

Unsupervised Learning for Salient Object Detection via Minimization of Bilinear Factor Matrix Norm

Min Li^{ID}, Yao Zhang, Mingqing Xiao^{ID}, Senior Member, IEEE, Weiqiang Zhang, and Xiaoli Sun^{ID}

Abstract—Saliency detection is an important but challenging task in the study of computer vision. In this article, we develop a new unsupervised learning approach for the saliency detection by an intrinsic regularization model, in which the Schatten-2/3 norm is integrated with the nonconvex sparse $l_{2/3}$ norm. The $l_{2/3}$ -norm is shown to be capable of detecting consistent values among sparse foreground by using image geometrical structure and feature similarity, while the Schatten-2/3 norm can capture the lower rank of background by matrix factorization. To improve effective performance of separation for Schatten-2/3-norm and $l_{2/3}$ -norm, a Laplacian regularization is adopted to the foreground for the smoothness. The proposed model essentially converts the required nonconvex optimization problem into the convex one, conducted by splitting the objective function based on singular value decomposition on one much smaller factor matrix and then optimized by using the alternating direction method of the multiplier. The convergence of the proposed algorithm is discussed in detail. Extensive experiments on three benchmark datasets demonstrate that our unsupervised learning approach is very competitive and appears to be more consistent across various salient objects than the current existing approaches.

Index Terms—Alternating direction method of multiplier (ADMM), $l_{2/3}$ -norm, low-rank matrix, matrix decomposition, salient object detection, Schatten- q -norm.

I. INTRODUCTION

SALIENT object detection is one of the fundamental tasks in the study of computer vision, aiming to localize and segment the foreground objects from the background. It appears in a wide range of applications, such as in the study of psychology, neurobiology, brain bionics, medical image, and so on.

Manuscript received June 20, 2020; revised January 25, 2021 and June 14, 2021; accepted August 11, 2021. The work of Min Li, Yao Zhang, Weiqiang Zhang, and Xiaoli Sun was supported in part by the National Nature Science Foundation of China under Grant 62072312, Grant 61972264, Grant 61872429, and Grant 61772343; in part by the National Nature Science Foundation of Guangdong province under Grant 2019A15100894; in part by the Shenzhen Basis Research Project under Grant JCYJ20180305125521534; and in part by the Interdisciplinary Innovation Team, Shenzhen University, under Grant SZUGS2021SEMINAR06. The work of Mingqing Xiao was supported by the National Science Foundation-Division of Mathematical Sciences (NSF-DMS) of the U.S. under Grant 1419028. (*Corresponding authors: Yao Zhang; Mingqing Xiao*)

Min Li, Weiqiang Zhang, and Xiaoli Sun are with the Shenzhen Key Laboratory of Advanced Machine Learning and Applications, College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, China (e-mail: limin800@szu.edu.cn; wzhang@szu.edu.cn; xlsun@szu.edu.cn).

Yao Zhang is with the Key Laboratory of Solar Activity, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China (e-mail: jaafar_zhang@163.com).

Mingqing Xiao is with the Department of Mathematics, Southern Illinois University, Carbondale, IL 62901 USA (e-mail: mxiao@math.siu.edu).

This article has supplementary material provided by the authors and color versions of one or more figures available at <https://doi.org/10.1109/TNNLS.2021.3105276>.

Digital Object Identifier 10.1109/TNNLS.2021.3105276

During past decades, many different approaches for salient object detection have been proposed in order to improve detection effectiveness. Among them, the deep learning-based method [1]–[5], the bottom-up method [6]–[11], the top-down method [1], [12], and the method by combining bottom-up cues with top-down priors [13]–[20] are the most commonly seen in the literature.

In this article, we present a new approach by formulating the low-rank and sparse decomposition for salient object detection. More specifically, let image X be represented as a combination of the background Z and the foreground E . The background Z often tends to be highly redundant and lies in a low-dimensional subspace. Hence, the background Z usually is characterized by a low-rank feature matrix. The foreground E is a distinctive salient object that is often sparse and deviates from the low-rank subspace. Therefore, the foreground E can be represented by a sparse sensory matrix. Mathematically, this problem generally can be formulated as (see [14])

$$\min_{Z, E} \text{rank}(Z) + \lambda \|E\|_0 \quad \text{s.t. } X = Z + E \quad (1)$$

where $\|\cdot\|_0$ denotes the l_0 -norm by convention, which represents the number of nonzero matrix entries, and λ is positive regularization parameter. However, model (1) by nature is a nonconvex NP-hard problem. Consequently, this results in the emergence of many relaxation models in which both nonconvex terms are replaced by their convex envelopes (i.e., the nuclear norm $\|\cdot\|_*$ and the l_1 -norm, respectively) so that the minimization problem becomes computationally feasible.

The drawback for adopting the convex envelope as relaxation is the existence of a gap between the rank function/sparse metric and its convex envelope, which may result in the solution being undesirable [20]–[23]. As shown in Fig. 1, for example, some common issues by convex approaches can be seen, such as the backgrounds appear to be cluttered and the foreground targets show to be scattered and incomplete (as shown in ULR¹ and SLR² of Fig. 1), or in some other cases, the salient target contains the entire image because the loss of the sparse property (as shown in SMD³ and WLRR⁴ of Fig. 1). While our proposed nonconvex approach (as shown ours in Fig. 1) in this article provides a desirable solution in these cases. Therefore, compared with the nonconvex approach,

¹Shen and Wu [15] proposed a unified low-rank matrix recovery, which is referred to as ULR for short.

²Zou *et al.* [17] presented the segmentation guided low-rank method for the matrix recovery, which is referred to as SLR for short.

³Peng *et al.* [18] gave a structured matrix decomposition approach, which is referred to as SMD for short.

⁴Tang *et al.* [19] proposed the weighted low-rank matrix recovery, which is referred to as WLRR for short.

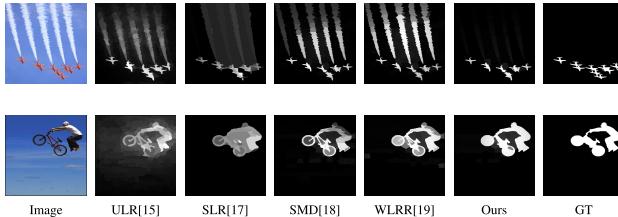


Fig. 1. Examples of salient object detection from typical existing models, compared with ours in this article. GT stands for the ground truth. The approaches under the convex framework appear to be problematic.

the convex relaxation setting for the low-rank approximation appears its limitation in capturing the salient object from its background effectively and often generates undesirable outcomes.

Besides, many numerical algorithms for the low-rank-based models are associated with high-computational complexity since they are based on singular value decomposition (SVD) on the full matrix [24]. This leads to a low efficiency for handling very large-scale datasets, which often is undesirable in real-time applications. To address the issue in terms of effectiveness and efficiency for salient object detection, in this article, we propose the unsupervised learning approach for the saliency detection, in which the nonconvex Schatten- $2/3$ norm as the regularization is used to retain the background detail, and the nonconvex $l_{2/3}$ -norm is exploited to capture the sparsity of foreground object. The reason for choosing these two particular norms is elaborated on next.

Comparing with the widely used nuclear norm, the Schatten- $2/3$ norm, which is equivalent to the $l_{2/3}$ -norm on singular values, is better in capturing the rank function structure since the $q = 2/3$ value scales each singular value accordingly, maintaining the convexity in its factors. This adopted regularization, as shown in this article, can make the background adequately lie in its corresponding subspace, allowing the low-rank approximation of the background and providing a cleaner background image than other existing low-rank approaches.

Regarding the l_1 -norm, the nonconvex $l_{2/3}$ -norm is tractable in the sparse approximation than the unfeasible l_0 -norm. And this setting appears to be effective in sharing the desired consistency within the same image patches due to taking into account both image geometrical structure and feature similarity among image patches. Moreover, from the experimental perspective, the similar appearance between the salient objects and the background could affect the accuracy of detection. To identify the difference between the foreground and the background, we adopt the Laplacian term, as shown in [18], that can promote the object completeness as much as possible (as shown in Fig. 1).

In addition, the unified surrogate for Schatten- $2/3$ quasi-norm can be represented as Frobenius/nuclear hybrid norm, and the nonconvex $l_{2/3}$ norm permits an analytic solution for the objective function by searching the roots of a quadratic polynomial as other problems [25], [26]. Therefore, the factorization formulation for Schatten- $2/3$ quasi-norm requires far less memory, only requiring to compute the SVD on a much

small factored matrix, as contrary to the full matrix in the other low-rank-based methods. Most recently, the bilinear factor matrix norm in literature is adopted for image inpainting, alignment, data completion, and recovery [24], [27], [28]. This article is the first study to apply Schatten- $2/3$ and $l_{2/3}$ norms to the salient object detection problem, which requires a different but subtle study, in particular, for the optimization process.

The remainder of this article is organized as follows: Section II provides some preliminary introduction and some recent related works. The proposed framework is illustrated in Section III. Section IV develops the corresponding minimization algorithm. The detailed analysis of both convergence and complexity of our proposed algorithm is given in Section V. Extensive experiment results are provided in Section VI. This article ends with concluding remarks in Section VII.

II. RELATED APPROACHES

Salient object detection has been extensively studied during recent years. Due to the inherent nature, low-rank representation (referred to as LRR for short) [29] is the most typical method, which uses a low-rank coefficient matrix multiplied by a dictionary to represent the background and utilize $l_{2,1}$ norm to characterize sparse components (i.e., the salient objects). The formulation is given by

$$\min_{Z, E} \|Z\|_* + \lambda \|E\|_{21} \quad \text{s.t. } X = XZ + E \quad (2)$$

where X is the input image, XZ represents the background detail, which can be reconstructed independently, Z is the set of reconstruction coefficients, E is the salient objects, $\|\cdot\|_*$ is the nuclear norm, and $\|\cdot\|_{21}$ is used to characterize the sparsity of the columns of E .

The *single* LRR approach, however, is not ideal in modeling visual *multiple* features, as shown in the literature, because of its simple structure. Thus, by integrating multifeature with top-down priors, LRR is generalized as a multitask sparsity pursuit method (referred to as MSP for short) [16]. The approach is formulated by combining multiple features as

$$\begin{aligned} & \min_{Z_1, \dots, Z_J, E_1, \dots, E_J} \sum_{i=1}^J \|Z_i\|_* + \lambda \|E\|_{21} \\ & \text{s.t. } X_i = X_i Z_i + E_i, \quad i = 1, \dots, J. \end{aligned} \quad (3)$$

However, the norm $\|\cdot\|_{21}$ of the sparse matrix E does not match nuclear norm well due to the absence of spatial relations and the feature affinities, leading to the generation of inaccurate saliency object. Hence, in order to promote the ULR to absorb the low-level features and to generate the high-level guidance, the following setting is proposed [15]:

$$\min_{Z, E} \|Z\|_* + \lambda \|E\|_1 \quad \text{s.t. } X = Z + E. \quad (4)$$

Experimentally, the ULR approach results in the nonuniformly highlighted salient object due to the lack of spatial relations and feature similarities of patches. To address this issue, SLR developed an unsupervised model by leveraging the low-rank matrix recovery and the segmentation priors [17]. SMD established a matrix decomposition approach based on a tree-structured sparsity and a Laplacian regularization [18].

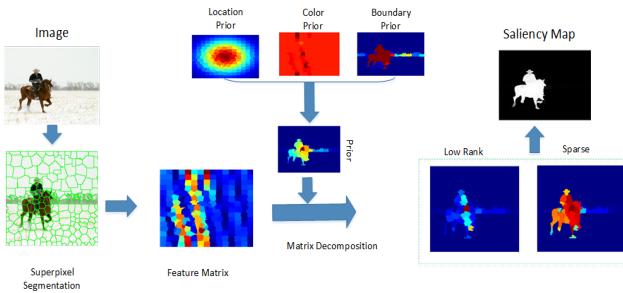


Fig. 2. Flowchart of the proposed approach, where background is characterized by the Schatten-2/3 quasi-norm and foreground is approximated by $l_{2/3}$ -norm.

As an extension of the SMD model, WLRR constructed a weighted low-rank regularization based on the high-level background prior map [19].

Although many improvements have been achieved in terms of the existing models, the background detail is mainly characterized by the nuclear norm. It is a common consensus that the minimization based on the nuclear norm usually over shrinks the background features when it is used to approximate the rank function. On the other hand, l_1 -norm for foreground over penalizes large entries of vector and leads to a biased solution, whereas the structured-tree sparse constraint on fine feature in SMD destroys the spatial relationship of the salient object. Therefore, to address these issues, the nonconvex Schatten-2/3 quasi-norm and the $l_{2/3}$ -norm, respectively, are adopted for background and foreground in this article. The advantages of this proposed approach will be illustrated in detail later. A flowchart of the proposed model can be shown in Fig. 2.

III. UNSUPERVISED DECOMPOSITION MODEL VIA MINIMIZATION OF BILINEAR FACTOR MATRIX NORM

A. Formulation of Unsupervised Decomposition Model

Given an input image, we conduct the oversegmentation and extract the basic features. Consequently, a feature matrix X can be decomposed into redundancy part L and spare one S and can be characterized with different regularizations respectively. Specifically, the optimization problem for the salient object detection can be formulated as the following general framework:

$$\begin{aligned} & \min_{L, S} f(L) + \lambda g(S) + \gamma h(L, S) \\ & \text{s.t. } X = L + S \end{aligned} \quad (5)$$

where $f(L)$ is the spectral regularization for background L which ensures low-rank, and $g(S)$ is a sparse constraint. $h(L, S)$ is an interactive regularization term which is introduced to improve performance when clutter appears in the background. And λ, γ are positive parameters that tradeoff among these three terms.

B. Schatten- q Quasi-Norm for Background Detail

It is known that background, in general, is located in a low-dimensional subspace, lying in the given structural datasets; thus, how to find the suitable rank to capture the background is the crux of this problem. Technically, the rank function can

be relaxed by its convex envelope [13]–[19], [29], i.e., nuclear norm, which is the sum of singular values, and also is the Schatten-1 norm (or the trace norm) [24]. Unfortunately, there is always a gap between the nuclear norm and the (nonconvex) rank function,⁵ which causes to over-penalize rank components of a low-rank matrix and results in a deviation from the original solution. Therefore, many new strategies based on nonconvex rank relaxation have been explored to fill this gap (see, [21], [24], [27], [28], [30], [31]).

Among them, the nonconvex Schatten- q prenorm can reduce the gap between rank function ($q = 0$) and nuclear norm ($q = 1$) by setting $0 < q < 1$, which has superiority in providing an increasingly tight approximation to low-rank matrix requiring only the weaker restricted isometry property. Since the background is dominated by a few factors, which always shows low-rank property. Thereupon, we propose this flexible nonconvex regularization for the background, which can effectively describe the lower rank structure of background. To be specific, the Schatten- q norm of a matrix $L \in \mathbb{R}^{m \times n}$ is defined as

$$\|L\|_{Sq} = \left(\sum_{j=1}^{\min\{m,n\}} \sigma_j^q(L) \right)^{1/q} \quad (6)$$

where $\sigma_j(L)$ is the j th singular value of L .

Obviously, the nonconvex Schatten- q norm ($0 < q < 1$) is in essence the l_q -norm on singular values. And the common forms of Schatten- q norm in existing models are summarized in Table I for $q \in [0, 2]$. To reduce computation complexity and better handle large-scale tasks of computer vision, matrix factorization has been proposed to solve the models based on Schatten- q norm [24], [27], [28], [31]. Especially, the unified surrogates for the Schatten- q norm can be characterized as follows.

Lemma 1 (Bi-Schatten- q Norm Surrogate [28]): Given matrices $U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}$, and $L \in \mathbb{R}^{m \times n}$ with $\text{rank}(L) = r \leq d$, the following equality holds:

$$\frac{1}{q} \|L\|_{Sq}^q = \min_{U, V: L=UV^T} \frac{1}{q_1} \|U\|_{S_{q_1}}^{q_1} + \frac{1}{q_2} \|V\|_{S_{q_2}}^{q_2} \quad (7)$$

where q, q_1 , and $q_2 > 0$, which satisfy $(1/q) = (1/q_1) + (1/q_2)$.

To the best of our knowledge, Schatten-1/2 norm and Schatten-2/3 norm [24], [27], [28], [31] have been recently used widely in matrix completion and robust principal component analysis, respectively. Here, we focus on the specific values of $q = 2/3$ due to its decomposable formulation and better experimental performance for the background. In fact, the Schatten-2/3 norm is, in essence, equivalent to Frobenius/nuclear hybrid norm. Hence, it follows from Lemma 1 that the Frobenius/nuclear hybrid norm can be described as the following theorem.

⁵The rank of a matrix is to count the number of nonzero singular values. Low-rank is equivalent to seeking the sparsity of the singular values of a matrix. Thus, rank minimization is an important regularization for a low-rank solution. However, rank minimization is hard. The common numerical strategy is to find the nuclear norm as a substitute, which in fact sums the amplitude of nonvanishing singular values. Thereupon, a gap between the nuclear norm and the rank function emerges naturally.

TABLE I
COMMON FORMS FOR SCHATTEN- q QUASI-NORM
AND l_p -NORM WHEN $p, q \in [0, 2]$

p, q	Low-rankness $\ L\ _{S_q}$	Sparsity $\ S\ _{l_p}$
0	$\ L\ _{S_{l_0}} = \#\{ \sigma_i \neq 0\}$	$\ S\ _0 = \#\{s_{ij} \neq 0\}$
(0, 1)	$\ L\ _{S_q} = (\sum \sigma_k^q)^{\frac{1}{q}}$	$\ S\ _{l_p} = (\sum \sum s_{ij} ^p)^{\frac{1}{p}}$
1	$\ L\ _* = \sum \sigma_k$	$\ S\ _{l_1} = \sum \sum s_{ij} $
2	$\ L\ _F = \sqrt{\sum \sigma_k^2}$	$\ S\ _F = \sqrt{\sum \sum s_{ij} ^2}$

Fig. 3. Comparison results between the nuclear norm commonly used for background in the literature and the Schatten-2/3 norm used in this article.

Lemma 2 (Frobenius/Nuclear [24], [28]): Given $L \in \mathbb{R}^{m \times n}$ with rank $(L) = r \leq d$, L can be factorized into two much smaller matrices $U \in \mathbb{R}^{m \times d}$ and $V \in \mathbb{R}^{n \times d}$ such that $L = UV^T$. The Frobenius/nuclear norm of L can be denoted as

$$\begin{aligned} \|L\|_{F/N} &= \|L\|_{S_{2/3}} := \min_{L=UV^T} \|U\|_* \|V\|_F \\ &= \min_{U, V: L=UV^T} \left[\frac{2}{3} \|U\|_* + \frac{1}{3} \|V\|_F^2 \right]^{\frac{3}{2}} \end{aligned} \quad (8)$$

where $\|\cdot\|_*$ is nuclear norm and $\|\cdot\|_F$ is the Frobenius norm.

Based on Lemmas 1 and 2, the nonconvex and nonsmooth regularization for background is transformed into a convex and smooth one, which is not only tractable but also feasible for the required optimization. The factorization formulation (8) only requires to compute SVD on a much smaller matrix U as compared with the full matrix L . Fig. 3 gives the intuitive results from ablation experiments for the background, respectively, in which the second column of images are obtained by using nuclear norm only and the third column of images are given via Schatten-2/3 regularization. It is clear that the decomposed background results of $\|\cdot\|_{S_{2/3}}^{2/3}$ are significantly better than that of nuclear norm visually.⁶

C. Nonconvex $l_{2/3}$ Norm for Foreground and Laplacian Constraint

It is expected that a valid regularization for the foreground can extract the complete information from the salient regions.

⁶To quantitatively highlight that Schatten-2/3 norm can approximate low-rank more accurately, we calculate the nuclear norm and Schatten-2/3 norm of the background for the second row of images in Fig. 3, respectively. Suppose that the rank of background is 40, which can be treated as the representative of low rank. By setting the saliency object to zero, we extract the background matrix L . The nuclear norm and Schatten-2/3 norm for the first 40 values of SVD of L , respectively, can be denoted as $\|L\|_* = 3.769 \times 10^4$ and $\|L\|_{\frac{2}{3}} = 3.2 \times 10^3$. Thus, we have $\|L\|_* - 40 \approx 3.77 \times 10^4$ and $\|L\|_{\frac{2}{3}} - 40 \approx 3.2 \times 10^3$. Obviously, $3.2 \times 10^3 < 3.77 \times 10^4$, which coincides with our original motivation of choosing Schatten-2/3 norm for background.

Due to being sparse, the foreground usually exhibits a heavy-tailed distribution, we impose the hyper-Laplacian priors to the salient part S , i.e., the nonconvex l_p norm, which can maintain the sparse feature of the corresponding matrix more effectively. Technically, this regularization can be uniformly formulated in the following form:

$$g(S) = \|S\|_{l_p}^p. \quad (9)$$

In the standard framework, for $p \in [0, 2]$, the regularized term $\|\cdot\|_{l_p}$ is shown in Table I. Compared with l_1 -norm, the nonconvex surrogates listed in Table I give not only a better but also tractable approximation than the l_0 norm. Thus, the nonconvex l_p -norm takes the full advantage of the spatial contiguity and feature similarity among all image patches so that they have more accurate and consistent representation. On the other hand, an analytic solution can be found for the specific values $p = 2/3$ by seeking the roots of a quadratic polynomial [25], [26], and $l_{2/3}$ -norm can obtain competitive performance for decomposition of foreground as verified in experiments VI. In addition, it has been known that the corresponding algorithm is more efficient than many existing algorithms, as shown in [25]. Thus, in this article, we focus on the nonconvex $l_{2/3}$ norm detecting the foreground which is represented by identifying spatially localized sparse residuals. Fig. 4 gives the comparative examples for sparsity term, in which the second column of images are obtained by $\|\cdot\|_1$ and the third column of images are acquired via $\|\cdot\|_{l_{2/3}}^{2/3}$. As demonstrated in our experiments later, $\|\cdot\|_{l_{2/3}}^{2/3}$ appears to be desirable in characterizing the sparsity.

To effectively promote the performance of Schatten- q norm under cluttered background, the Laplacian regularization $Tr(S\mathcal{M}_F S^T)$ is introduced for the salient object S , where a Laplacian matrix \mathcal{M}_F consists of an affinity matrix characterizing the similar features among image patches. This term is based on the local invariance assumption, which can improve the consistency of foreground, as shown in Fig. 5.

In summary, the proposed nonconvex model for the salient object detection can be formulated as

$$\begin{aligned} &\min_{L, S} \|L\|_{S_{2/3}}^{2/3} + \lambda \|S\|_{l_{2/3}}^{2/3} + \gamma Tr(S\mathcal{M}_F S^T) \\ &\text{s.t. } X = L + S. \end{aligned} \quad (10)$$

The problem (10) can be solved by various algorithms. However, most of existing algorithms often rely on iteration involving SVD, which incurs the largest computational complexity [21], [24], [27], [28]. Therefore, using factorization formulation (8), the model (10) can be further reformulated as

$$\begin{aligned} &\min_{U, V, S} \frac{1}{3} (2\|U\|_* + \|V\|_F^2) + \lambda \|S\|_{l_{2/3}}^{2/3} + \gamma Tr(S\mathcal{M}_F S^T) \\ &\text{s.t. } X = L + S, L = UV^T. \end{aligned} \quad (11)$$

It follows from the model (11) that the norms of bilinear factor matrix (i.e., $\|U\|_*$ and $\|V\|_F$) are convex, which corresponds to the problem of minimizing norm combination on two much smaller factor matrices. Obviously, the model (11) based on factorization has a lower computation cost than the minimization of (10) by directly computing the SVD on a full matrix.

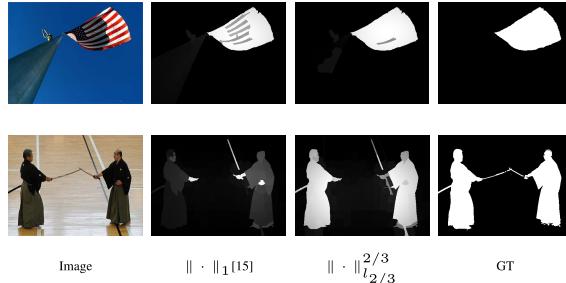


Fig. 4. Comparison results from using the different sparse regularization: l_1 -norm and $l_{2/3}$ -norm.

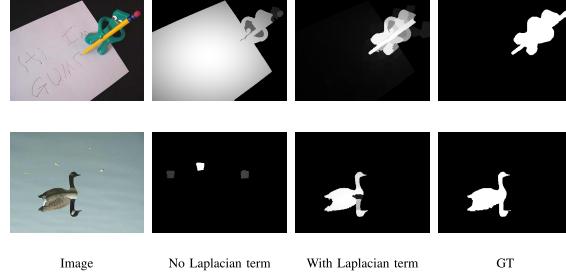


Fig. 5. Comparison of using Laplacian and non-Laplacian.

IV. OPTIMIZATION

In this section, the proposed model (11) can be optimized by the alternating direction method of multipliers (ADMMs) [32]. To solve (11), we introduce the auxiliary variables to split the interdependent terms so that they can be solved independently. Based on that, the problems (11) can be formulated as the following equivalent form:

$$\begin{aligned} & \min_{U, V, S, T, M} \frac{1}{3} (2\|T\|_* + \|V\|_F^2) \lambda \|S\|_{l_{2/3}}^{2/3} + \gamma \text{Tr}(M \mathcal{M}_F M^T) \\ & \text{s.t. } X = L + S, \quad L = U V^T, \quad T = U, \quad M = S \end{aligned} \quad (12)$$

where T and M are the corresponding auxiliary variables.

To remove the equality constraints, Lagrange multipliers Y_1 , Y_2 , Y_3 , and Y_4 are introduced. Thus, the augmented Lagrangian function for problem (12) is constructed as

$$\begin{aligned} \mathcal{L}_{2/3}(U, V, L, S, T, M, Y_1, Y_2, Y_3, Y_4, \mu) &= \frac{1}{3} (2\|T\|_* + \|V\|_F^2) + \lambda \|S\|_{l_{2/3}}^{2/3} + \gamma \text{Tr}(M \mathcal{M}_F M^T) \\ &+ \langle Y_1, F - L - S \rangle + \langle Y_2, L - U V^T \rangle + \langle Y_3, T - U \rangle \\ &+ \langle Y_4, S - M \rangle + \frac{\mu}{2} (\|F - L - S\|_F^2 + \|L - U V^T\|_F^2 \\ &+ \|T - U\|_F^2 + \|S - M\|_F^2) \end{aligned} \quad (13)$$

where $\mu > 0$ is a penalty parameter. To solve Eq.(13), we seek the optimal U , V , L , S , T , and M by a suitable iteration as shown in Algorithm 1. To be specific, we consider the following subproblems in sequence.

A. Updating U

With respect to U , the corresponding optimization problem is

$$\arg \min_U \|L - U V^T + \frac{Y_2}{\mu}\|_F^2 + \|T - U + \frac{Y_3}{\mu}\|_F^2. \quad (14)$$

Since (14) is a least squares problem, the optimal solution can be written as

$$U^* = \left(L V + M + \frac{Y_2 V + Y_3}{\mu} \right) (V^T V + I)^{-1}. \quad (15)$$

B. Updating V

Fixing the other variables, V can be given by

$$\arg \min_V \frac{1}{3} \|V\|_F^2 + \frac{\mu}{2} \left\| L^T - V U^T + \frac{Y_2^T}{\mu} \right\|_F^2. \quad (16)$$

Thus, we have

$$V^* = (\mu L^T U + Y_2^T U) \left(\frac{2}{3} I + \mu U^T U \right)^{-1}. \quad (17)$$

C. Updating L

With respect to L , we consider the following subproblem:

$$\arg \min_L \left\| F - L - S + \frac{Y_1}{\mu} \right\|_F^2 + \left\| L - U V^T + \frac{Y_2}{\mu} \right\|_F^2. \quad (18)$$

Hence, we obtain the corresponding solution as follows:

$$L^* = \frac{1}{2} \left(U V^T + F - S + \frac{Y_1 - Y_2}{\mu} \right). \quad (19)$$

D. Updating S

To update S , the minimization problem becomes

$$\begin{aligned} \arg \min_S \lambda \|S\|_{l_{2/3}}^{2/3} + \frac{\mu}{2} &\left(\left\| S - \left(F - L + \frac{Y_1}{\mu} \right) \right\|_F^2 \right. \\ &\left. + \left\| S - \left(M - \frac{Y_4}{\mu} \right) \right\|_F^2 \right). \end{aligned} \quad (20)$$

Set $A = (F - L + H/2) + (Y_1 + Y_4/2\mu)$ and $\tau = (\lambda/\mu)$, and the problem (20) can be translated into the following form:

$$\arg \min_S \frac{\lambda}{\mu} \|S\|_{l_{2/3}}^{2/3} + (\|S - A\|_F^2).$$

Thus, this problem can be solved by Theorem 1, as shown in the following.

Theorem 1: [24], [33] For any matrix A , $X \in \mathbb{R}^{m \times n}$, solution of the flowing minimization

$$\min_X \|X - A\|_F^2 + \tau \|X\|_{l_{2/3}}^{2/3} \quad (21)$$

is $\mathcal{T}(A)$, where the two-thirds-thresholding operator $\mathcal{T}(\cdot)$ is

$$\mathcal{T}_\tau(a) = \begin{cases} \frac{\text{sgn}(a) \left(\varphi_\tau(a) + \sqrt{\frac{2|a|}{\varphi_\tau(a)} - \varphi_\tau^2(a)} \right)^3}{8}, & |a| > \frac{2\sqrt[4]{3\tau^3}}{3} \\ 0, & \text{else} \end{cases} \quad (22)$$

where

$$\varphi_\tau(a) = \frac{2}{\sqrt{3}} \sqrt{\sqrt{\tau} \cosh \left(\arccos h \frac{\left(\frac{27a^2}{16}\tau - \frac{3}{2} \right)}{3} \right)}.$$

Algorithm 1 Problem (13) Solved by ADMM

Input: F, d, α, β and μ $M_0 = Y_2^0 = 0, N_0 = Y_3^0 = 0, S_0 = H_0 = Y_1^0 = Y_4^0 = 0, \mu_{\max} = 10^{10}, \rho = 1.1$ and $k = 0$

- 1: **while** not converged **do**
- 2: update U_{k+1} by (15)
- 3: update V_{k+1} by (17)
- 4: update L_{k+1} by (19)
- 5: update S_{k+1} by (22)
- 6: update T_{k+1} by (25)
- 7: update M_{k+1} by (28)
- 8: update $Y_i, 1 \leq i \leq 4$ by the following forms, i.e.

$$\begin{aligned} Y_1 &\leftarrow Y_1 + \mu(F - L - S) \\ Y_2 &\leftarrow Y_2 + \mu(L - UV^T) \\ Y_3 &\leftarrow Y_3 + \mu(T - U) \\ Y_4 &\leftarrow Y_4 + \mu(S - M) \end{aligned}$$

9: μ_{k+1} by $\mu \leftarrow \min(\rho\mu, \mu_{\max})$

10: k by $k \leftarrow k + 1$

11: **end while**

Output: L and S

E. Updating T

Fixing the other variables, with respect to T , we have

$$\arg \min_T \frac{2}{3} \|T\|_* + \frac{\mu}{2} \left\| T - \left(U - \frac{Y_3}{\mu} \right) \right\|_F^2. \quad (23)$$

In general, (23) can be represented as

$$\min_T \tau \|T\|_* + \frac{1}{2} \|T - C\|_F^2 \quad (24)$$

where $C = U - (Y_3/\mu)$ and $\tau = (2/3\mu)$. Thus, the optimization (24) is nuclear norm regularized least squares problem, which can be solved by the singular value thresholding (SVT) [34](refer Theorem 2).

Theorem 2 (SVT [34]): Let the SVD of C can be written as $C = U\Sigma V^T$, then the optimal solution for problem (24) is given by

$$T = \mathcal{D}_\tau(C) = U\mathcal{S}_\tau(\Sigma)V^T \quad (25)$$

where \mathcal{D}_ε is the SVT operator and \mathcal{D}_ε is the soft thresholding operator, which is defined as

$$\mathcal{S}_\varepsilon(x) = \begin{cases} x - \varepsilon, & x > \varepsilon \\ x + \varepsilon, & x < -\varepsilon \\ 0, & \text{else.} \end{cases} \quad (26)$$

F. Updating M

Finally, fixing the other variables, and updating M leads to the following problem:

$$\arg \min_M \gamma \text{Tr}(M\mathcal{M}_F M^T) + \frac{\mu}{2} \left\| S - M + \frac{Y_4}{\mu} \right\|_F^2. \quad (27)$$

Then, we have

$$M^* = (\mu S + Y_4)(2\gamma\mathcal{M}_F + \mu I)^{-1}. \quad (28)$$

For Algorithm 1, we perform SVD on a much smaller factored-matrix as a contrary to the full matrix used in WLRR [19], SMD [18], ULR [15], SLR [17], and LRR [16]. Thus, complexity analysis for Algorithm 1 can be summarized in the following.

Remark 1: Given an $m \times n$ matrix ($m \geq n$), the cost for the inverse of $d \times d$ matrix, the multiplication of $m \times d$ matrix, and $d \times n$ matrix are $O(mn^2)$, $O(d^3)$, and $O(mdn)$, respectively. In terms of WLRR [19], SMD [18], ULR [15], SLR [17], and LRR [16], the cost is $O(mn^2)$ since the thin SVD of an $m \times n$ matrix with $m \geq n$ is performed. While for Algorithm 1, the dominant cost of each iteration for updating $U \in \mathbb{R}^{m \times d}$, $V \in \mathbb{R}^{n \times d}$, and $T \in \mathbb{R}^{m \times d}$ using (15), (17), and (25) is $O(6mnd + 2d^3 + md^2 + nd^2)$. Thus, it is easy to observe that $O(6mnd + 2d^3 + md^2 + nd^2) \ll O(mn^2)$ for $m, n \gg d$, which implies that our proposed algorithm is more efficient.

V. ALGORITHM ANALYSIS

According to the definition of the critical point for non-convex functions [53], the variable sequence generated by Algorithm 1 can converge to a limiting point satisfying the Karush–Kuhn–Tucker (KKT) condition. Specifically, the detailed description can be summarized as Theorem 29.

Theorem 3: Let $\{(U^k, V^k, L^k, S^k, T^k, M^k, \{Y_i^k\})\}$ be a sequence generated by Algorithm 1. Suppose that the sequence $\{Y_i^k\}$ is bound, μ_k is nondecreasing, and $\sum_{k=0}^{\infty} (\mu_{k+1}/\mu_k^M) < \infty$, then the limit point $(U_\infty, V_\infty, L_\infty, S_\infty, T_\infty, M_\infty)$ of the above-mentioned sequence satisfies the following KKT conditions:

$$\begin{aligned} 0 &\in \frac{2}{3}\partial\|T_\infty\|_* + Y_3^\infty, \\ 0 &\in \lambda\partial\|S_\infty\|_{2/3}^{2/3} + Y_1^\infty + Y_4^\infty, \\ 0 &\in \gamma(M_\infty\mathcal{M}_F^T + M_\infty\mathcal{M}_F) + Y_4^\infty, \\ L_\infty &= U_\infty V_\infty^T, T_\infty = U_\infty, S_\infty = M_\infty, L_\infty + S_\infty = X_\infty. \end{aligned}$$

The proof of theorem can be found in the Appendix of this article. The conditions for nondecreasing μ_k and the boundedness of sequence $\{Y_i^k\}$ are given in [24]. Theorem 29 is established through an iteration algorithm in the inner loop (i.e., ADMM algorithm). When the inner loop is solvable exactly, the proof of convergence becomes relatively simple.

However, Theorem 29 cannot guarantee the global convergence, which is an essential issue in terms of the effectiveness of the proposed solvers for Algorithm 1. Recent studies [54], [55] provide the global convergence analysis under certain assumptions named coercivity, feasibility, Lipschitz subminimization paths, and objective regularity (see [55, Th. 1]). Inspired by these approaches and our own analysis, the global convergence for the proposed nonconvex model is shown in the following theorem.

For the ease of analysis, we first reformulate the proposed model (10) as a standard form coinciding with [54] and [55]. Hence, we have the augmented Lagrangian function

$$\mathcal{L}_\beta(L, S, \omega) = \Phi(L, S) + \langle \omega, X - L - S \rangle + \frac{\beta}{2} \|X - L - S\|_F^2 \quad (29)$$

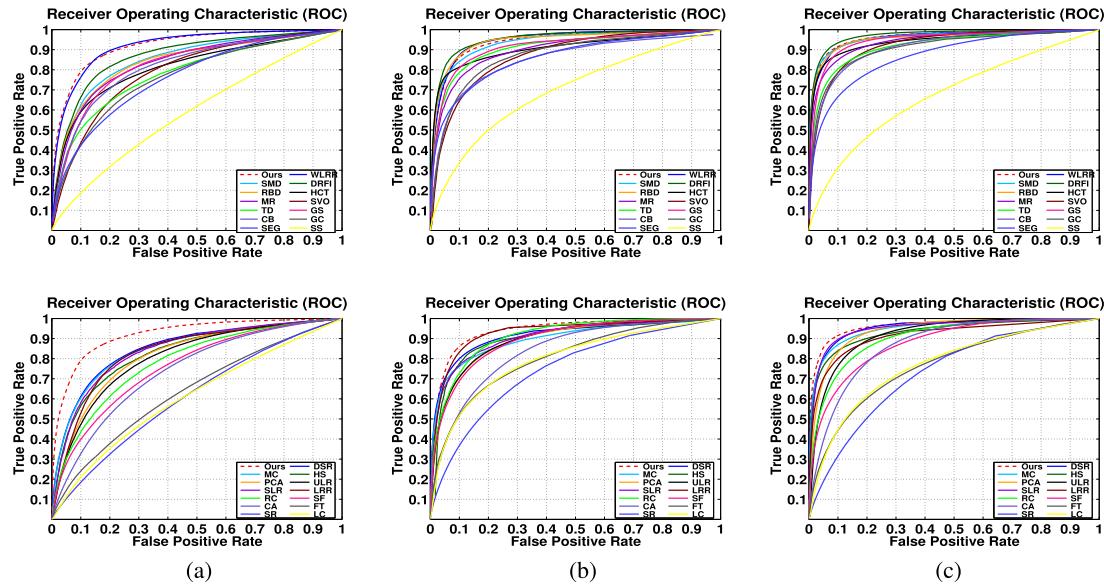


Fig. 6. ROC curve comparison.

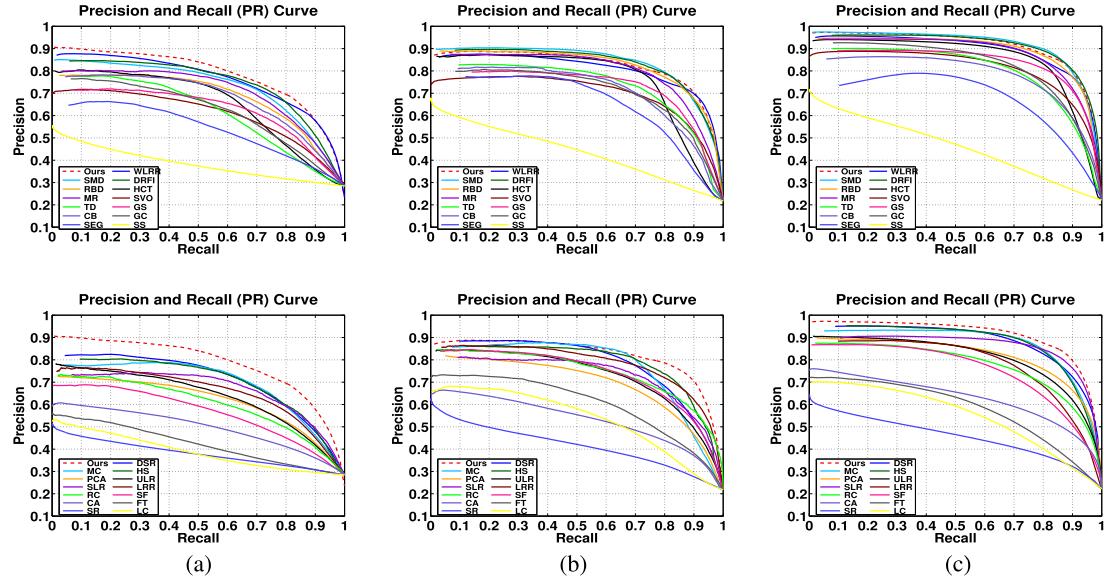


Fig. 7. PR curve comparison.

methods. The ROC and PR curves on three datasets are shown in Figs. 6 and 7, respectively. The scores about WF, OR, AUC, and MAE are given in Tables II–IV. Besides, Fig. 8 shows the overall qualitative comparisons.

Compared with the competing methods, the proposed model achieves the best results in terms of the four metrics (i.e., WF, AUC, OR, and MAE) on the ECSSD datasets, which is shown in Table II. In the ROC curves (the first row of Fig. 6), our model and WLRR [19] are the best two among those competing models. However, for PR curves (the first row of Fig. 7), our method gives the best performance. These results demonstrate that our model has the desired capacity in detecting the complete foreground from various complex scenes.

From Table III, we can find that our method achieves the best results with the highest OR and the lowest MAE, while SMD [18] and WLRR [19] obtain the highest WF and AUC,

respectively. From ROC and PR curves (the second row of Figs. 6 and 7), we can see that our model outperforms the low-rank-based methods. It is worth noting that our approach is also comparable to the other representative models in most cases across different criteria.

The last comparison is conducted on MSRA10K [9] with a single object. As shown in Table IV, our model provides the second-best in terms of WF, AUC, and OR, the third in MAE. The ROC and PR curves are entangled with DRFI [35] and SMD [18] in a wide range (the third rows of Figs. 6 and 7). Especially, the result of DRFI [35] is only a reference, which is a supervised top-down model. These results represent that the proposed model has better consistency in terms of performance.

Fig. 8 gives the qualitative comparisons from the best methods in the experiments. For single-object images (e.g., the images in the first three rows), salient maps of the

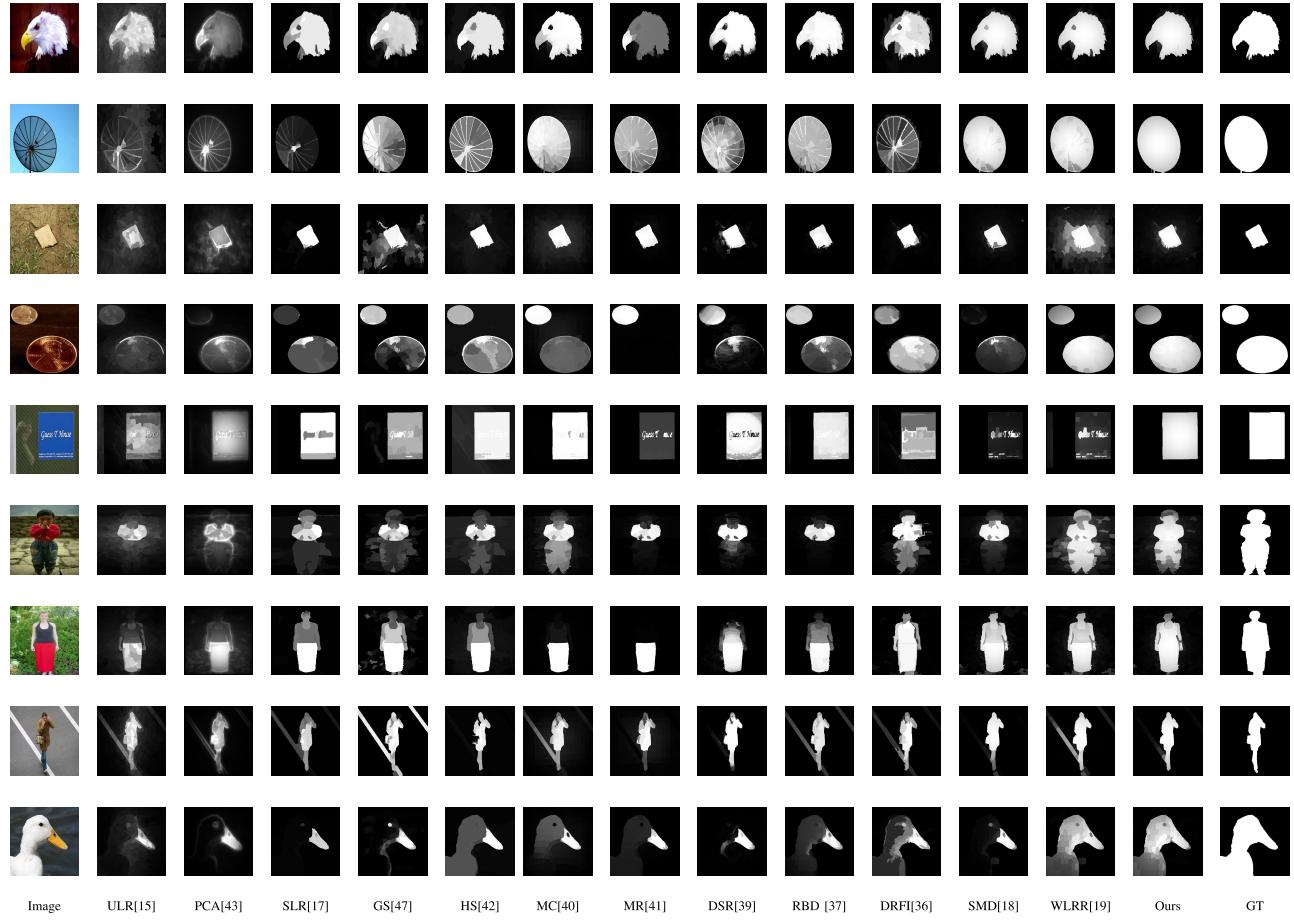


Fig. 8. Visual quality comparison of saliency maps from 13 methods.

new model represent the complete salient object with a few scattered and incomplete patches. For the object with different representations (e.g., the images from the fourth row to the sixth row), salient maps of the new model possess consistent values within the same foreground. For the images with various complex scenes (e.g., image in the last three rows), several other methods miss the detecting parts (such as ULR [15], SLR [17], DSR [38], and SMD [18]) or confuse background with the foreground (e.g., DRFI [35] and WLRR [19]). By contrast, the proposed method successfully identifies all desirable salient objects with consistent values. For the images whose foreground and background share similar appearance (such as images in the third and the fourth row), our new approach is more capable of detecting the complete foreground from the background in a consistent way. These outcomes clearly demonstrate the effectiveness of our algorithm in handling images with various complex scenes.

VII. CONCLUSION

In this article, we propose an unsupervised learning model for salient object detection by using the bilinear factor matrix norm. Different from the currently existing literature, the nonconvex Schatten- $2/3$ norm is adopted to characterize the background information and is shown to be better in capturing the low-rank structure of the background details, leading to a better background representation. The nonconvex $l_{2/3}$ norm is incorporated to formulate the sparsity and consistency of

foreground. To promote the further performance of Schatten- $2/3$ norm and nonconvex $l_{2/3}$ norm, a Laplacian constraint is adopted to the salient object. It is worth noting that the choice of $p = q = 2/3$ makes the optimization to be tractable and reduces the required computational complexity. In addition, under our framework by splitting the balanced variables, the optimization process, associated with the ADMM, requires computing the SVD on a much small factored matrix, and thus, greatly reduces the computational cost. Meanwhile, the convergence of the algorithm is mathematically analyzed and validated. Experiments on the MSRA10K, ECSSD, and iCoSeg datasets show that the proposed model has performed well and appears to be highly consistent for detecting various salient objects.

A further direction of our future work is to study the salient object detection problem under the tensor framework and to explore the tensor rank by conducting multiscale low-rank decomposition. Besides, the adaptive selection of tradeoff parameters is also an important subject of study. Such studies will be reported elsewhere.

APPENDIX PROOF OF THEOREM 29

Proof: We first show that $\{U^k\}$, $\{V^k\}$, $\{L^k\}$, $\{S^k\}$, $\{T^k\}$, and $\{M^k\}$ are convergent Cauchy sequences.

According to boundedness of multipliers and some variables produced by alternating direction method of multiplier

(ADMM) (we can refer to [24]), the corresponding Lagrange parameter $\{Y_i^k\}$ is bounded. Due to μ_k is nondecreasing, $\sum_{k=0}^{\infty} (\mu_{k+1}/\mu_k^2) < \infty$ and

$$Y_3^{k+1} = Y_3^k + \mu_k(T^{k+1} - U^{k+1})$$

one can have

$$\begin{aligned} & \sum_{k=0}^{\infty} \|T^{k+1} - U^{k+1}\|_F \\ &= \left\langle \frac{Y_3^{k+1} - Y_3^k}{\mu_k}, \frac{Y_3^{k+1} - Y_3^k}{\mu_k} \right\rangle^{1/2} \\ &= \sum_{k=0}^{\infty} \frac{1}{\mu_k} \|Y_3^{k+1} - Y_3^k\|_F \leq \sum_{k=0}^{\infty} \frac{\mu_{k+1}}{\mu_k^2} \|Y_3^{k+1} - Y_3^k\|_F < \infty. \end{aligned}$$

Similarly, we can obtain

$$\begin{aligned} & \sum_{k=0}^{\infty} \|L^{k+1} + S^{k+1} - X^{k+1}\|_F \\ &\leq \sum_{k=0}^{\infty} \frac{\mu_{k+1}}{\mu_k^2} \|Y_1^{k+1} - Y_1^k\|_F < \infty \\ & \sum_{k=0}^{\infty} \|U^{k+1}(V^{k+1})^T - L^{k+1}\|_F \\ &\leq \sum_{k=0}^{\infty} \frac{\mu_{k+1}}{\mu_k^2} \|Y_2^{k+1} - Y_2^k\|_F < \infty \\ & \sum_{k=0}^{\infty} \|S^{k+1} - M^{k+1}\|_F \\ &\leq \sum_{k=0}^{\infty} \frac{\mu_{k+1}}{\mu_k^2} \|Y_4^{k+1} - Y_4^k\|_F < \infty \end{aligned}$$

which implies that

$$\begin{aligned} & \lim_{k \rightarrow \infty} \|L^{k+1} + S^{k+1} - X^{k+1}\|_F = 0 \\ & \lim_{k \rightarrow \infty} \|U^{k+1}(V^{k+1})^T - L^{k+1}\|_F = 0 \\ & \lim_{k \rightarrow \infty} \|T^{k+1} - U^{k+1}\|_F = 0 \\ & \lim_{k \rightarrow \infty} \|S^{k+1} - M^{k+1}\|_F = 0. \end{aligned} \quad (30)$$

Thus, $\{(U^k, V^k, L^k, S^k, T^k, M^k)\}$ approximates to a feasible solution.

Since

$$Y_2^k = Y_2^{k-1} + \mu_{k-1}(L^k - U^k(V^k)^T) \quad (31)$$

and

$$Y_3^k = Y_3^{k-1} + \mu_{k-1}(T^k - U^k) \quad (32)$$

the minimization problem (14) can be converted to seek the corresponding first-order optimal condition for U . That is

$$\begin{aligned} & (L^k - U^{k+1}(V^k)^T + \mu_k^{-1}Y_2^k)V^k + (T^k - U^{k+1} + \mu_k^{-1}Y_3^k) \\ &= T^k - U^{k+1} + U^k - U^k + \mu_k^{-1}Y_3^k \\ &+ \left(U^k(V^k)^T + \frac{Y_2^k - Y_2^{k-1}}{\mu_{k-1}} - U^{k+1}(V^k)^T + \mu_k^{-1}Y_2^k \right)V^k \\ &= \frac{Y_3^k - Y_3^{k-1}}{\mu_{k-1}} - U^{k+1} + U^k + \mu_k^{-1}Y_3^k \\ &+ \left(U^k(V^k)^T + \frac{Y_2^k - Y_2^{k-1}}{\mu_{k-1}} - U^{k+1}(V^k)^T + \mu_k^{-1}Y_2^k \right)V^k \\ &= (U^k - U^{k+1})\left(I + V^k(V^k)^T\right) \\ &+ \frac{Y_3^k - Y_3^{k-1}}{\mu_{k-1}} + \mu_k^{-1}Y_3^k + \left(\frac{Y_2^k - Y_2^{k-1}}{\mu_{k-1}} + \mu_k^{-1}Y_2^k \right)V^k \\ &= 0. \end{aligned}$$

Therefore, we arrive at

$$\begin{aligned} & U^{k+1} - U^k \\ &= \left(I + (V^k)^T V^k \right)^{-1} \\ & \cdot \left[\frac{Y_3^{k-1} - Y_3^k}{\mu_{k-1}} - \mu_k^{-1}Y_3^k + \left(\frac{Y_2^{k-1} - Y_2^k}{\mu_{k-1}} - \mu_k^{-1}Y_2^k \right)V^k \right]. \end{aligned}$$

Let

$$\begin{aligned} r_1 &= \max_{k \in \mathbb{Z}_+, k=1, 2, \dots} \left\{ \left\| \left(I + (V^k)^T V^k \right)^{-1} \right\|_F \right. \\ &\quad \cdot \left(\|Y_3^k\|_F + \mu_{k-1}^{-1} \mu_k^{-1} \|Y_3^k - Y_3^{k-1}\|_F \right. \\ &\quad \left. \left. + (\mu_{k-1}^{-1} \mu_k^{-1} \|Y_2^{k-1} - Y_2^k\|_F + \|Y_2^k\|_F) \|V^k\|_F \right) \right\}, \end{aligned}$$

due to $\sum_{k=0}^{\infty} (\mu_{k+1}/\mu_k^2) < \infty$, we have the estimates

$$\sum_{k=0}^{\infty} \|U^{k+1} - U^k\|_F \leq \sum_{k=0}^{\infty} \mu_k^{-1} r_1 \leq \sum_{k=0}^{\infty} \frac{\mu_{k+1}}{\mu_k^2} r_1 < \infty.$$

Thus, we obtain

$$\lim_{k \rightarrow \infty} \|U^{k+1} - U^k\|_F = 0. \quad (33)$$

This shows that $\{U^k\}$ is convergent Cauchy sequence. Similarly, for $\{V^k\}$, $\{L^k\}$, $\{S^k\}$, $\{T^k\}$, and $\{M^k\}$, one can get

$$\begin{aligned} & \lim_{x \rightarrow \infty} \|V^{k+1} - V^k\|_F = 0 \\ & \lim_{x \rightarrow \infty} \|L^{k+1} - L^k\|_F = 0 \\ & \lim_{x \rightarrow \infty} \|S^{k+1} - S^k\|_F = 0 \\ & \lim_{x \rightarrow \infty} \|T^{k+1} - T^k\|_F = 0 \\ & \lim_{x \rightarrow \infty} \|M^{k+1} - M^k\|_F = 0 \end{aligned} \quad (34)$$

which indicates that they are also convergent Cauchy sequences.

Next, we prove that the limit point of sequence $\{(U^k, V^k, L^k, S^k, T^k, M^k)\}$ satisfies the Karush–Kuhn–Tucker conditions.

Suppose $(U_*, V_*, L_*, S_*, T_*, M_*)$ is the critical point of optimization problem (13). According to Fermat's rule [24], the first-order optimal conditions for subproblems (19), (22), and (26) are given by

$$\begin{aligned} 0 &\in \frac{2}{3}\partial\|T_*\|_* + Y_3^* \\ 0 &\in \lambda\partial\|S_*\|_{l_{2/3}}^{2/3} + Y_1^* + Y_4^* \\ 0 &\in \gamma(M_*\mathcal{M}_F^T + M_*\mathcal{M}_F) + Y_4^* \\ L_* &= U_*V_*^T, T_* = U_*, S_* = M_*, L_* + S_* = X_*. \end{aligned} \quad (35)$$

Let $U_\infty, V_\infty, L_\infty, S_\infty, T_\infty$, and M_∞ be the limit points of the above-mentioned sequence, the limit points satisfy the following first-order optimal conditions:

$$\begin{aligned} 0 &\in \frac{2}{3}\partial\|T_\infty\|_* + Y_3^\infty \\ 0 &\in \lambda\partial\|S_\infty\|_{l_{2/3}}^{2/3} + Y_1^\infty + Y_4^\infty \\ 0 &\in \gamma(M_\infty\mathcal{M}_F^T + M_\infty\mathcal{M}_F) + Y_4^\infty \\ L_\infty &= U_\infty V_\infty^T, T_\infty = U_\infty, S_\infty = M_\infty, L_\infty + S_\infty = X_\infty \end{aligned}$$

and thus, the proof is completed. \square

PROOF OF THEOREM 4

Proof: It is well known that $\mathcal{L}_\beta(L^k, S^k, \omega^k)$ decreases monotonically due to $\mathcal{L}_\beta(L^k, S^k, \omega^k) \geq \mathcal{L}_\beta(L^{k+1}, S^{k+1}, \omega^{k+1})$. Next, we first give that $\mathcal{L}_\beta(L^k, S^k, \omega^k)$ is lower bounded.

Since $\Phi(L, S)$ is the restricted prox-regular function, some properties named, respectively, coercivity, feasibility, and Lipschitz subminimization paths hold for this objective function [55]. By Lipschitz subminimization paths, there exists S' such that $X = L^k + S'$ and $S' = H(S')$ (where H is a Lipschitz continuous map). Using coercivity, feasibility, and Lipschitz subminimization paths of objective function, for $d_S^k \in \partial\Phi_S$, we have

$$\begin{aligned} \mathcal{L}_\beta(L^k, S^k, \omega^k) &= \Phi(L^k, S^k) + \langle \omega^k, X - L^k - S^k \rangle \\ &\quad + \frac{\beta}{2}\|X - L^k - S^k\|_F^2 \\ &= \Phi(L^k, S^k) + \langle d_S^k, S' - S^k \rangle + \frac{\beta}{2}\|S' - S^k\|_F^2 \\ &\geq \Phi(L^k, S') + \frac{\beta}{4}\|X - L^k - S'\|_F^2 > -\infty. \end{aligned} \quad (36)$$

This shows that $\mathcal{L}_\beta(L^k, S^k, \omega^k)$ is low bounded, where upper bound is $\mathcal{L}_\beta(L^0, S^0, \omega^0)$. From the property of coercivity, we know that (L^k, S^k, ω^k) is bounded. Hence, there exist a convergent subsequence and a limit point, denoted by $(L^{k_i}, S^{k_i}, \omega^{k_i})_{i \in N} \rightarrow (L^*, S^*, \omega^*)$ as $i \rightarrow +\infty$.

Second, we show that $\mathcal{L}_\beta(L^k, S^k, \omega^k)$ is descent

$$\begin{aligned} \mathcal{L}_\beta(L^k, S^k, \omega^k) - \mathcal{L}_\beta(L^{k+1}, S^{k+1}, \omega^{k+1}) &= \Phi(L^k, S^k) - \Phi(L^{k+1}, S^{k+1}) + \langle \omega^{k+1}, L^k - L^{k+1} \rangle \\ &\quad + \langle \omega^{k+1}, S^k - S^{k+1} \rangle + \langle \omega^k - \omega^{k+1}, X - L^k - S^k \rangle \\ &\quad + \frac{\beta}{2}\|L^{k+1} - L^k\|_F^2 - \frac{\beta}{2}\|S^{k+1} - S^k\|_F^2 \\ &\geq -\frac{1}{\beta}\|\omega^k - \omega^{k+1}\|_F^2 + \frac{\beta - L_\Phi M}{2} \\ &\quad \times \left\{ \|L^k - L^{k+1}\|_F^2 + \|S^k - S^{k+1}\|_F^2 \right\} \end{aligned} \quad (37)$$

where M and L_Φ are Lipschitz constants. Since $\omega^k = -\partial\Phi_S$ for any k , we obtain the following inequality:

$$\|\omega^k - \omega^{k+1}\|_F \leq L_\Phi(\|L^k - L^{k+1}\|_F + \|S^k - S^{k+1}\|_F). \quad (38)$$

In summary, we have

$$\begin{aligned} \mathcal{L}_\beta(L^k, S^k, \omega^k) - \mathcal{L}_\beta(L^{k+1}, S^{k+1}, \omega^{k+1}) &\geq \left(\frac{\beta - L_\Phi M}{2} - \frac{L_\Phi M}{\beta} \right) \|L^k - L^{k+1}\|_F^2 \\ &\quad + \left(\frac{\beta - L_\Phi M}{2} - \frac{L_\Phi M}{\beta} \right) \|S^k - S^{k+1}\|_F^2 \end{aligned} \quad (39)$$

where $\beta > \max\{1, 3L_\Phi M\}$. From (39), we know that $\mathcal{L}_\beta(L^k, S^k, \omega^k)$ is descent. Therefore, $\|L^k - L^{k+1}\|_F \rightarrow 0$ and $\|S^k - S^{k+1}\|_F \rightarrow 0$ as $k \rightarrow 0$.

Third, we give subgradient bound and limiting continuity for $\mathcal{L}_\beta(L^k, S^k, \omega^k)$. Since Φ is the Lipschitz differentiable, these two properties are trivial. In fact, subgradient bound can be written as

$$\begin{aligned} &\|\partial_L \mathcal{L}_\beta(L^{k+1}, S^{k+1}, \omega^{k+1})\|_F \\ &\leq (\beta + 2L_\Phi M)(\|L^{k+1} - L^k\|_F + \|S^{k+1} - S^k\|_F) \\ &\|\partial_\omega \mathcal{L}_\beta(L^{k+1}, S^{k+1}, \omega^{k+1})\|_F \\ &\leq \frac{L_\Phi M}{\beta}(\|L^{k+1} - L^k\|_F + \|S^{k+1} - S^k\|_F) \\ &\|\partial_S \mathcal{L}_\beta(L^{k+1}, S^{k+1}, \omega^{k+1})\|_F \\ &\leq L_\Phi M(\|L^{k+1} - L^k\|_F + \|S^{k+1} - S^k\|_F). \end{aligned} \quad (40)$$

Therefore, there exists $d^k \in \partial\mathcal{L}_\beta(L^k, S^k, \omega^k)$ such that $\|d^k\|_F \rightarrow 0$. In particular, $\|d_i^k\|_F \rightarrow 0$ as $i \rightarrow 0$. Based on the Lipschitz differentiable function, if (L^*, S^*, ω^*) is the limit point of a subsequence $(L^{k_i}, S^{k_i}, \omega^{k_i})_{i \in N}$, then $\mathcal{L}_\beta(L^*, S^*, \omega^*) = \lim_i \mathcal{L}_\beta(L^{k_i}, S^{k_i}, \omega^{k_i})$. By definition of subgradient, we obtain $0 \in \partial\mathcal{L}_\beta(L^*, S^*, \omega^*)$.

Finally, we check that \mathcal{L}_β is a Kurdyka–Lojasiewicz function (abbreviated as KL function), which is central to the global convergence analysis. We know that functions satisfying the KL inequality are, respectively, real-analytic functions, semi-algebraic functions, and locally strongly convex functions [59].⁹ Since \mathcal{L}_β is the sum of real-analytic and semi-algebraic functions, \mathcal{L}_β satisfies the KL inequality. Hence, the sequence (L^k, S^k, ω^k) converges globally to the limit point under the KL assumption.¹⁰ \square

ACKNOWLEDGMENT

The authors, Min Li, Yao Zhang, and Xiaoli Sun, would like to thank the hospitality from the Department of Mathematics, Southern Illinois University, Carbondale, IL, USA, during their visit from December 2017 to December 2019, and this joint work was conducted during the visit.

⁹The definition of KL function and some examples are introduced in detail in [59].

¹⁰The proofs of (37) and (40) are similar to the Lemmas 4, 5, and 10, respectively, in [55]. Some properties used in the above-mentioned proof can be found in [55].

REFERENCES

- [1] A. Borji, M.-M. Cheng, Q. Hou, H. Jiang, and J. Li, "Salient object detection: A survey," *Comput. Vis. Media*, vol. 5, no. 2, pp. 117–150, Jun. 2019.
- [2] G. Li and Y. Yu, "Contrast-oriented deep neural networks for salient object detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 12, pp. 6038–6051, Dec. 2018.
- [3] Q. Hou, M. Cheng, X. Hu, A. Borji, Z. Tu, and P. H. S. Torr, "Deeply supervised salient object detection with short connections," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 4, pp. 815–828, Apr. 2019.
- [4] X. Hu, L. Zhu, J. Qin, C.-W. Fu, and P.-A. Heng, "Recurrently aggregating deep features for salient object detection," in *Proc. AAAI*, 2018, pp. 6943–6950.
- [5] S. Huo, Y. Zhou, W. Xiang, and S. Y. Kung, "Semisupervised learning based on a novel iterative optimization model for saliency detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 225–241, Jan. 2019.
- [6] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
- [7] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915–1926, Oct. 2012.
- [8] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 194–201, Jan. 2012.
- [9] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, Mar. 2015.
- [10] S. Huo, Y. Zhou, J. Lei, N. Ling, and C. Hou, "Iterative feedback control-based salient object segmentation," *IEEE Trans. Multimedia*, vol. 20, no. 6, pp. 1350–1364, Jun. 2018.
- [11] W. Wang, C. Yang, H. Chen, and X. Feng, "Unified discriminative and coherent semi-supervised subspace clustering," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2461–2470, May 2018.
- [12] J. Yang and M.-H. Yang, "Top-down visual saliency via joint CRF and dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 576–588, Mar. 2017.
- [13] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.
- [14] E. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, pp. 1–39, Jun. 2011.
- [15] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 853–860.
- [16] C. Lang, G. Liu, J. Yu, and S. Yan, "Saliency detection by multitask sparsity pursuit," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1327–1338, Mar. 2012.
- [17] W. Zou, K. Kpalma, Z. Liu, and J. Ronsin, "Segmentation driven low-rank matrix recovery for saliency detection," in *Proc. Brit. Mach. Vis. Conf.*, 2013, pp. 1–13.
- [18] H. Peng, B. Li, H. Ling, W. Hu, W. Xiong, and S. J. Maybank, "Salient object detection via structured matrix decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 818–832, Apr. 2017.
- [19] C. Tang, P. Wang, C. Zhang, and W. Li, "Salient object detection via weighted low rank matrix recovery," *IEEE Signal Process. Lett.*, vol. 24, no. 4, pp. 490–494, Apr. 2017.
- [20] M. Li, Y. Zhang, M. Xiao, C. Xu, and W. Zhang, "On Schatten- q quasi-norm induced matrix decomposition model for salient object detection," *Pattern Recognit.*, vol. 96, pp. 1–12, Dec. 2019.
- [21] C. Lu, J. Tang, S. Yan, and Z. Lin, "Generalized nonconvex nonsmooth low-rank minimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 4130–4137.
- [22] F. Nie, H. Wang, X. Cai, H. Huang, and C. Ding, "Robust matrix completion via joint Schatten p -norm and ℓ_p -norm minimization," in *Proc. IEEE 12th Int. Conf. Data Mining*, Dec. 2012, pp. 566–574.
- [23] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Statist. Assoc.*, vol. 96, no. 456, pp. 1348–1360, 2001.
- [24] F. Shang, J. Cheng, Y. Liu, Z. Luo, and Z. Lin, "Bilinear factor matrix norm minimization for robust PCA: Algorithms and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 9, pp. 2066–2080, Sep. 2018.
- [25] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-Laplacian priors," in *Proc. 22nd Int. Conf. Neural Inf. Process. Syst. (NIPS)*, 2009, pp. 1033–1041.
- [26] Z. Xu, F. X. X. Chang, and H. Zhang, " $L_{1/2}$ regularization: A thresholding representation theory and a fast solver," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1013–1027, Jul. 2012.
- [27] R. Sun and Z.-Q. Luo, "Guaranteed matrix completion via non-convex factorization," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6535–6579, Nov. 2016.
- [28] C. Xu, Z. Lin, and H. Zha, "A unified convex surrogate for the Schatten- p norm," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 926–932.
- [29] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 663–670.
- [30] Q. Yao, J. T. Kwok, T. Wang, and T. Y. Liu, "Large-scale low-rank matrix learning with nonconvex regularizers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2628–2643, Nov. 2019.
- [31] H. Zhang, J. Yang, F. Shang, C. Gong, and Z. Zhang, "LRR for subspace segmentation via tractable Schatten- p norm minimization and factorization," *IEEE Trans. Cybern.*, vol. 49, no. 5, pp. 1722–1734, May 2019.
- [32] R. Vidal, Y. Ma, and S. Sastry, *Generalized Principal Component Analysis*. New York, NY, USA: Springer, 2016.
- [33] W. C. J. Sun and Z. Xu, "Fast image deconvolution using closed-form thresholding formulas of L_q ($q = \frac{1}{2}, \frac{2}{3}$) regularization," *J. Vis. Commun. Image Represent.*, vol. 24, pp. 31–41, Jan. 2013.
- [34] J. Cai, E. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [35] J. Wang, H. Jiang, Z. Yuan, M.-M. Cheng, X. Hu, and N. Zheng, "Salient object detection: A discriminative regional feature integration approach," *Int. J. Comput. Vis.*, vol. 123, no. 2, pp. 251–268, 2017.
- [36] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2814–2821.
- [37] J. Kim, D. Han, Y.-W. Tai, and J. Kim, "Salient region detection via high-dimensional color transform," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 883–890.
- [38] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2976–2983.
- [39] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang, "Saliency detection via absorbing Markov chain," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1665–1672.
- [40] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3166–3173.
- [41] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1155–1162.
- [42] R. Margolin, A. Tal, and L. Zelnik-Manor, "What makes a patch distinct?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1139–1146.
- [43] C. Scharfenberger, A. Wong, K. Fergani, J. S. Zelek, and D. A. Clausi, "Statistical textural distinctiveness for salient region detection in natural images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 979–986.
- [44] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and N. Crook, "Efficient salient region detection with soft image abstraction," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1529–1536.
- [45] K.-Y. Chang, T.-L. Liu, H.-T. Chen, and S.-H. Lai, "Fusing generic objectness and visual saliency for salient object detection," in *Proc. 14th IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 914–921.
- [46] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 29–42.
- [47] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 733–740.
- [48] H. Jiang, J. Wang, Z. Yuan, T. Liu, and N. Zheng, "Automatic salient object segmentation based on context and shape prior," in *Proc. Brit. Mach. Vis. Conf.*, 2011, pp. 110.1–110.12.
- [49] E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä, "Segmenting salient objects from images and videos," in *Proc. 11th Eur. Conf. Comput. Vis.*, 2010, pp. 366–379.
- [50] R. Achanta, S. Hemami, F. Estrada, and S. Sussstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1597–1604.
- [51] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

- [52] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proc. 14th Annu. ACM Int. Conf. Multimedia (MULTIMEDIA)*, 2006, pp. 815–824.
- [53] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Math. Program.*, vol. 146, nos. 1–2, pp. 459–494, 2014.
- [54] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM J. Optim.*, vol. 26, no. 1, pp. 337–364, 2016.
- [55] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *J. Sci. Comput.*, vol. 78, pp. 29–63, Jan. 2019.
- [56] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen, "Interactively co-segmentating topically related images with intelligent scribble guidance," *Int. J. Comput. Vis.*, vol. 93, no. 3, pp. 273–292, Jul. 2011.
- [57] G. Li and Y. Yu, "Deep contrast learning for salient object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 478–487.
- [58] Z. Luo, A. Mishra, A. Achkar, J. Eichel, S. Li, and P.-M. Jodoin, "Non-local deep features for salient object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6609–6617.
- [59] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1758–1789, Jan. 2013.



Min Li received the Ph.D. degree in applied mathematics from Xidian University, Xi'an, China, in 2008.

She is currently an Associate Professor with the College of Mathematics and Statistics, Shenzhen University, Shenzhen, China. Her research interests include the field of partial differential equations, geometric multiscale analysis, subspace clustering, machine learning, saliency detection, and computer vision.



Yao Zhang received the master's degree from the College of Mathematics Statistics, Shenzhen University, Shenzhen, China, in 2019. He is currently pursuing the Ph.D. degree with the Key Laboratory of Solar Activity, National Astronomical Observatories, Chinese Academy of Sciences, Beijing, China.

His research interests include mathematics, computer vision, and solar radio astronomy.



Mingqing Xiao (Senior Member, IEEE) received the Ph.D. degree in applied mathematics from the University of Illinois at Urbana–Champaign, Champaign, IL, USA, in 1997.

He is currently a Professor with Southern Illinois University, Carbondale, IL, USA. His current research interests are optimization, image process, control theory, and numerical analysis.

Dr. Xiao serves in many editorial boards of various top journals, such as the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, *Automatica*, and the *European Journal of Control*.



Weiqiang Zhang received the Ph.D. degree in applied mathematics from Xidian University, Xi'an, China, in 2007.

He is currently an Associate Professor with the College of Mathematics and Statistics, Shenzhen University, Shenzhen, China. His research interests include the field of wavelet, geometric multiscale analysis, and computer vision.



Xiaoli Sun received the Ph.D. degree in applied mathematics from Xidian University, Xi'an, China, in 2007.

She is currently an Associate Professor with the College of Mathematics and Statistics, Shenzhen University, Shenzhen, China. Her research interests include wavelet analysis, subspace clustering, and computer vision.