On Schatten-q quasi-norm induced matrix decomposition model for salient object detection

Min Li[a], Yao Zhang[a], Mingqing Xiao[b,*], Chen Xu[a,*], Weiqiang Zhang[a]

[a] College of Mathematics and Statistics, Shenzhen Key Laboratory of Advanced Machine Learning and Applications, Shenzhen University, Shenzhen 518060, China
[b] Department of Mathematics, Southern Illinois University, Carbondale, IL 62901, US

ARTICLE INFO

ABSTRACT

Salient object detection is not only important but also challenging tasks in the study of computer vision. In this paper, different from existing approaches, we propose a novel regularization model for the salient object detection, which integrates a weighted group sparsity with the convex Schatten-1 or the non-convex Schatten-2/3 and Schatten-1/2 norm, respectively. A weighted group sparsity induced norm developed in this paper is shown to be able to make the foreground being consistent within the same image patches by effectively absorbing the image geometrical structure as well as the feature similarity. The Schatten quasi-norm is successfully used to capture the lower rank of background via factorization technique, and an alternative non-convex formulation for nuclear norm is studied. The corresponding alternative direction method of multiplier (ADMM) with derived solutions are discussed in detail, and the convergence of algorithm is validated. Extensive experiments on the six widely used datasets show that the proposed approach has capacity in outperforming most of state-of-the-art models in current literature.

1. Introduction

It is instinctive that human vision system can effectively capture important information from visual scenes. Inspired by this biological capability, salient object detection, which aims to localize and segment the most conspicuous foreground objects from background, has received considerable attention in cognitive psychology, neurobiology and computer vision. Simultaneously, this feature has been widely used in image cropping [1], adaptive image display on mobile devices [2], extracting dominant colors on the object of interest for web image filter [3], video surveillance [4], motion segmentation [5,6], and among others.

In general, most of existing researches on the problem of salient object detection [7] falls into two main categories: bottom-up and top-down, which respectively focuses either on independent goals without prior knowledge or on specific goals with prior knowledge. The bottom-up models [8–14] are stimulus-driven, which are shown to be effective in predicting human fixations and highlighting the informative regions of images. The top-down models [15–26] are task-driven, which are suitable in using image contexts and specific visual priors. However, both top-down and bottom-up approaches have the limited ability in dealing with complex scenes. For example, the bottom-up approach is only to be able to detect certain part of target objects and may not be able to distinguish the background from the salient object in complex scenes effectively. Due to the high diversity of various object type in complex scenes, top-down models usually are not suitable for scalability and generalization.

Therefore, approaches which combine bottom-up cues with top-down priors become more practical. In literature, the most representative work was by Candès, in which he introduced the low rank matrix recovery theory into the study of salient object detection (denoted as LRMR) problems [27,28]. Motivated by LRMR method, Shen et al. [29] proposed a unified approach based on low rank matrix recovery (ULR), which effectively incorporated traditional low-level feature with high-level guidance. Lang et al. [30] introduced a multi-task sparsity pursuit method (MSP) based on low-rank representation (LRR) [31] which seamlessly integrates multiple features with top-down priors to produce jointly the saliency map with a single inference step. Zou et al. [32] presented a fully unsupervised model that exploited bottom-up segmentation as a guidance cue of the matrix recovery (SLR). Peng et al. gave a structured matrix decomposition approach with a tree-structured sparsity-inducing regularization and a Laplacian
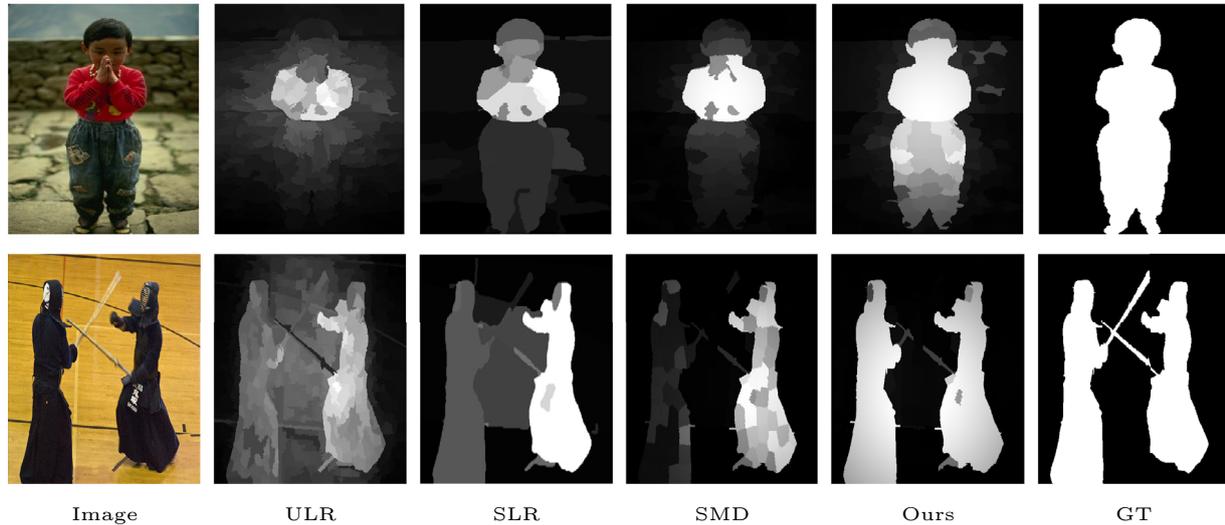
Fig. 1. Examples of our method performs close to the ground true.

regularization (SMD) [33]. As an extension of the SMD model, Sun et al. [34] elaborated the matrix decomposition model based on diversity which mainly considered the half regularization for background. The shared characteristic of these methods is to use low rank to characterize the background. Technically, it is a common consensus that the background should be a highly redundant part of information which has visually consistent representation in computer vision. Hence, according to linear algebra theory, this potential redundant part should lie in some much lower dimensional subspace, which can be efficiently characterized by a low rank feature matrix.

It is well known that aforementioned low rank approaches can be formulated as a rank minimization problem, which not only is non-convex but also can be an NP-hard problem. To make such a problem be tractable, a standard approach is to replace the rank function by its lower convex envelop, i.e. the nuclear norm [27–31,33]. However, such an approach may over-penalize the large singular values due to the gap between the rank function and its convex envelop, and result in deviating from the original solution, as discussed in [35–37]. More specifically, convex relaxation by using nuclear norm for salient object detection suffers from the following limitations:

1. Since the nuclear norm is a linear sum of singular values, minimizing the nuclear norm may lead to ignoring the inter-correlation between elements in foreground objects due to the lack of distinction of singular values, and thus the generated saliency object is scattered and incomplete, as shown in Fig. 1 (ULR).
2. When the background is complex or cluttered (for instance when it has similar appearance with the salient object), it is likely to be confused by the foreground target due to insufficient accuracy caused by the gap, as shown in Fig. 1 (ULR and SLR).
3. When the object possesses different representations, the conspicuous values with the same object could appear to be inconsistent, again due to insufficient accuracy caused by the gap, as shown in Fig. 1 (ULR, SLR and SMD).

In addition, the existing algorithms for current low-rank based models have to be solved iteratively and involve singular value decomposition (SVD) of a large scale matrix in each iteration, which generally requires high computation costs [38,39]. This also could lead to a limitation for handling large scale data that often occurs in the study of computer vision.

In order to address these problems, a novel Schatten-q quasi-norm induced matrix decomposition model (SQNMD) in this paper is proposed. First, motivated by the bilinear factorization with two factor matrix norm regularizers used in low-level vision processing [38,39], a tractable low-rank regularizer is used to characterize the background part. In fact, this regularization is the non-convex Schatten-q norm for $0 < q < 1$ that can take advantage of the heavy-tailed distribution of singular values in matrices. Compared with the nuclear norm, it is better in capturing the rank structure and provides more adequate low rank approximation, as shown in our experiments later on. In other words, this regularization is more suitable to describe the background in a corresponding subspace associated with lower rank properties. Thus the adoption of non-convex Schatten-q norm usually leads to a cleaner background than convex setting, illustrated in our experiments. Second, by taking into account image geometrical structure and feature similarity between image patches, we use a weighted group sparsity norm to characterize the foreground object. This setting allows our process to share the consistency within the same image patches. Moreover, in order to distinguish thoroughly between saliency target and background, we continue to adopt the Laplacian term similar to [33], and it plays a complementary role when there is a similar appearance between the salient objects and the background, which can guarantee the object completeness as much as possible. By integrating aforementioned techniques together, our proposed model can separate the salient objects from complex scenes much more effectively, compared with current available approaches.

In particular, we pay attention to two specific cases of q, 1/2 and 2/3 in this paper, which are in essence bi-nuclear quasi-norm and Frobenius/nuclear hybrid norm respectively. In our each individual algorithm, we only need to perform SVDs on two much smaller factor matrices as contrary to the larger ones used in existing low rank methods. Therefore, the proposed model is also well suitable for big data processing. To the best of our knowledge, this is the first study which pursues Schatten quasi-norm solvers in salient object detection (the latest and the most representative studies based on Schatten quasi-norm are mainly to consider the low-level vision problems, e.g., see [38,40–45]). In addition, we evaluate the SQNMD method on six well-known benchmark datasets that involve various complex scenarios, and compare our

method with 12 state-of-the-art methods. Through some standard evaluation index, our proposed SQNMD method in this paper generates more competitive results than current existing methods, as shown in our experiment section.

The remainder of this paper is organized as follows: Section 2 provides some related works about salient object detection problems. Section 3 provides the development of our proposed SQNMD model. Section 4 develops the corresponding optimization algorithm and computational complexity. Algorithm analysis is given in Section 5. The experiment results are shown in Section 6. Finally, the conclusion is given in Section 7.

2. Related works

During recent years, we have witnessed the significant advances in the study of salient object detection. In this section, we mainly discuss the most influential models which are closely related to low rank approximation.

LRMR [28] aims to recover a low-rank matrix L (i.e. the background without salient object) and a sparse matrix S (i.e. the foreground with salient object) from the given feature matrix of an input image $F = L + S$ as follows:

$$\min_{L,S} rank(L) + \lambda \Phi(S) \quad s.t. \quad F = L + S \qquad (1)$$

where $\Phi(S)$ is a sparsity regulation function and λ is positive trade-off parameters. Unfortunately, solving (1) is intractable and may be NP-hard in many cases. Thus, this motivates us to seek the corresponding relaxed approximation models.

LRR [31] assumes that there exists strong correlation among background patches, which leads to consider a low-rank coefficient matrix multiplied by the feature matrix to represent the background and utilize the sum of l_2 norm of the columns of matrix as the sparse regularization. That is,

$$\min_{Z_0,E_0} \|Z_0\|_* + \lambda \|E_0\|_{21} \, s.t. X = XZ_0 + E_0$$

where XZ_0 denotes the background, which can be reconstructed by itself, Z_0 denotes th reconstrucion coefficients, E_0 denotes the salient objects, $\|\cdot\|_*$ is the nuclear norm, i.e. the sum of the singular values which is a convex approximation to the rank function; $\|\cdot\|_{21}$ can make columns of E to be sparse. However, the LRR can only model a certain type of visual features, which can not be directly used for multifeature cases. To combine together multiple features, MSP [30] introduce the multi-task extension of LRR by incorporating multi-feature collaborative enhancement and top-down priors. Thus, MSP model can be formulated as the following form:

$$\min_{\substack{Z_1,\cdots,Z_K \\ E_1,\cdots,E_K}} \sum_{i=1}^{K} \|Z_i\|_* + \lambda \|E\|_{21} \quad s.t. X_i = X_i Z_1 + E_i \quad i = 1, \ldots, K \qquad (2)$$

But the sparse matrix E may not be accurate in MSP model, since it ignores the spatial relations and the feature affinities, that may cause inaccurate salient object detection.

To overcome this shortage, the ULR [29] model resorts to feature transformation and sparse representation, which integrates the low-level features and high-level semantic priors. Thus, the ULR can be written as:

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1 \quad s.t. \quad F = L + S \qquad (3)$$

However, this model usually produces non-uniformly highlighted salient object, because the spatial relations and feature similarities of patches are lacking. To address this problem, the SLR [32] exploits the bottom-up segmentation prior H_c from the connectivity between regions and borders as guidance clue of detection (i.e. adds the condition $F = AH_c$ in (3)), which can yield a small weight

to the regions approximating image borders (which leads to the approximated object being more accurate). Although the SLR obtains more effectively salient objects than the ULR, some important cases are ignored, such as when the background is clustered, or when the background has similar appearance with the salient object. So, the SMD [33] is proposed to further address the problem.

Comparing with the SLR model, the SMD method utilizes structured sparsity regularized term $\|\cdot\|_{2,\infty}$ instead of $\|\cdot\|_1$ to capture the potential structure of image and forces patches from the same object to have similar saliency values. Simultaneously, a Laplacian regularization under the assumption of the local invariance is also considered, which can preserves the local structure of the image feature and enlarge the difference between the background and salient object. However, in the SMD model, the background information is still characterized by the nuclear norm. As we mentioned before, the nuclear norm may lead to the over-penalization for large singular vues, that makes solution deviate from the original background. Thus different from current literature, in this paper, the non-convex Schatten-q quasi-norms for $0 < q < 1$ is considered for the salient object detection problem in order to achieve a better approximation to the rank function of background. Moreover, we focuses on Schatten-1/2 and 2/3, which appear to be suitable constraint on the background so that it can be compressed adequately.

Here what must be specifically stated is that 1/2-norm for background appeared in matrix decomposition model based on diversity [34] (DIMD). However, DIMD has to be solved iteratively, and involves singular value decomposition (SVD) in each iteration, which means that the corresponding algorithm has high per-iteration complexity. The proposed Schatten-q quasi-norm induced matrix decomposition (SQNMD) utilizes two tractable bilinear factor matrix norms (i.e. bi-nuclear quasi-norm and Frobenius/nuclear hybrid norm) for background, which only requires SVDs on two much smaller factor matrices as contrary to the larger ones used in DIMD [34] (one can refer to computational complexity analysis of Section 5, such as Remarks 1–3). Thus, the new SQNMD model is more general, tractable and scalable, which subsumes the DIMD method as a special case to some extent. The new SQNMD model not only inherits the main advantages of the DIMD method, i.e. it decomposes a given matrix into structured parts with diversity, but it is also armed with the new ability to remedy the gap between the rank function and its convex envelop. The experimental results (Section 6) shows clearly that the resulting saliency maps of new method are more visually favorable.

3. Schatten-q quasi-norm induced decomposition model for salient object detection

3.1. Problem formulation

Inspired by Peng et al. [33], a given nature image I is over-segmented into N non-overlapping patches $\mathcal{P} = \{P_1, P_2, \ldots P_N\}$. For each patch P_i, i-th D-dimensional low-level feature vector can be denoted as $f_i \in \mathbb{R}^D$. Consequently, a feature matrix of I is formed, which can be written as $F = [f_1, f_2, \ldots, f_N] \in \mathbb{R}^{D \times N}$. The problem of salient object detection is to find an efficient model to decompose the feature matrix F into a low-rank part L (background) and a sparse part S (salient object), respectively.

In order to address the salient object detection problem, we formulate this problem into the following more general form:

$$\min_{L,S} f(L) + \alpha g(S) + \beta h(L, S) \quad s.t. \quad F = L + S \qquad (4)$$

where $f(L)$ is a low-rank constraint for background L, $g(S)$ is a sparse regularized term, $h(L, S)$ is an interactive regularization term

Table 1
The norms of low-rank: $\|X\|_{S_q} (q \in [0, 2])$.

q	$\|X\|_{S_q}$	
0	$\|X\|_{S_0} = \|\sigma\|_{l_0}$	Rank
(0, 1)	$\|X\|_{S_q} = \left(\sum \sigma_k^q\right)^{\frac{1}{q}}$	Schatten-q norm
1	$\|X\|_* = \sum \sigma_k$	Nuclear norm
2	$\|X\|_F = \sqrt{\sum \sigma_k^2}$	Frobenius norm

between L and S, and α, β are positive trandeoff parameters used in the optimization.

3.2. Schatten-q quasi-norm term for background

The extended Schatten-q quasi-norm of a matrix $X \in \mathbb{R}^{m \times n}$ is defined as:

$$\|X\|_{Sq} = \left(\sum_{i=1}^{\min\{m,n\}} \sigma_i^q(X) \right)^{1/q} \tag{5}$$

where $\sigma_i(X)$ is the $i-th$ singular value of X. For $q \in [0, 2]$, the regularized term $\|X\|_{S_q}$ can be summarized in Table 1.

Since the non-salient background in a image is similar, it thus always lies in low dimensional subspace. Therefore, there is the consensus that one only need to minimize the rank of this subspace for background. In order to approximate more effectively the rank function, we consider the second choice in Table 1, i.e. Shatten-q quasi-norm with $0 < q < 1$. Throughout this paper, without loss (of) generality, we mainly focus on two specific values of q, 1/2 and 2/3 for the purpose of implementations. According to the theory of unified surrogate for Schatten-p norm [38], Schatten-1/2 and 2/3 are essentially bi-nuclear quasi-norm and Frobenius/Nuclear hybrid norm, whose characterizations are given below:

Theorem 1 (Frobenius/Nuclear and Bi-Nuclear Norm Surrogate [38]). *Given matrices $U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}$, and $X \in \mathbb{R}^{m \times n}$ with $rank(X) = r \leq d$, the flowing hold:*

$$\|X\|_{S_{2/3}}^{2/3} = \min_{U,V:X=UV^T} \frac{2}{3}\|U\|_* + \frac{1}{3}\|V\|_F^2$$

$$\|X\|_{S_{1/2}}^{1/2} = \min_{U,V:X=UV^T} \frac{1}{2}\|U\|_* + \frac{1}{2}\|V\|_* \tag{6}$$

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

In addition, we will also consider $q = 1$. Although it is nuclear norm and convex (i.e. it reduces to SMD [33]), we propose an alternative non-convex bilinear spectral penalty for the nuclear norm described in Theorem 2.

Theorem 2 [42]. *Given matrices $U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}$, and $X \in \mathbb{R}^{m \times n}$ with $rank(X) = r \leq d$, the flowing holds:*

$$\|X\|_* = \min_{U,V,X=UV^T} \frac{\|U\|_F^2}{2} + \frac{\|V\|_F^2}{2}. \tag{7}$$

Due to the requirement of factorization of matrix norm during the approach, regardless of whether the value of p is 1, 1/2, or 2/3, we develop an approach in which it is only required to perform SVDs on two much smaller factor matrices for solving minimization as contrary to the larger ones used in SMD. This is particularly useful for many big datasets. To the best of our knowledge, the use of non-convex Schatten quasi-norm solvers for the salient object detection has not been seen yet, and the latest and most representative works based on Schatten quasi-norm are all considered for the low-level vision problems, such as matrix completion and image inpainting [38,40–45].

3.3. Weighted group sparsity regularization for salient region and Laplacian constraint

It is well-known that a valid segmentation result has to bear some of the potential information of the image. Inspired by Peng et al. [33], Liu and Ye [46], Jia et al. [47], we continue to impose a weighted group sparsity induced norm to the salient part S, which can be written in the following form:

$$\wedge(S) = \sum_{j=1}^{n} v_j \|S_{G_j}\|_p, \tag{8}$$

where G_j is the j-th node of graph cut, $v_j > 0$ is a prior weight for the node G_j, S_{G_j} is a sub-matrix of S, $S_{G_j} \in \mathbb{R}^{D \times |G_j|}$ ($|\cdot|$ is the cardinality of a set), n is the number of nodes, $\|\cdot\|_p$ is the l_p-norm, which is used to characterize relationships among the corresponding patches within the same group. In general, the p value is set to be $1 \leq p \leq \infty$. In order to produce more precise and structurally consistent results, we focus on $p = \infty$ which can take the advantage of the spatial contiguity and feature similarity among image patches. In fact, l_∞ norm is the maximum saliency value of patches within the group that determines if the group belongs to saliency or not. More specifically, (8) is a weight group sparsity norm over a graph, where

$$v_j = 1 - \max\left(\{\pi_k : k \in G_j\}\right). \tag{9}$$

π_i in (9) indicates the likelihood that patch P_i belongs to a salient object based on high-level information [29]. This weight group sparsity norm, on one hand, can forces the patches within a same group to have similar salient values, and on the other hand, it can emphasize patches from different groups to have differentiable representations, as shown in Fig. 2.

In addition, to enlarge the distance between salient object and background in image patch, we also append the Laplacian regularization $Tr(SQ_F S^T)$ [33,48] based on the local invariance assumption for salient object S ($Tr(\cdot)$ is the trace of matrix). Here the entry of a Laplacian matrix Q_F can be given as the following form:

$$(Q_F)_{i,j} = \begin{cases} -w_{ij}, & \text{if } i \neq j \\ \sum_{j \neq i} w_{i,j}, & \text{otherwise}. \end{cases} \tag{10}$$

$w_{i,j}$ is the entry of an affinity matrix which represents the feature similarity of patches (P_i, P_j), i.e.

$$w_{i,j} = \begin{cases} \exp\left(-\frac{\|f_i - f_j\|^2}{2\delta^2}\right) & \text{if } (P_i, P_j) \in V \\ 0 & \text{otherwise}. \end{cases} \tag{11}$$

where V denotes the set of adjacent patch.

In summary, the problem (4) can be specifically constructed as (which is simply referred to as SQNMD model in short notation)

$$\min_{L,S} \|L\|_{S_q}^q + \alpha \sum_{j=1}^{n} v_j \|S_{G_j}\|_\infty + \beta Tr(SQ_F S^T) \quad s.t. \quad F = L + S \tag{12}$$

According to the Theorems listed in the part 3.2, the model (12) can be rewritten as the following three forms:
- $q = 1$,

$$\min_{U,V,S,H} \frac{1}{2}(\|U\|_F^2 + \|V\|_F^2) + \alpha \sum_{j=1}^{n} v_j \|S_{G_j}\|_\infty + \beta Tr(HQ_F H^T)$$

$$s.t. \quad F = UV^T + S, \ S = H. \tag{13}$$

- $q = 2/3$,

$$\min_{U,V,S,M,H} \frac{1}{3}(2\|M\|_* + \|V\|_F^2) + \alpha \sum_{j=1}^{n} v_j \|S_{G_j}\|_\infty + \beta Tr(HQ_F H^T)$$

| Image | G_4 | G_3 | G_2 | G_1 |
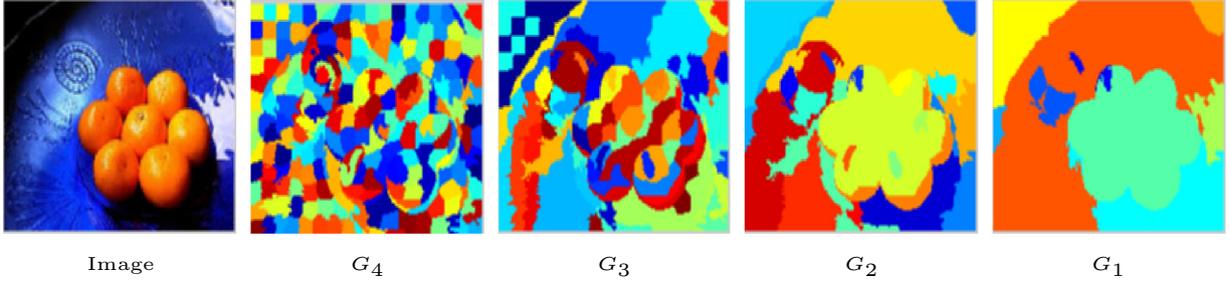
Fig. 2. Results from weighted group sparse acting.

$$s.t.\ F = UV^T + S,\ M = U,\ S = H. \tag{14}$$

- $q = 1/2$,

$$\min_{U,V,S,M,N,H} \frac{1}{2}(||M||_* + ||N||_*) + \alpha \sum_{j=1}^{n} v_j ||S_{G_j}||_\infty + \beta Tr(HQ_F H^T)$$

$$s.t.\ F = UV^T + S,\ M = U,\ N = V,\ S = H. \tag{15}$$

It is easy to see that the model in $q = 1$ is equivalent to SMD [33]. For $q = 1/2$, the proposed model degenerates as DIMD [34]. But the difference is that the model (13) is based on an alternative non-convex formulation. On the other hand, the ablation study of each regularization term, not shown here, also demonstrates the effectiveness from our proposed approach.

4. Optimization

Motivated by the alternating direction method of multipliers (ADMM) [49,50], we provide those related optimization algorithms for the aforementioned problems.

4.1. Solving (13)

The augmented Lagrangian function for problem (13) is given by

$$\mathcal{L}_1(U, V, S, H, Y_1, Y_2, \mu) = \frac{1}{2}(||U||_F^2 + ||V||_F^2) + \alpha \sum_{j=1}^{n} v_j ||S_{G_j}||_\infty$$

$$+ \beta Tr(HQ_F H^T) + \langle Y_1, F - UV^T - S \rangle + \langle Y_2, S - H \rangle$$

$$+ \frac{\mu}{2}(||F - UV^T - S||_F^2 + ||S - H||_F^2) \tag{16}$$

where Y_1, Y_2 are the Lagrange multipliers, and $\mu > 0$ is a penalty parameter. The solution of (16) is equivalent to minimizing the augment Lagrange function $\mathcal{L}_1$. The complete algorithm is shown in Algorithm 2. Next, the detailed updating processes of variables are shown in the following subsections.

Updating U. First, fixing V, S and H and seeking U to minimize the function $\mathcal{L}_1$. Thus, we consider the following optimization problem:

$$U^* = \arg\min_{U} \frac{1}{2}||U||_F^2 + \frac{\mu}{2}\left\|F - UV^T - S + \frac{Y_1}{\mu}\right\|_F^2. \tag{17}$$

Obviously, (17) is a least squares problem, thus the optimal solution is given by

$$U^* = (\mu FV + Y_1 V - \mu SV)(I + \mu V^T V)^{-1}. \tag{18}$$

Updating V. Fixing U, S and H, V can be obtained by solving the following problem:

$$V^* = \arg\min_{V} \frac{1}{2}||V||_F^2 + \frac{\mu}{2}\left\|F - UV^T - S + \frac{Y_1}{\mu}\right\|_F^2. \tag{19}$$

Similarly, we have

$$V^* = (\mu F^T U + Y_1^T U - \mu S^T U)(I + \mu U^T U)^{-1}. \tag{20}$$

Updating S. To update S, we arrive at the problem below:

$$S^* = \arg\min_{S} \frac{\alpha}{2\mu} \sum_{j=1}^{n} ||S_{G_j}||_\infty + \frac{1}{2}\left\|S - \frac{F - UV^T + H + \frac{Y_1 - Y_2}{\mu}}{2}\right\|_F^2. \tag{21}$$

To solving (21), we simplify this problem as:

$$S^* = \min_{S} \gamma \sum_{j=1}^{n} v_j \left\|S_{G_j}\right\|_\infty + \frac{1}{2}\|S - A\|_F^2. \tag{22}$$

Similar to the approach given in [51], we provide an optimization algorithm based on the hierarchical proximal operator, which is summarized in Algorithm 1.

Algorithm 1 Solving S by hierarchical proximal operator.

Input: F and γ
1: Set $S = A$
2: **for** $i = d$ **to** 1 **do**
3:     **for** $j = 1$ **to** n **do**
4:         $S_{G_j}^{k+1} = \begin{cases} \frac{\left\|S_{G_j}\right\|_1 - \gamma v_j}{\left\|S_{G_j}\right\|} S_{G_j} & if\ \left\|S_{G_j}\right\| > \gamma v_j \\ 0, & else \end{cases}$
5:     **end for**
6: **end for**
Output: S^k

Updating H. Finally, fixing U, V, and S, updating H is obtained by solving the following problem:

$$H^* = \arg\min_{H} \beta Tr(HQ_F H^T) + \frac{\mu}{2}\left\|S - H + \frac{Y_2}{\mu}\right\|_F^2. \tag{23}$$

then we have

$$H^* = (\mu S + Y_2)(2\beta Q_F^T + \mu I)^{-1}. \tag{24}$$

In summary, the algorithm of (16) based on ADMM can be outlined as Algorithm 2. And we also give the per-iteration complexity analysis for Algorithm 2, that is Remark 1.

Remark 1. It is well known that the computation complexity of thin SVD for an $m \times n$ matrix with $m \geqslant n$ is $O(mn^2)$. The cost of computing the inverse for $d \times d$ matrix is $O(d^3)$, and the expense of multiplication for $m \times d$ matrix and $d \times n$ matrix is $O(mdn)$. In ULR [29], DIMD [34] and SMD [33], the cost is dominated by the computation of the thin SVD of an $m \times n$ matrix with $m \geqslant n$, and is $O(mn^2)$ respectively. But for Algorithm 2, the dominant cost of each iteration for updating $U \in \mathbb{R}^{m \times d}$ and $V \in \mathbb{R}^{n \times d}$ using Eqs. (18) and (20) is $O(6mnd + 2d^3 + md^2 + nd^2)$. Therefore, we deduce that $O(6mnd + 2d^3 + md^2 + nd^2) \ll O(mn^2)$ for m, $n \gg d$.

Algorithm 2 Problem (16) solved by ADMM.

Input: F, d, α, β and $\mu_0, S_0 = H_0 = Y_1^0 = Y_2^0 = 0, \mu_{\max} = 10^{10}, \rho = 1.1$ and $k = 0$
1: **while** not converged **do**
2:     update U_{k+1} by (18)
3:     update V_{k+1} by (20)
4:     update S_{k+1} by (22)
5:     update H_{k+1} by (24)
6:     update Y_1^{k+1} by $Y_1 \leftarrow Y_1 + \mu\left(F - UV^T - S\right)$
7:     update Y_2^{k+1} by $Y_2 \leftarrow Y_2 + \mu(S - H)$
8:     update μ_{k+1} by $\mu \leftarrow \min(\rho\mu, \mu_{\max})$
9:     update k by $k \leftarrow k + 1$
10: **end while**
Output: U, V and S

4.2. The procedure for solving (14)

The augmented Lagrangian function for problem (14) is represented as

$$\mathcal{L}_{2/3}(U, V, S, M, H, Y_1, Y_2, Y_3, \mu) = \frac{1}{3}(2||M||_* + ||V||_F^2) + \alpha\sum_{j=1}^n v_j||S_{G_j}||_\infty$$

$$+ \beta Tr(HQ_F H^T) + \langle Y_1, F - UV^T - S\rangle + \langle Y_2, M - U\rangle + \langle Y_3, S - H\rangle$$

$$+ \frac{\mu}{2}(||F - UV^T - S||_F^2 + ||M - U||_F^2 + ||S - H||_F^2), \quad (25)$$

where Y_1, Y_2, Y_3 are the Lagrange multipliers, and $\mu > 0$ is a penalty parameter. The detailed updating processes of variables are shown as follows.

- **Updating U**: First of all, fixing V, S, M, and H, updating U is obtained by solving the following problem:

$$U^* = \arg\min_U \frac{\mu}{2}\left(\left\|F - UV^T - S + \frac{Y_1}{\mu}\right\|_F^2 + \left\|M - U + \frac{Y_2}{\mu}\right\|_F^2\right). \quad (26)$$

Thus, we have

$$U^* = \left(M + \frac{Y_2}{\mu} + FV + \frac{Y_1}{\mu}V - SV\right)(I + V^T V)^{-1}. \quad (27)$$

- **Updating V**: Secondly, fixing U, S, M, and H, updating V is obtained by

$$V^* = \arg\min_V \frac{1}{3}||V||_F^2 + \frac{\mu}{2}\left\|F - UV^T - S + \frac{Y_1}{\mu}\right\|_F^2. \quad (28)$$

Then we have

$$V^* = \left(\mu F^T U - \mu S^T U + Y_1^T U\right)\left(\frac{2}{3}I + \mu U^T U\right)^{-1}. \quad (29)$$

- **Updating S**: Third, fixing U, V, M, and H, updating S can be represented as

$$S^* = \arg\min_S \frac{\alpha}{2\mu}\sum_{i=1}^d\sum_{j=1}^{n_i}||S_{G_j^i}||_\infty + \frac{1}{2}\left\|S - \frac{F - UV^T + H + \frac{Y_1 - Y_3}{\mu}}{2}\right\|_F^2. \quad (30)$$

Therefore, (30) can be solved according to Algorithm 1.

- **Updating M**: Fourth, fixing U, V, S, and H, M is obtained by

$$M^* = \arg\min_M \frac{2}{3\mu}||M||_* + \frac{1}{2}\left\|U - \frac{Y_2}{\mu} - M\right\|_F^2. \quad (31)$$

Without loss of generality, the Eq. (31) can be simplified as

$$\min_B \tau||B||_* + \frac{1}{2}||C - B||_F^2. \quad (32)$$

Thus, the problem (32) is a regularized least squares problem associated with the nuclear norm, of which closed-form solution can be represented by the singular value thresholding (SVT) [52] (refers Theorem 3).

Theorem 3 (SVT [52]). *Let the singular value decomposition of C can be written as $C = U\Sigma V^T$, then the optimal solution for B is given by*

$$B = \mathcal{D}_\tau(C) = U\mathcal{S}_\tau(\Sigma)V^T, \quad (33)$$

where $\mathcal{D}_\varepsilon$ is the singular value thresholding operator and $\mathcal{D}_\varepsilon$ is the soft thresholding operator, which is defined as

$$\mathcal{S}_\varepsilon(x) = sign(x)\max(|x| - \varepsilon, 0) = \begin{cases} x - \varepsilon & x > \varepsilon \\ x + \varepsilon & x < -\varepsilon \\ 0 & \text{else} \end{cases}. \quad (34)$$

- **Updating H**: Finally, fixing U, V, S and M, updating H is set by

$$H^* = \arg\min_H \beta Tr(HQ_F H^T) + \frac{\mu}{2}\left\|S - H + \frac{Y_2}{\mu}\right\|_F^2. \quad (35)$$

Then we have

$$H^* = (\mu S + Y_3)(2\beta Q_F^T + \mu I)^{-1}. \quad (36)$$

Therefore, when $q = \frac{2}{3}$, we obtain the optimization procedure as given by Algorithm 3. Similarly, the complexity analysis can be summarized as Remark 2.

Algorithm 3 Problem (25) solved by ADMM.

Input: F, d, α, β and μ_0, $M_0 = Y_2^0 = 0$, $S_0 = H_0 = Y_1^0 = Y_3^0 = 0, \mu_{\max} = 10^{10}, \rho = 1.1$ and $k = 0$
1: **while** not converged **do**
2:     update U_{k+1} by (27)
3:     update V_{k+1} by (29)
4:     update S_{k+1} by (30)
5:     update M_{k+1} by (32)
6:     update H_{k+1} by (36)
7:     update Y_1^{k+1} by $Y_1 \leftarrow Y_1 + \mu\left(F - UV^T - S\right)$
8:     update Y_2^{k+1} by $Y_2 \leftarrow Y_2 + \mu(M - U)$
9:     update Y_3^{k+1} by $Y_3 \leftarrow Y_3 + \mu(S - H)$
10:    update μ_{k+1} by $\mu \leftarrow \min(\rho\mu, \mu_{\max})$
11:    update k by $k \leftarrow k + 1$
12: **end while**
Output: U, V and S

Remark 2. In Algorithm 3, the per-iteration complexity of updating $U \in \mathbb{R}^{m\times d}$, $V \in \mathbb{R}^{n\times d}$ and $M \in \mathbb{R}^{m\times d}$ using Eqs. (27), (29) and (32) is $O(6mnd + 2d^3 + 2md^2 + nd^2)$. Thus, one has $O(6mnd + 2d^3 + 2md^2 + nd^2) \ll O(mn^2)$ for $m, n \gg d$, where $O(mn^2)$ is respectively the cost of ULR [29], DIMD [34] and SMD [33].

4.3. The procedure for solving (15)

Similar to the discussion of subsection B, the augmented Lagrangian function for problem (15) is

$$\mathcal{L}_{1/2}(U, V, S, M, N, H, Y_1, Y_2, Y_3, Y_4, \mu) = \frac{1}{2}(||M||_* + ||N||_*)$$

$$+ \alpha\sum_{j=1}^n v_j||S_{G_j}||_\infty + \beta Tr(HQ_F H^T) + \langle Y_1, F - UV^T - S\rangle$$

$$+ \langle Y_2, M - U\rangle + \langle Y_3, N - V\rangle + \langle Y_4, S - H\rangle$$

$$+ \frac{\mu}{2}(||F - UV^T - S||_F^2 + ||M - U||_F^2 + ||N - V||_F^2 + ||S - H||_F^2) \quad (37)$$

Algorithm 4 Problem (37) solved by ADMM.

Input: F, d, α, β and $\mu_0, M_0 = Y_2^0 = 0, N_0 = Y_3^0 = 0, S_0 = H_0 = Y_1^0 = Y_4^0 = 0, \mu_{\max} = 10^{10}, \rho = 1.1$ and $k = 0$

1: **while** not converged **do**
2:     update U_{k+1} by (38)
3:     update V_{k+1} by (39)
4:     update S_{k+1} by (40)
5:     update M_{k+1} by (41)
6:     update N_{k+1} by (42)
7:     update H_{k+1} by (43)
8:     Y_1^{k+1} by $Y_1 \leftarrow Y_1 + \mu\left(F - UV^T - S\right)$
9:     Y_2^{k+1} by $Y_2 \leftarrow Y_2 + \mu(M - U)$
10:    Y_3^{k+1} by $Y_3 \leftarrow Y_3 + \mu(N - V)$
11:    Y_4^{k+1} by $Y_4 \leftarrow Y_4 + \mu(S - H)$
12:    μ_{k+1} by $\mu \leftarrow \min(\rho\mu, \mu_{\max})$
13:    k by $k \leftarrow k + 1$
14: **end while**
Output: U, V and S

---

where Y_1, Y_2, Y_3, Y_4 are the Lagrange multipliers, and $\mu > 0$ is a penalty parameter. Thus, we here also provide an efficient ADMM algorithm for (37) (i.e. Algorithm 4). Since the update scheme of U, V, S, M, N and H are very similar to that of Algorithm 3, we only need to solve the following convex optimization problems for each step in the alternating iteration.

1. **Updating U** :

$$U^* = \left(FV + \frac{Y_1}{\mu}V + M + \frac{Y_2}{\mu} - SV\right)(V^T V + I)^{-1}, \quad (38)$$

2. **Updating V** :

$$V^* = \left(N + \frac{Y_3}{\mu} + F^T U + \frac{Y_1^T}{\mu}U - S^T U\right)(I + U^T U)^{-1}, \quad (39)$$

3. **Updating S** :

$$S^* = \arg\min_S \frac{\alpha}{2\mu}\sum_{j=1}^n ||S_{G_j}||_\infty + \frac{1}{2}\left\| S - \frac{F - UV^T + H + \frac{Y_1 - Y_4}{\mu}}{2} \right\|_F^2, \quad (40)$$

4. **Updating M** :

$$M^* = \arg\min_M \frac{1}{2\mu}||M||_* + \frac{1}{2}\left\|\left(U - \frac{Y_2}{\mu}\right) - M\right\|_F^2, \quad (41)$$

5. **Updating N** :

$$N^* = \arg\min_N \frac{1}{2\mu}||N||_* + \frac{1}{2}\left\|V - \frac{Y_3}{\mu} - N\right\|_F^2, \quad (42)$$

6. **Updating H** :

$$H^* = \arg\min_H \beta Tr(HQ_F H^T) + \frac{\mu}{2}\left\|S - H + \frac{Y_2}{\mu}\right\|_F^2. \quad (43)$$

Remark 3. In Algorithm 4, the dominant complexity of updating $U \in \mathbb{R}^{m \times d}$, $V \in \mathbb{R}^{n \times d}$, $M \in \mathbb{R}^{m \times d}$ and $N \in \mathbb{R}^{n \times d}$ using Eqs. (38), (39), (41) and (42) is $O\left(6mnd + 2d^3 + 2md^2 + 2nd^2\right)$. Therefore, $O\left(6mnd + 2d^3 + 2md^2 + 2nd^2\right) \ll O\left(mn^2\right)$ for $m, n \gg d$, where $O\left(mn^2\right)$ is respectively the cost of ULR [29], DIMD [34] and SMD [33].

5. Convergence analysis

According to the above Algorithms, each sub-problem has a closed-form solution in the proposed SQNMD model and the objective value is always decreasing with respect to the primal variables optimized in each sub-problem. Now we provide the convergence property for the most complex Algorithm 4 under mild conditions, and the similar results can be shared with the Algorithms 2 and 3.

Theorem 4. *Let* $\{U_k\}, \{V_k\}, \{S_k\}, \{M_k\}, \{N_k\}, \{H_k\}$ *be a sequence generated by Algorithm 4. Suppose that the sequence* $\{Y_i^k\}$ *($1 \leqslant i \leqslant 4$) are bounded, and* μ_k *is non-decreasing and* $\sum_{k=0}^\infty \frac{\mu_{k+1}}{\mu_k^2} < \infty$, *then the sequences* $\{U_k\}, \{V_k\}, \{S_k\}, \{M_k\}, \{N_k\}, \{H_k\}$ *are all bounded.*

Proof. Let $X_k \triangleq (U_k, V_k, S_k, M_k, N_k, M_k)$ and $Y^k \triangleq (Y_1^k, Y_2^k, Y_3^k, Y_4^k)$. By the iterative scheme of Algorithm 4, we have that

$$\min_X \mathcal{L}_{1/2}\left(X, Y^k, \mu_k\right) = \mathcal{L}_{1/2}\left(X_{k+1}, Y^k, \mu_k\right). \quad (44)$$

Thus, it yields

$$\mathcal{L}_{1/2}\left(X_{k+1}, Y^k, \mu_k\right) \leq \mathcal{L}_{1/2}\left(X_k, Y^k, \mu_k\right)$$
$$= \mathcal{L}_{1/2}\left(X_k, Y^{k-1}, \mu_{k-1}\right) + \frac{\mu_k + \mu_{k-1}}{2(\mu_{k-1})^2}\sum_{i=1}^4 \left\|Y_i^k - Y_i^{k-1}\right\|_F^2. \quad (45)$$

Due to $\mu_{k+1} = \rho\mu_{k-1}, \rho > 1$ and $\sum_{k=0}^\infty \frac{\mu_k}{\mu_{k+1}^2} < \infty$, and μ_k is non-decreasing, one can obtain the following result:

$$\frac{\mu_k + \mu_{k-1}}{2(\mu_{k-1})^2} \leqslant \frac{2\mu_k}{2(\mu_{k-1})^2} = \frac{\mu_k}{(\mu_{k-1})^2} \to 0, \quad (46)$$

which implies that $\mathcal{L}_{1/2}((X_{k+1}, Y^k, \mu_k))$ is bounded.

Therefore, the problem (37) can be rewritten as

$$\frac{1}{2}\left(\|M_k\|_* + \|N_k\|_*\right) + \alpha\sum_{j=1}^n v_j \left\|S_{G_j}^k\right\|_\infty + \beta Tr(HQ_F H^T)$$
$$= \mathcal{L}_{1/2}\left(X_k, Y^{k-1}, \mu_{k-1}\right) - \frac{3}{2\mu_{k-1}}\sum_{i=1}^4 \left(\left\|Y_i^k\right\|_F^2 - \left\|Y_i^{k-1}\right\|_F^2\right), \quad (47)$$

which indicates that $\{M_k\}, \{N_k\}$ and $\{S_k\}$ are bounded.

Simultaneously, since $U_k = M_k - \frac{Y_2^k - Y_2^{k-1}}{\mu_{k-1}}$, we can deduce that $\{U_k\}$ is bounded. Similarly, $\{V_k\}$ and and $\{H_k\}$ are also bounded. $\square$

Theorem 5. *Let* $\{U_k\}, \{V_k\}, \{S_k\}, \{M_k\}, \{N_k\}, \{H_k\}$ *be a sequence generated by Algorithm 4. Suppose that the sequence* $\{Y_i^k\}$ *($1 \leqslant i \leqslant 4$) are bounded, and* μ_k *is non-decreasing and* $\sum_{k=0}^\infty \frac{\mu_{k+1}}{\mu_k^2} < \infty$, *then* $\{U_k\}, \{V_k\}, \{S_k\}, \{M_k\}, \{N_k\}, \{H_k\}$ *are all Cauchy sequences.*

Proof. Since

$$M_{k+1} - U_{k+1} = \frac{Y_2^{k+1} - Y_2^k}{\mu_k}, \quad (48)$$

by using Theorem 4, we have

$$\sum_{k=0}^\infty \|M_{k+1} - U_{k+1}\|_F = \sum_{k=0}^\infty \frac{1}{\mu_k}\left\|Y_2^{k+1} - Y_2^k\right\|_F$$
$$\leq \sum_{k=0}^\infty \frac{\mu_{k+1}}{\mu_k^2}\left\|Y_2^{k+1} - Y_2^k\right\|_F < \infty, \quad (49)$$

which yields

$$\lim_{k\to\infty} \|M_{k+1} - U_{k+1}\|_F = 0. \quad (50)$$

On the other hand, Lagrange parameter Y_1^k can be rewritten as the following form:

$$Y_1^k = Y_1^{k-1} + \mu_{k-1}\left(F - U_k V_k^T - S_k\right) \Rightarrow F - S_k = U_k V_k^T + \frac{Y_1^k - Y_1^{k-1}}{\mu_{k-1}}. \quad (51)$$

And the minimization of sub-problem (38) can be converted to seek the corresponding first order optimal condition regarding to U, that is,

$$\left(F - U_{k+1}V_k^T - S_k + \frac{Y_1^k}{\mu_k}\right)V_k + \left(M_k - U_{k+1} + \frac{Y_2^k}{\mu_k}\right) = 0. \tag{52}$$

Therefore, Eq. (52) can be rewritten as the following form:

$$(U_k - U_{k+1})\left(V_k^T V_k + I\right) + Z = 0. \tag{53}$$

where $Z = (\frac{Y_1^{k-1} - Y_1^k}{\mu_{k-1}} - \frac{Y_1^k}{\mu_k})V_k + \frac{Y_2^{k-1} - Y_2^k}{\mu_{k-1}} - \frac{Y_2^k}{\mu_k}$, the above Eq. (54) can be simplified as

$$U_k - U_{k+1} = Z\left(V_k^T V_k + I\right)^{-1}. \tag{54}$$

Accordingly, one has the following inequalities

$$\sum_{k=0}^{\infty} \|U_{k+1} - U_k\|_F \leqslant \sum_{k=0}^{\infty} \left(\frac{\left\|Y_1^{k-1} - Y_1^k\right\|_F \|V_k\|_F}{\mu_{k-1}} + \frac{\left\|Y_1^k\right\|_F \|V_k\|_F}{\mu_k}\right.$$
$$+ \left.\frac{\left\|Y_2^{k-1} - Y_2^k\right\|_F}{\mu_{k-1}} + \frac{\left\|Y_2^k\right\|_F}{\mu_k}\right)\left\|\left(V_k^T V_k + I\right)^{-1}\right\|_F$$
$$\leqslant \sum_{k=0}^{\infty} \left(\frac{\left\|Y_1^{k-1} - Y_1^k\right\|_F \|V_k\|_F}{\mu_{k-1}} + \frac{\left\|Y_1^k\right\|_F \|V_k\|_F}{\mu_{k-1}} + \frac{\left\|Y_2^{k-1} - Y_2^k\right\|_F}{\mu_{k-1}}\right.$$
$$+ \left.\frac{\left\|Y_2^k\right\|_F}{\mu_{k-1}}\right)\left\|\left(V_k^T V_k + I\right)^{-1}\right\|_F. \tag{55}$$

Set $\theta = \max\{(\|Y_1^{k-1} - Y_1^k\|_F \|V_k\|_F + Z_1)Z_2\}_{k=1}^{\infty}$, where $Z_1 = \|Y_1^k\|_F \|V_k\|_F + \|Y_2^{k-1} - Y_2^k\|_F \|Y_2^k\|_F$ and $Z_2 \triangleq \|(V_k^T V_k + I)^{-1}\|_F$, we get

$$\sum_{k=0}^{\infty} \|U_k - U_{k+1}\|_F \leqslant \sum_{k=0}^{\infty} \frac{1}{\mu_k}\theta \leqslant \sum_{k=0}^{\infty} \frac{\mu_{k+1}}{\mu_k^2}\theta < \infty, \tag{56}$$

which shows that $\{U_k\}$ is a Cauchy sequence.

Similar proofs lead to that $\{V_k\}$, $\{S_k\}$, $\{M_k\}$, $\{N_k\}$, $\{H_k\}$ are all Cauchy sequences. □

In fact, the proof of the Theorem 5 shows that under mild conditions, any limit point of the sequences generated by Algorithm 4 is a critical point of (37) due to the point whose sub-differential contains 0 only.

6. Experiments

6.1. Datasets,salient object detection algorithms and parameter settings

We use six standard benchmark datasets to represent various typical scenarios, i.e. ECSSD[16], MSRA10K [19], THUR15K [53], DUT-OMRON [25], iCoSeg [54] and PASCAL-S [55]. ECSSD dataset contains total 1000 images with various structurally complex objects, and MSRA10K dataset includes 10,000 images with single object. THUR15K and DUT-OMRON datasets involve a large scale single object with relative complex background, while iCoSeg and PASCAL-S datasets are related to multiple objects and different complex backgrounds.

The proposed method is compared with twelve recent state-of-the-art solutions, including four LR based methods which are DIMD [34], SMD [33], ULR [29] and SLR [32], respectively and eight recently developed prominent methods which are GS [15], HS [16], PCA [17], TD [11], DRFI [23], RBD [26], MR [25] and MC [56], respectively. It should be noted that these eight methods are not related to bilinear factorization we discussed. The main reason why we compare with these methods is to further demonstrate the generality of new model and the effectiveness of new algorithm.

Here we use initials to represent each individual method for the purpose of easy description throughout this section.

We consider a four-layer depth for group sparsity and set the bandwidth $\delta^2 = 0.05$, and the parameter $d = 25$, $\alpha = 0.3, 0.04, 0.06$, $\beta = 0.925, 0.6, 0.1125$ for $q = 1, \frac{2}{3}, \frac{1}{2}$ respectively. To get a fair comparison with other competing methods, we fix the parameters of our model for all conducted experiments.

6.2. Evaluation metrics

For the evaluation of experimental performances, we introduce several common metrics including the $F-$measure curve, the area under the ROC curve (AUC) [57], overlapping ratio (OR), and the mean absolute error (MAE), respectively.

It is well-known that precision is defined as the ratio of salient pixels correctly assigned, while recall is the percentage of correctly detected salient pixels to those all true salient pixels. Simultaneously, the $F-$measure curve is based on the weighted harmonic mean of precision (P) and recall (R), which is given by

$$F_\beta = \left(1 + \beta^2\right)\frac{P \cdot R}{\beta^2 P + R}, \tag{57}$$

where β^2 is set to be 0.3 similar to [9]. Whenever either the PR curve or the $F-$ measure curve is applicable, they rely on the variation of the saliency threshold that determines the attribution of salient objects. In addition, the receiver operating characteristic (ROC) curve is generated from true positive rates as well as false positive rates obtained by calculating the PR curve. The overlapping ratio (OR) is defined as the ratio between the segmented object mask S' and the ground truth, i.e.,

$$OR = \frac{|S' \cap G|}{|S' \cup G|}, \tag{58}$$

The mean absolute error (MAE) [10] characterizes the mean absolute difference between the saliency map S and the ground true G by the following:

$$MAE = mean(|S - G|). \tag{59}$$

For the weighted F-measure(WF) metric, we adopt definition introduced in [58].

6.3. Comparision with the state-of-the-arts approaches

We first compare our models with ULR [29], SLR [32], DIMD [34] and SMD [33]. Tables 2 and 3 show evaluation metric results based on six datasets. Note that the proposed models always can achieve the best performances on ECSSD, PASCAL-S and iCoSeg. For THUR15K, the proposed models obtain the best results in terms of WF, OR, AUC, and the second best in MAE. On MSRA10K, the proposed models give the best in OR and AUC, and the second best in WF and MAE. Finally, for DUT-OMRON, the proposed models indicate the second best results in terms of WF, OR and AUC. Thus, these metrics demonstrate that the proposed models (SQNMD) have potentiality in performing significantly better than the other LR-based methods.

Tables 4 and 5 report the comparison of new model and the other non-low-rank methods (i.e. PCA [17], GS [15], HS [16], MC [56], MR [25], DSR [24], RBD [26], DRFI [23]) in evaluation metrics. Obviously, the proposed model gives the best performance on ECSSD dataset. For PASCAL-S, iCoSeg and MSRA10K, the new method provides the same results, which are the best in terms of WF, OR and MAE, the second best in AUC. On THUR15K, the proposed model presents the best in WF, the second best in OR and the third best in AUC. For DUT-OMRON, the proposed method only obtains the second best in OR and the third best in WF. But lots of good evaluation indexes confirm that the proposed model

Table 2

Comparison with the Other Low-rank methods and performance boost with different baselines on Datasets.

Dataset	ECSSD				THUR15K				PASCAL-S			
Metric	WF ↑	OR ↑	AUC ↑	MAE ↓	WF ↑	OR ↑	AUC ↑	MAE ↓	WF ↑	OR ↑	AUC ↑	MAE ↓
ULR [29]	0.351	0.347	0.755	0.312	0.259	0.325	0.801	0.249	0.351	0.295	0.718	0.320
SLR [32]	0.442	0.474	0.764	0.252	0.387	0.433	0.823	0.167	0.398	0.390	0.711	0.275
DIMD [34]	0.534	0.548	0.819	0.181	0.430	0.458	0.829	0.166	0.500	0.451	0.738	0.245
SMD [33]	0.517	0.523	0.775	0.227	0.434	**0.466**	0.823	**0.159**	0.486	0.446	0.730	0.246
Ours $q=1$	**0.542**	**0.563**	0.813	0.174	**0.436**	**0.466**	0.824	0.160	0.494	0.447	0.733	0.245
Ours $q=2/3$	**0.542**	0.562	0.813	**0.173**	0.435	0.463	0.825	0.163	0.500	0.449	0.734	**0.244**
Ours $q=1/2$	0.540	0.554	**0.820**	0.179	0.432	0.459	**0.831**	0.167	**0.503**	**0.453**	**0.740**	0.245

The ↑ indicates the larger value achieved, the better performance is, while ↓ indicates the smaller, the better.

Table 3

Comparison with the Other Low-rank methods and performance boost with different baselines on Datasets.

Dataset	iCoSeg				MSRA10K				DUT-OMRON			
Metric	WF ↑	OR ↑	AUC ↑	MAE ↓	WF ↑	OR ↑	AUC ↑	MAE ↓	WF ↑	OR ↑	AUC ↑	MAE ↓
ULR [29]	0.379	0.443	0.814	0.222	0.425	0.524	0.831	0.141	0.254	0.318	0.805	0.260
SLR [32]	0.473	0.505	0.805	0.179	0.601	0.691	0.840	0.224	0.392	0.429	**0.822**	**0.161**
DIMD [34]	0.603	0.594	0.827	0.141	**0.761**	0.733	0.836	**0.087**	0.398	0.415	0.807	0.183
SMD [33]	0.611	0.598	0.822	0.138	0.704	0.741	0.847	0.104	**0.424**	**0.441**	0.809	0.166
Ours $q=1$	0.613	0.597	0.824	0.138	0.701	0.741	0.848	0.105	0.422	0.440	0.812	0.170
Ours $q=2/3$	**0.626**	**0.608**	0.829	**0.133**	0.705	**0.744**	0.848	0.103	0.423	0.440	0.812	0.171
Ours $q=1/2$	0.612	0.600	**0.833**	0.139	0.695	0.734	**0.852**	0.109	0.405	0.424	0.815	0.183

The ↑ indicates the larger value achieved, the better performance is, while ↓ indicates the smaller, the better.

Table 4

Comparison with Non-Low-rank method and performance boost with different baselines on Datasets.

Dataset	ECSSD				THUR15K				PASCAL-S			
Metric	WF ↑	OR ↑	AUC ↑	MAE ↓	WF ↑	OR ↑	AUC ↑	MAE ↓	WF ↑	OR ↑	AUC ↑	MAE ↓
PCA [17]	0.358	0.371	0.759	0.291	0.298	0.362	0.822	0.198	0.353	0.352	0.719	0.296
GS [15]	0.436	0.435	0.758	0.255	0.370	0.387	0.814	0.176	**0.456**	0.418	**0.734**	0.262
HS [16]	0.449	0.432	0.766	0.269	0.365	0.402	0.801	0.218	*0.451*	0.349	*0.733*	0.286
MC [56]	0.441	*0.495*	*0.779*	0.251	0.349	*0.444*	0.834	0.184	0.423	0.412	0.740	0.272
MR [25]	0.480	0.491	0.761	0.235	0.378	0.426	0.796	0.178	0.446	0.431	0.722	0.265
DSR [24]	0.489	0.480	0.754	0.227	*0.423*	0.426	0.803	*0.142*	0.439	0.409	0.712	*0.258*
RBD [26]	*0.490*	0.494	0.752	**0.225**	0.421	0.431	0.804	**0.150**	0.474	**0.442**	0.725	**0.247**
DRFI [23]	**0.517**	**0.527**	0.780	0.217	0.432	*0.481*	*0.856*	0.147	0.449	*0.432*	*0.749*	0.258
Ours	*0.542*	*0.562*	*0.813*	*0.173*	*0.435*	**0.463**	**0.825**	0.163	*0.500*	*0.449*	**0.734**	*0.244*

The ↑ indicates the larger value achieved, the better performance is, while ↓ indicates the smaller, the better. The best three results are highlighted with *italic*, **bold** and ***bold italic*** fonts, respectively, Ours $q=2/3$.

Table 5

Comparison with Non-Low-rank method and performance boost with different baselines on Datasets.

Dataset	iCoSeg				MSRA10K				DUT-OMRON			
Metric	WF ↑	OR ↑	AUC ↑	MAE ↓	WF ↑	OR ↑	AUC ↑	MAE ↓	WF ↑	OR ↑	AUC ↑	MAE ↓
PCA [17]	0.407	0.427	0.798	0.201	0.473	0.576	0.839	0.185	0.287	0.341	**0.827**	0.207
GS [15]	0.519	0.520	0.819	0.167	0.606	0.664	0.839	0.139	0.363	0.372	0.814	0.173
HS [16]	0.536	0.537	0.812	0.176	0.604	0.656	0.833	0.149	0.350	0.397	0.801	0.227
MC [56]	0.461	0.543	0.807	0.179	0.576	0.694	*0.843*	0.145	0.347	0.425	*0.820*	0.186
MR [25]	0.554	0.573	0.795	0.162	0.642	0.693	0.824	0.125	0.381	0.420	0.779	0.187
DSR [24]	0.548	0.514	0.801	0.153	0.656	0.654	0.825	0.121	0.419	0.408	0.803	**0.139**
RBD [26]	**0.599**	**0.588**	*0.827*	**0.138**	**0.685**	*0.716*	0.834	**0.108**	*0.427*	*0.432*	0.814	*0.144*
DRFI [23]	*0.592*	*0.582*	0.839	*0.139*	*0.666*	0.723	0.857	*0.114*	0.424	0.444	*0.839*	0.138
Ours	*0.626*	*0.608*	**0.829**	*0.133*	*0.705*	*0.744*	**0.848**	*0.103*	**0.423**	**0.440**	0.812	0.171

The ↑ indicates the larger value achieved, the better performance is, while ↓ indicates the smaller, the better. The best three results are highlighted with *italic*, **bold** and ***bold italic*** fonts, respectively, Ours $q=2/3$.

(SQNMD) is even more competitive than most of the non-low-rank methods.

Fig. 3 shows the F-measure curve comparison of the proposed model and ten competitive methods on six datasets. Note that the proposed model has significantly better performance than the other methods for ECSSD, iCoSeg, MSRA10k and PASCAL-S. For THUR15K, the new method, SMD [33] and DRFI [23] yield equal shares whose results are better than ones of the other methods. For DUT-OMRON, it is clear that our method can outperform the other methods except DRFI [23]. In fact, DRFI is a semi-supervised method with deep learning. Thus, inspired by this, we will consider combination of bilinear factorization norm and learning task.

(a) ECSSD (b) THUR15K (c) PASCAL-S

(d) iCoSeg (e) MSRA10K (f) DUT-OMRON
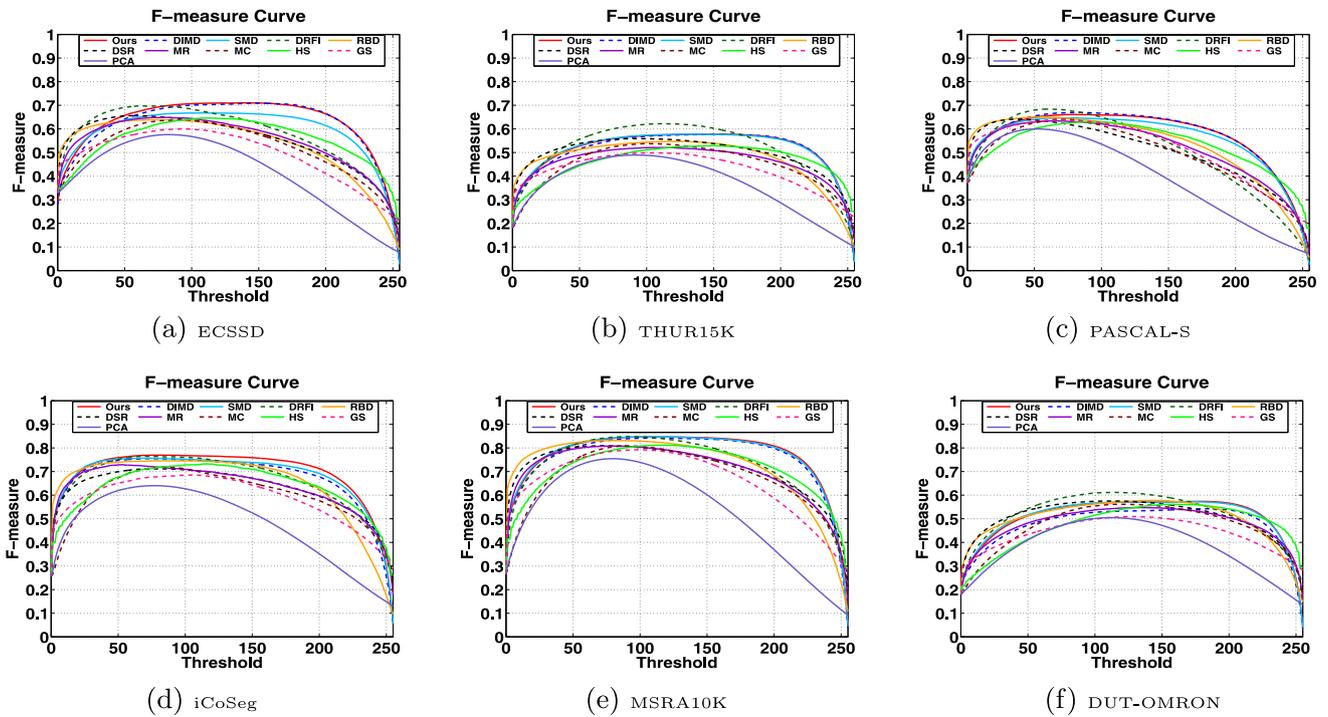
Fig. 3. Quantitative comparison on six datasets in terms of F-measure curve.

IMG   ULR   PCA   SLR   GS   HS   MC   MR   DSR   RBD   DRFI   SMD   DIMD   Ours   GT
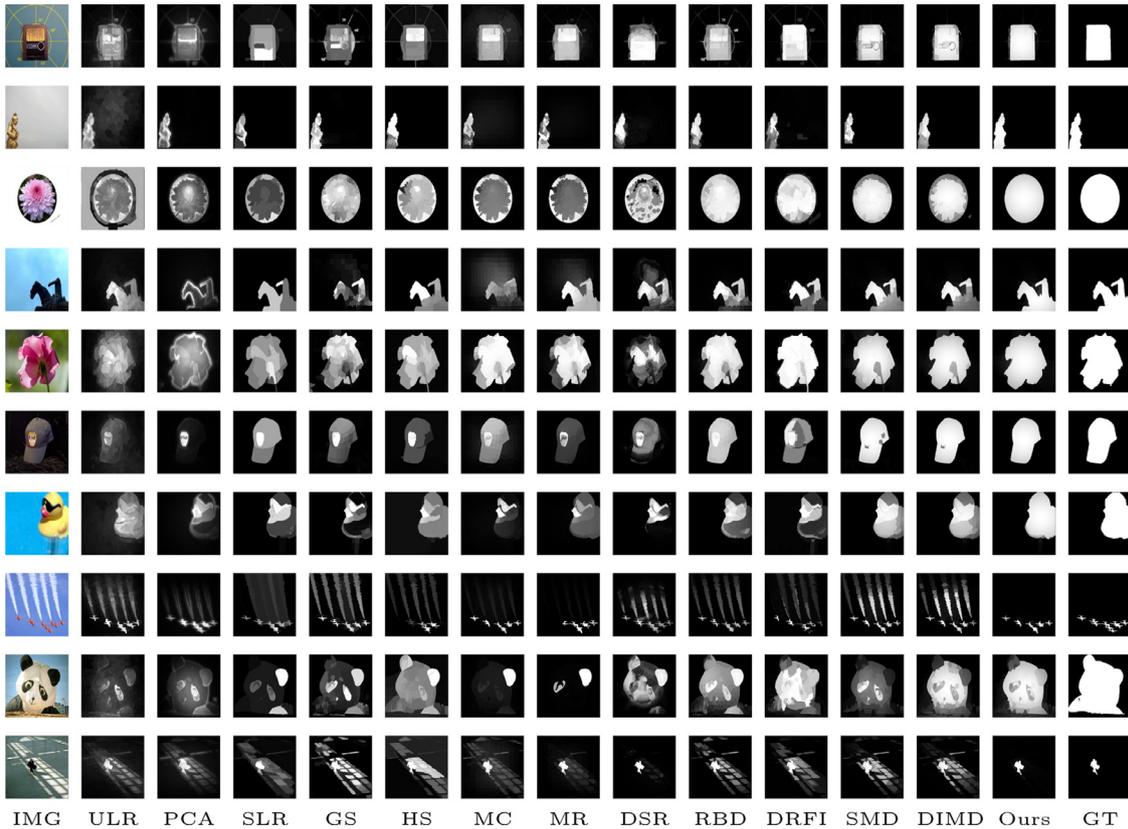
Fig. 4. Visual comparison of saliency maps, in which the results from our method appear to be very close to the ground truth (IMG and GT are short for image and ground truth).

Finally, Fig. 4 gives some quantitative visual comparisons based on the twelve state-of-the-art models. It is easy to find that the maps from our model not only possess the consistent salient values of pixels within the same salient objects, but also assign suc- cessfully all the salient objects with consistent values. However, map images from the other methods are more or less incomplete and inconsistent. Thus, these results illustrate clearly the effective- ness of our proposed algorithms.

7. Conclusion and future work

In this paper, a Schatten-q quasi-norm induced matrix decomposition model for conducting the salient object detection is presented. In the proposed method, the non-convex Schatten-1/2 or Schatten-2/3 norm is applied to formulate the background, that provides cleaner solution. A weighted group sparsity induced norm is imposed on foreground to share consistent within the same image patches. Meanwhile, an alternative non-convex formulation is proposed for nuclear norm (i.e. $q = 1$). Efficient numerical algorithms with closed-form solutions are established to solve the proposed model under the optimization framework, whose convergence and complexity are also discussed. Experiments on the six competitive datasets show that the proposed model (SQNMD) can outperform most of the state-of-the-art models.

An interesting direction of future work is how to extract high-level semantic information of salient object using tensor analysis. In addition, the adaptive selection of trade-off parameters is worthy of a further study for the purpose of more efficient implementation.

Acknowledgments

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.patcog.2019.106975.

References

[1] Y. Chen, T. Huang, K. Chang, Y. Tsai, H. Chen, B. Chen, Quantitative analysis of automatic image cropping algorithms:a dataset and comparative study, in: 2017 IEEE Winter Conference on Applications of Computer Vision, 2017, pp. 226–234.
[2] W. Yin, J. Luo, C.W. Chen, Event-based semantic image adaptation for user-centric mobile display devices, IEEE Trans. Pattern Anal. Mach. Intell. 13 (2011) 432–442.
[3] P. Wang, D. Zhang, G. Zeng, J. Wang, Contextual dominant color name extraction for web image search, in: 2012 IEEE International Conference on Multimedia and Expo Workshops, 2012, pp. 319–324.
[4] W. Hou, X. Gao, X. Li, Visual saliency detection using information divergence, Pattern Recognit. 46 (2013) 2658–2669.
[5] C. Liu, P. Yuen, G. Qiu, Object motion detection using information theoretic spatio-temporal saliency, Pattern Recognit. 22 (2009) 2897–2906.
[6] Y. Zhang, M. Li, Motion segmentation using collaborative low-rank and sparse subspace clustering, in: 2017 13th International Conference on Computational Intelligence and Security, 2017, pp. 409–414.
[7] A. Borji, M. Cheng, H. Jiang, J. Li, Salient object detection: a benchmark, IEEE Trans. Image Process. 24 (2015) 5706–6572.
[8] J. Kim, D. Han, Y. Tai, J. Kim, Salient region detection via high-dimensional color transform, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 883–890.
[9] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 1597–1604.
[10] F. Perazzi, P. Krhenbhl, Y. Pritch, A. Hornung, Saliency filters: contrast based filtering for salient region detection, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 733–740.
[11] C. Scharfenberger, A. Wong, K. Fergani, J.S. Zelek, D.A. Clausi, Statistical textural distinctiveness for salient region detection in natural images, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 979–986.
[12] Y. Niu, Y. Geng, X. Li, F. Liu, Leveraging stereopsis for saliency analysis, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 454–461.
[13] J. Zhang, S. Sclaroff, Saliency detection: a boolean map approach, in: 2013 IEEE International Conference on Computer Vision, 2013, pp. 153–160.
[14] H.-H. Yeh, K.-H. Liu, C.-S. Chen, Salient object detection via local saliency estimation and global homogeneity refinement, Pattern Recognit. 47 (2014) 1740–1750.
[15] Y. Wei, F. Wen, W. Zhu, J. Sun, Geodesic saliency using background priors, in: Proceedings of the 12th European Conference on Computer Vision, 2012, pp. 29–42.
[16] Q. Yan, L. Xu, J. Shi, J. Jia, Hierarchical saliency detection, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 1155–1162.
[17] R. Margolin, A. Tal, L. Zelnik-Manor, What makes a patch distinct? in: 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 1139–1146.
[18] K. Chang, T. Liu, H. Chen, S. Lai, Fusing generic objectness and visual saliency for salient object detection, in: 2011 International Conference on Computer Vision, 2011, pp. 914–921.
[19] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, H. Shum, Learning to detect a salient object, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2011) 353–367.
[20] J. Yang, M. Yang, Top-down visual saliency via joint crf and dictionary learning, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 576–588.
[21] A. Kocak, K. Cizmeciler, A. Erdem, E. Erdem, Top down saliency estimation via superpixel-based discriminative dictionaries, Br. Mach. Vision Conf. (2014).
[22] Y. Yi, L. Su, Q. Huang, Z. Wu, C. Wang, Saliency detection with two-level fully convolutional networks, in: 2017 IEEE International Conference on Multimedia and Expo, 2017, pp. 271–276.
[23] J. Wang, H. Jiang, Z. Yuan, M. Cheng, X. Hu, N. Zheng, Salient object detection: a discriminative regional feature integration approach, Int. J. Comput. Vision 123 (2017) 251–268.
[24] X. Li, H. Lu, L. Zhang, X. Ruan, M. Yang, Saliency detection via dense and sparse reconstruction, in: 2013 IEEE International Conference on Computer Vision, 2013, pp. 2976–2983.
[25] C. Yang, L. Zhang, H. Lu, X. Ruan, M. Yang, Saliency detection via graph-based manifold ranking, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3166–3173.
[26] W. Zhu, S. Liang, Y. Wei, J. Sun, Saliency optimization from robust background detection, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2814–2821.
[27] E. Candès, B. Recht, Exact matrix completion via convex optimization, Found. Comput. Math. 9 (2009) 717–772.
[28] E. Candès, X. Li, Y. Ma, J. Wright, Robust principal component analysis, J. ACM 58 (2011) 1–39.
[29] X. Shen, Y. Wu, A unified approach to salient object detection via low rank matrix recovery, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 853–860.
[30] C. Lang, G. Liu, J. Yu, S. Yan, Saliency detection by multitask sparsity pursuit, IEEE Trans. Image Process. 21 (2012) 1327–1338.
[31] G. Liu, Z. Lin, Y. Yu, Robust subspace segmentation by low-rank representation, Int. Conf. Mach. Learn. (2010) 663–670.
[32] W. Zou, K. Kpalma, Z. Liu, J. Ronsin, Segmentation driven low-rank matrix recovery for saliency detection, 24th Br. Mach. Vision Conf. (2013) 1–13.
[33] H. Peng, B. Li, H. Ling, W. Hu, W. Xiong, S.J. Maybank, Salient object detection via structured matrix decomposition, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 818–832.
[34] X. Su, Z. He, C. Xu, X. Zhang, W. Zou, G. Baciu, Diversity induced matrix decomposition model for salient object detection, Pattern Recognit. 66 (2017) 253–267.
[35] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, J. Am. Stat. Assoc. 96 (2001) 1348–1361.
[36] C. Lu, J. Tang, S. Yan, Z. Lin, Generalized nonconvex nonsmooth low-rank minimization, in: IEEE 2014 Conference on Computer Vision and Pattern Recognition, 2014, pp. 4130–4137.
[37] F. Nie, H. Wang, X. Cai, H. Huang, C. Ding, Robust matrix completion via joint Schatten p-norm and l_p-norm minimization, in: 12th IEEE International Conference on Data Mining, 2012, pp. 566–574.
[38] H. Zhang, J. Yang, F. Shang, C. Gong, Z. Zhang., Tractable and scalable Schatten quasi-norm approximations for rank minimization, IEEE Trans. Cybern. (2018) 1–14.
[39] F. Shang, J. Cheng, Y. Liu, Z. Luo, Z. Lin, Bilinear factor matrix norm minimization for robust pca: algorithms and applications, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2018) 2066–2080.
[40] E. Kim, M. Lee, S. Oh, Elastic-net regularization of singular values for robust subspace learning, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 915–923.
[41] K. Jin, J. Ye, Annihilating filter-based low-rank Hankel matrix approach for image inpainting, IEEE Trans. Image Process. 24 (2015) 3498–3511.
[42] B. Recht, M. Fazel, P. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, SIAM Rev. 52 (2010) 471–501.

[43] R. Cabral, F. Torre, J. Costeira, A. Bernardino, Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition, in: 2013 IEEE International Conference on Computer Vision, 2013, pp. 2488–2495.

[44] J. Feng, H. Xu, S. Yan, Online robust pca via stochastic optimization, Adv. Neural Inf. Process. Syst. 26 (2013) 404–412.

[45] R. Sun, Z. Luo, Guaranteed matrix completion via non-convex factorization, IEEE Trans. Inf. Theory 62 (2016) 6535–6579.

[46] J. Liu, J. Ye, Moreau–Yosida regularization for grouped tree structure learning, in: Proceedings of the 23rd International Conference on Neural Information Processing Systems, 2010, pp. 1459–1467.

[47] K. Jia, T. Chan, Y. Ma, Robust and practical face recognition via structured sparsity, in: Proceedings of the 12th European Conference on Computer Vision, 2012, pp. 331–344.

[48] D. Cai, X. He, J. Han, T.S. Huang, Graph regularized nonnegative matrix factorization for data representation, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2011) 1548–1560.

[49] R. Vidal, Y. Ma, S. Sastry, Generalized Principal Component Analysis, Springer New York (2016).

[50] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn. 3 (2011) 1–122.

[51] R. Jenatton, J. Mairal, G. Obozinski, F.R. Bach, Proximal methods for hierarchical sparse coding, J. Mach. Learn. Res. 12 (2011) 2297–2334.

[52] J. Cai, E. Candès, Z. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optim. 20 (2010) 1956–1982.

[53] M. Cheng, G. Zhang, N. Mitra, X. Huang, S. Hu, Global contrast based salient region detection, in: 2011 IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 409–416.

[54] D. BatraEmail, A. Kowdle, D. Parikh, J. Luo, T. Chen, Interactively co-segmentating topically related images with intelligent scribble guidance, Int. J. Comput. Vision 93 (2011) 273–292.

[55] Y. Li, X. Hou, C. Koch, J. Rehg, A.L. Yuille, The secrets of salient object segmentation, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 280–287.

[56] B. Jiang, L. Zhang, H. Lu, C. Yang, M. Yang, Saliency detection via absorbing Markov chain, in: 2013 IEEE International Conference on Computer Vision, 2013, pp. 1665–1672.

[57] Z. Bylinskii, T. Judd, A. Oliva, A. Torralba, F. Durand, What do different evaluation metrics tell us about saliency models?, arXiv:1604.03605 (2016).

[58] R. Margolin, L. Zelnik-Manor, A. Tal, How to evaluate foreground maps? in: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 248–255.

Min Li received the Ph.D. degree in applied mathematics from Xidian University, Xi'an, China, in 2008. She is currently an associate processor with the college of Mathematics and Statistics, Shenzhen University, Shenzhen, China. Her research interests include the field of partial differential equations, geometric multi-scale analysis, subspace clustering, machine learning, saliency detection and computer vision.

Yao Zhang is a master degree candidate at the college of Mathematics Statistics, Shenzhen University, Shenzhen. His research focuses on the mathematics and computer vision.

Mingqing Xiao received his Ph.D. in Applied Mathematics from University of Illinois at Urbana-Champaign in 1997. He is professor of Southern Illinois University at Carbondale and a Senior IEEE member. He has been served in many editorial broads of various top journals such as IEEE Transaction on Automatic Control, Automatica, European Journal of Control, etc. His current research interests are optimization, image processing, control theory, and numerical analysis.

Chen Xu received the B.S. and M.S. degree in mathematics form Xidian University, Xi'an, China, in 1986 and 1989, respectively, and the Ph.D. degree in Mathematics from Xian Jiaotong University, Xi'an, China, in 1992. He is currently a professor of Mathematics and an Advisor for doctoral students at Shenzhen University, Shenzhen, China. His research interests include information and computational science, and analysis and application of wavelet.

Weiqiang Zhang received his Ph. D. degree in applied mathematics from Xidian University, Xi'an, China, in 2007. He is currently a associate processor with the college of Mathematics and Statistics, Shenzhen University, Shenzhen, China. His research interests include the field of Wavelet, geometric multi-scale analysis and computer vision.