Phi(\rho^k, \hat{w})} \left\{ C_6^k * L^k + C_7^k * S^k + C_8^k * D + \frac{Y_1^k}{\mu^k} \right\}, \quad (18)$$

where $*$ denotes the convolution operator, $C_1^k, C_2^k, \dots, C_8^k$ are the convolution kernels learned in the k^{th} layer of DUST-RPCA network. The Lagrange multipliers of Y_1^k, Y_2^k and Y_3^k in (10), (11), (17) and (18) are also learned by the network at every layer together with the parameters $1/\mu^k, \lambda^k$ and ρ^k . Fig. 1 shows the k^{th} layer of the architecture of DUST-RPCA.

IV. EXPERIMENTAL EVALUATIONS

A. Experiment Settings

We evaluated the performance of our proposed algorithm on two different datasets: the moving MNIST dataset and the CDNet dataset. CDNet dataset consists of real world videos with camera jitter, dynamic backgrounds, and occlusion. The moving MNIST dataset [23] contains 10,000 moving digit sequences, each 20 frames long. Each of these frames is resized to 32×32 pixels and then stacked as columns to create the sparse matrix S of size 1024×20 . Furthermore, we sample from a standard Gaussian distribution to generate matrices $U \in \mathbb{R}^{1024 \times r}$ and $V \in \mathbb{R}^{20 \times r}$, with the rank r set to 5. These matrices are in turn used to generate the low-rank matrix $L = UV^T$. We superimpose the MNIST digits in S on the low-rank matrix L to create the data matrix D . Finally, with the data sequences normalized to the range $[0, 1]$, we split the dataset into 9000 and 1000 sequences for training and testing respectively. The loss function \mathcal{L} is kept as

$$\mathcal{L} = \frac{1}{2} \left\{ f(L_i, \bar{L}_i) + f(S_i, \bar{S}_i) \right\}, \quad (19)$$

$$f(X_i, \bar{X}_i) = \frac{1}{N} \sum_{i=1}^N \|X_i - \bar{X}_i\|_F^2,$$

where $f(X_i, \bar{X}_i)$ is the mean-square error (MSE) loss function, N is the total number of training sequences in the dataset, and $\{\bar{L}_i, \bar{S}_i\}_{i=1}^N$ are the low-rank and the sparse components predicted by the network against their groundtruths

$\{L_i, S_i\}_{i=1}^N$. Network configurations such as kernel size and weight initialization are kept identical to [11], for a fair comparison.

The second experiment is conducted on the CDNet dataset. We resize images from the dynamic background category to 32×32 pixels and construct a single sample by stacking 20 such frames as columns. We split the dataset into 450 training samples and 125 validation samples. Networks for both datasets were trained using the Adam optimizer with a learning rate of 2×10^{-3} (MNIST) and 2.2×10^{-3} (CDNet) respectively for 30 epochs. Since the CDNet ground truth labels do not give us information regarding the actual background behind the sparse foreground object, we modify our loss function \mathcal{L} as follows:

$$\mathcal{L} = \frac{1}{2} \left\{ f(L_i \circ (1 - W_i), \bar{L}_i \circ (1 - W_i)) + f(S_i, \bar{S}_i) \right\} \quad (20)$$

$$f(X_i, \bar{X}_i) = \frac{1}{N} \sum_{i=1}^N \|X_i - \bar{X}_i\|_F^2,$$

where W is a masking matrix corresponding to the sparse component in each frame. The Hadamard product $L_i \circ (1 - W_i)$ yields the low rank background only. This approach is similar to the one adopted in [24].

B. Results and Analysis

Fig. 2 shows the visual separation results of DUST-RPCA on a sample frame from the moving MNIST dataset. DUST-RPCA accurately recovers both the foreground and the background. The results presented in Fig. 3 shows that the DUST-RPCA achieves significantly lower MSE for both synthetic (moving MNIST), as well as real world (CDNet) datasets, than that achieved by other networks. It also converges earlier than other networks. The refRPCA, building upon the framework of

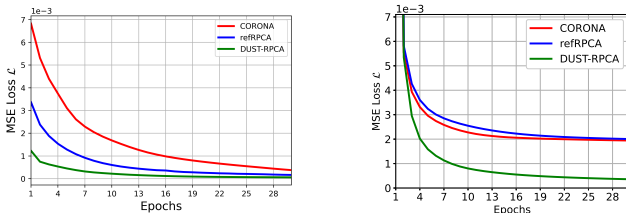


Fig. 3: (Left) MSE loss for moving MNIST dataset, (Right) MSE loss for CDNet dataset

the CORONA network, formulates its problem in terms of S_P - a matrix representing reference frames, Q - a weighting matrix and η_1, η_2 and η_3 - tuning regularizers. S_P is constructed as $S_P = [s_1, Ps_1, \dots, Ps_{q-1}]$, where s_q represents the q^{th} column of the corresponding sparse matrix S and P represents the projection matrix. However, it makes the simplifying assumption that the projection matrix P does not significantly vary from one frame to the other. As such, it cannot fully incorporate the more complex correlation between frames. More importantly, both P and Q , which are learned by the network during training, remain invariant at inference time to

TABLE I: Accuracy and efficiency comparison of DUST-RPCA with other deep unfolded RPCA networks.

Method	MNIST Dataset		CDNet Dataset	
	MSE ($\times 10^{-4}$)	Time (ms)	MSE ($\times 10^{-4}$)	Time (ms)
CORONA [11]	3.77	1.19	19.43	1.40
refRPCA [12]	1.66	2.47	20.04	0.73
DUST-RPCA	0.61	5.38	3.60	4.31

the data matrix D , which may vary significantly from one input sample to another. The proximal operator of DUST-RPCA, on the other hand, adapts for every input at each layer of the network, making it more robust to the varying input data sequences. Its performance would hold even in applications such as cloud removal, where there is little to no correlation between the successive sparse foregrounds in S . Although the correlation between successive sparse foregrounds may vary from one application to the other, the correlation between the background frames remains high. In fact, RPCA makes use of this phenomenon in its $D = L + S$ model to solve the background-foreground separation problem. However, amidst noise and other phenomena such as occlusion, clutter and jitter, the high correlation between background frames starts to decrease. To solve this problem, unlike other deep unfolded networks, DUST-RPCA jointly learns the continuity across both spatial and temporal dimensions of the background thus achieving better MSE performance than both CORONA and refRPCA.

Table I shows the average MSE achieved on the testing dataset by CORONA, refRPCA and DUST-RPCA along with the average inference time, in milli-seconds, per frame of the MNIST sequence. While DUST-RPCA achieves the lowest MSE, it also takes more time at inference due to the one-time computation of A_s and A_t for every input and the addition of more convolutional layers. However, it is the SVD computation that makes up the computational bottleneck for DUST-RPCA much like CORONA and refRPCA. Moreover, the inference time is still much less than that of traditional RPCA optimization methods which use relatively more layers/iterations than the model-aided RPCA networks [11].

V. CONCLUSION

Deep unfolded networks have demonstrated improved convergence and robustness over their iterative counterparts. Unlike other unfolded RPCA networks, the proposed DUST-RPCA scheme incorporates spatial and temporal consistency constraints by way of regularized graph Laplacian matrices achieving better accuracy. It also reweights the thresholds for the proximal operator for every input at each layer of the network making the proposed DUST-RPCA more adaptive. Moreover, the proposed DUST-RPCA can also be used in applications, for instance medical imaging, where there may be little to no correlation between foregrounds in successive frames.

REFERENCES

- [1] Thierry Bouwmans, Sajid Javed, Hongyang Zhang, Zhouchen Lin, and Ricardo Otazo. On the Applications of Robust PCA in Image and Video Processing. *Proceedings of the IEEE*, 106(8):1427–1457, 2018.
- [2] Pengfei Sun and Jun Qin. Low-rank and sparsity analysis applied to speech enhancement via online estimated dictionary. *IEEE Signal Processing Letters*, 23(12):1862–1866, 2016.
- [3] John Wright and Yi Ma. *High-Dimensional Data Analysis with Low-Dimensional Models: Principles, Computation, and Applications*. Cambridge University Press, 2022.
- [4] Namrata Vaswani, Yuejie Chi, and Thierry Bouwmans. Rethinking PCA for Modern Data Sets: Theory, Algorithms, and Applications [Scanning the Issue]. *Proceedings of the IEEE*, 106(8):1274–1276, 2018.
- [5] Tianzhu Zhang, Bernard Ghanem, Si Liu, Changsheng Xu, and Narendra Ahuja. Robust visual tracking via exclusive context modeling. *IEEE Transactions on Cybernetics*, 46(1):51–63, 2016.
- [6] Tao Zhou, Fanghui Liu, Harish Bhaskar, and Jie Yang. Robust visual tracking via online discriminative and low-rank dictionary learning. *IEEE Trans. Cybern.*, 48(9):2643–2655, 2018.
- [7] Yehui Yang, Wenrui Hu, Yuan Xie, Wensheng Zhang, and Tianzhu Zhang. Temporal restricted visual tracking via reverse-low-rank sparse learning. *IEEE Transactions on Cybernetics*, 47(2):485–498, 2017.
- [8] Sajid Javed, Arif Mahmood, Thierry Bouwmans, and Soon Ki Jung. Background foreground modeling based on spatiotemporal sparse subspace clustering. *IEEE Transactions on Image Processing*, 26(12):5840–5854, 2017.
- [9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [10] Nir Shlezinger, Yonina C Eldar, and Stephen P Boyd. Model-based deep learning: On the intersection of deep learning and optimization. *arXiv preprint arXiv:2205.02640*, 2022.
- [11] 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.
- [12] Huynh Van Luong, B. Joukovsky, Yonina C. Eldar, and N. Deligiannis. A Deep-Unfolded Reference-Based RPCA Network For Video Foreground-Background Separation. *2020 28th European Signal Processing Conference (EUSIPCO)*, pages 1432–1436, 2021.
- [13] 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.
- [14] Shoaib Imran, Muhammad Tahir, Zubair Khalid, and Momin Uppal. A Deep Unfolded Prior-Aided RPCA Network for Cloud Removal. *IEEE Signal Processing Letters*, 29:2048–2052, 2022.
- [15] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis. *J. ACM*, 58(3), Jun 2011.
- [16] Stephen Boyd. *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*. 01 2010.
- [17] Zhouchen Lin, Risheng Liu, and Zhixun Su. Linearized alternating direction method with adaptive penalty for low-rank representation. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011.
- [18] 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*, 34:16977–16989, 2021.
- [19] Vishal Monga, Yuelong Li, and Yonina C Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44, 2021.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] Nitish Srivastava, Elman Mansimov, and Ruslan Salakhutdinov. Unsupervised Learning of Video Representations Using LSTMs. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, ICML’15*, page 843–852. JMLR.org, 2015.
- [24] Boris Joukovsky, Nikos Deligiannis, and Yonina C Eldar. Interpretable neural networks for video separation: Deep unfolding RPCA with foreground masking. 2022.