Superpixel Segmentation Based on Spatially Constrained Subspace Clustering

Hua Li, Yuheng Jia, Runmin Cong, Member, IEEE, Wenhui Wu, Sam Kwong, Fellow, IEEE, and Chuanbo Chen

Abstract—Superpixel segmentation aims at dividing the input image into some representative regions containing pixels with similar and consistent intrinsic properties, without any prior knowledge about the shape and size of each superpixel. In this paper, to alleviate the limitation of superpixel segmentation applied in practical industrial tasks that detailed boundaries are difficult to be kept, we regard each representative region with independent semantic information as a subspace, and correspondingly formulate superpixel segmentation as a subspace clustering problem to preserve more detailed content boundaries. We show that a simple integration of superpixel segmentation with the conventional subspace clustering does not effectively work due to the spatial correlation of the pixels within a superpixel, which may lead to boundary confusion and segmentation error when the correlation is ignored. Consequently, we devise a spatial regularization and propose a novel convex locality-constrained subspace clustering model that is able to constrain the spatial adjacent pixels with similar attributes to be clustered into a superpixel and generate the content-aware superpixels with more detailed boundaries. Finally, the proposed model is solved by an efficient alternating direction method of multipliers (ADMM) solver. Experiments on different standard datasets demonstrate that the proposed method achieves superior performance both quantitatively and qualitatively compared with some state-of-the-art methods.

Index Terms—Superpixel segmentation, subspace clustering, locality-constrained, spatial correlation

This work was supported by the Key Project of Science and Technology Innovation 2030 supported by the Ministry of Science and Technology of China under Grant 2018AAA0103101, in part by the General Science Foundation of China under Grants 61772344, 62020104, in part by the Hong Kong RGC General Research Funds under 9042816 (CityU 11209819), in part by the Beijing Nova Program under Grant Z201100006B20016, in part by the Fundamental Research Funds for the Central Universities under Grant 2019RC039, in part by the Beijing Nova Program under Grant Z2012030050, Grant 2019M660438. (Corresponding author: Sam Kwong.)

H. Li is with the School of Software Engineering, Huazhong University of Science and Technology, Wuhan 430074, China, and also with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong (e-mail: huaili27-c@my.cityu.edu.hk). Y. Jia is with the School of Computer Science and Engineering, Southeast University, Nanjing, 211189, China (e-mail: yjia@seu.edu.cn). R. Cong is the Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China, and also with the Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing Jiaotong University, Beijing 100044, China (rmcong@bjtu.edu.cn). W. Wu is with the College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: wuwen.hui@szu.edu.cn). S. Kwong is with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong, and also with the City University of Hong Kong Shenzhen Research Institute, Shenzhen 518057, China (e-mail: csshankl@cityu.edu.hk). C. Chen is with the School of Software Engineering, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: chuanboc@163.com).

I. INTRODUCTION

Superpixels are alternative primitives to represent the original pixels in an image based on certain intrinsic properties, such as coherent spatial relationship, and similar appearance. Instead of utilizing pixel entities directly, superpixels can provide semantic representation and reduce the dimension of primitives [1], [2], which have been widely applied in practical and industrial tasks of computer vision, such as saliency detection [3]–[20], feature selection [21], object recognition [22], background subtraction [23], image enhancement and reconstruction [24]–[26]. In recent years, numerous methods with diverse problem formulations have been proposed for superpixel segmentation [27]–[31] and achieved excellent performance in theoretical study. However, one obvious weakness of the existing superpixel methods is the limitation to adapt to the local image details, such as the content boundaries and object contours. And in many cases of practical problems and industrial tasks [3]–[5], [22], the local details are the key information, which should not be ignored. For most methods, the detailed boundaries of image content are still hard to be well preserved, due to the lack of the prior knowledge about the shape and size of the superpixels in an image. These methods have to increase the number of superpixels and reduce the superpixel size to achieve the detailed boundaries. This may usually lead to large data redundancy in the sparse area, which is unwillingness for practical tasks. How to balance the detailed information and superpixel number is a challenging problem for superpixel segmentation to be applied in industrial tasks.

To alleviate this problem, we suppose each superpixel with independent information comes from one distinct subspace, and formulate the superpixel segmentation as a subspace clustering problem [22]–[24]. Subspace clustering is an effective way to segment data drawn from different dimensions or subspaces, and has achieved impressive performance in some real-world tasks, such as face clustering [34] and foreground extraction [35]. Moreover, it also has been adopted in industrial applications, e.g., hyperspectral image classification based on superpixel-level subspace learning [36]–[38]. For superpixel segmentation, each superpixel is regarded as a subspace containing pixels with similar intrinsic attributes, such as similar texture, boundary, color and coherent spatial information, while pixels in different subspace have different properties. Then the solution to the subspace clustering corresponds to the segmentation of superpixels. By this way, the pixels in the unit containing independent semantic information are more probably segmented into one superpixel, no matter...
how small the unit area is. At the same time, the size of the superpixels in sparse area are relatively large according to the similar attributes of pixels. Consequently, the detailed content boundaries and the superpixel number can be better balanced, i.e., using less superpixels to preserve more detailed information.

Nevertheless, it is still a challenging work by inducing subspace clustering theory to superpixel segmentation directly. Different from the general sample data in clustering, pixels in an image have certain spatial correlation that only the similar pixels in adjacency tend to be clustered as one superpixel. Directly applying the conventional subspace clustering models to superpixel segmentation task can not excellently work since the spatial correlation of pixels is not taken into consideration. This may cause the confusion and segmentation error of content boundaries. In other words, we should consider the spatial constraints when using the subspace clustering for superpixel segmentation task. To this end, we devise a spatial regularization term to emphasize the spatial correlation, and propose a spatially constrained subspace clustering based superpixel segmentation model to generate superpixels with more accurate and detailed boundaries, which is more appropriate for practical and industrial tasks.

In this paper, considering the limitation of superpixel segmentation in the application of practical industrial tasks, we propose a novel spatially constrained subspace clustering based superpixel segmentation model to produce more adaptive superpixels. In summary, the contributions of this paper include:

- Considering the challenge of pursuing detailed boundaries, we formulate the superpixel segmentation task as a subspace clustering problem to generate content-aware superpixels with more detailed boundaries and fewer primitives. To the best of our knowledge, this is the first attempt to induce subspace clustering theory to solve the superpixel segmentation problem, and effectively alleviate the challenge in superpixel segmentation.
- To enforce the spatial adjacent pixels with similar attributes to be clustered into a superpixel corresponding to a subspace, we design a spatial regularization term in the loss function to constrain the local adjacency when clustering pixels. The ablation study has verified the effectiveness of the spatial regularization term.
- The extensive experiments and discussions on the widely used benchmarks and datasets demonstrate that the proposed method achieves superior performance both quantitatively and qualitatively.

The paper is organized as follows. The related works are presented in Section II. In Section III, we propose a novel superpixel segmentation model. The comparative experimental results and analyses are presented in Section IV. Finally, Section V concludes this paper.

II. RELATED WORK

Various superpixel segmentation algorithms have been proposed for different motivations and tasks. They can be classified into two categories: unsupervised approach and supervised approach.

A. Unsupervised Superpixel Segmentation Methods

Unsupervised approach mainly includes two categories: graph-based methods and clustering-based methods.

Normalized cut (Ncut) algorithm [39] is one of the representative graph-based methods, which first groups pixels into several large areas by normalized cut, and then adopts K-means to segment the large areas into small units. However, it suffers from high computational complexity cost and low boundary adherence, which limit its utilization. Considering the image topological structure, topology preserved regular superpixel (TPS) [2] algorithm was put out to produce relatively regular superpixel. However, it suffers from low adherence to boundaries due to the demand of regular superpixel shape. In contrast to TPS, lazy random walk (LRW) [27] algorithm achieves high object boundary adherence by iteratively optimizing an energy function based on texture measurement. Recently, approximately structural superpixels (ASS) [31] was proposed to generate approximately structural superpixels by an asymmetrical square-wise superpixel segmentation way, which can largely reduce data amount as well as preserve image content boundaries.

Turbopixel [40] is one of the early clustering-based methods, which adopts local image gradient and geometric flow to generate regular superpixel, but it suffers from the time consuming problem in practice. Simple linear iterative clustering (SLIC) [29] is widely used due to its high efficiency. SLIC adopts K-means clustering approach and changes the search area into a local area rather than the global searching in K-means, which largely reduces the computation complexity. As an extension of SLIC, Fast linear iterative clustering (FLIC) [30] is an active search method which emphasizes the neighboring continuity. Linear spectral clustering (LSC) [29] generates superpixels by mapping pixels into a high-dimensional space through a kernel function. Recently, Bayesian Adaptive Superpixel Segmentation (BASS) [41] is proposed to generate adaptive superpixels by a Bayesian mixture model. However, the number of superpixel cannot be predefined, which is not convenient for certain tasks that demand to initialize the number manually.

B. Supervised Superpixel Segmentation Methods

Supervised superpixel segmentation algorithms are mainly based on deep-learning based approaches. Recently, deep learning based methods for superpixel segmentation have been proposed to incorporate deep learning technique with traditional segmentation. There are three representative deep learning based superpixel segmentation methods, Segment-aware Loss(SEAL) [42], Superpixel Sampling Networks(SSN) [43], and Superpixel Segmentation with Fully Convolutional Networks (FCN) [44].

SEAL adopts deep neural network and designs a segmentation-aware affinity learning method to keep weak object boundary in superpixel segmentation. SSN presents a differentiable superpixel sampling model which can be integrated into end-to-end trainable networks. FCN algorithm adopts a simple fully connected network to predict superpixel on regular image grids. Then, on the basis of the predicted
superpixel segmentation by constructing an affinity matrix by learning effective representation coding. For efficiency, we use the extracted features from the raw pixel units rather than the raw pixel features as input. Here, a self-expressive model \cite{Li2013} is adopted to learn the affinity matrix for efficiency, and the data is explained by itself: a data point can be represented as a linear combination of other data points in the same subspace, \textit{i.e.}, \( X = X Z + E \), where \( X = [x_1, x_2, \cdots, x_n] \) is the feature matrix, \( Z \) is the coding matrix, and \( E \) is the error matrix. Consequently, we first formulate the superpixel segmentation problem as:
\[
\min_{Z,E} \frac{1}{2} \|E\|_F^2 + \lambda \|Z\|_0, \quad \text{s.t.} \ X = XZ + E, \ diag(Z) = 0_n, \tag{1}
\]
where \( \lambda \) is a trade-off parameter, and \( 0_n \in \mathbb{R}^{n \times 1} \) is an all zeros vector. The constraint \( diag(Z) = 0_n \) is used to eliminate the trivial solution, \textit{i.e.}, \( Z \) equals to an identity matrix. The term \( \|Z\|_0 \) is a regularization to enforce the sparsity, representing the global subspace structure in \( X \). \( \ell_0 \)-norm means to find the sparsest solution. However, minimizing \( \|Z\|_0 \) is extremely difficult since it corresponds to an NP-hard problem. To find a nontrivial sparse representation efficiently, the \( \ell_0 \)-norm is usually relaxed to its convex surrogate, \textit{i.e.}, the \( \ell_1 \)-norm \cite{Li2013}.

Nevertheless, directly applying the conventional subspace clustering models to superpixel segmentation task does not work well. Fig. 2 shows the segmentation results of some state-of-the-art superpixel methods. From Fig. 2(b)-(d), we can see that the spots in the butterfly wing are almost unrecognizable. In Fig. 2(e), we adopt the subspace clustering model in \cite{Li2013} to generate superpixels, and the boundaries of the spots are also not achieved. The reason is that different from the general sample data in clustering, image pixels have inherent spatial correlations, which is not taken in the consideration of the conventional models. Therefore, we design a locality constraint to pursue the spatial correlation.

### B. Superpixels via Locality-Constrained Subspace Clustering

Considering the spatial correlation of pixels in an image that only the similar pixels in adjacency tend to be clustered as one superpixel, we devise a spatial regularization term to
Therefore, we propose a locality-constrained subspace clustering model to generate content-aware superpixels with more detailed boundaries.

**Spatial Regularization:** Considering a data set $X = \{x_1, x_2, \ldots, x_n\}$ drawn from input image units, for image superpixel segmentation, we devise a spatial regularization term $f(Z)$ to incorporate the neighbor relationship information in $X$. The neighbors of the $i$-th column $x_i$ (e.g., $x_{i-1}, x_{i+1}$) could be similar to $x_i$. In other words, in the coding space, the corresponding neighbors of $z_i$, which is a new representation for $x_i$, could be close to $z_i$. The spatial regularization term $f(Z)$ is defined as:

\[
f(Z) = \|ZW\|_{2,1}. \tag{2}\]

The \(\ell_{2,1}\)-norm adopted here aims to achieve the similarity between neighboring pixels. \(W\) is the weight matrix enforcing the neighboring correlations in $X$, which is defined as:

\[
W = \bar{\omega} - D, \tag{3}\]

where $\omega$ is the spatial adjacent matrix of $X$, $\bar{\omega}$ is the lower triangular matrix of $\omega$, and $D_{ij} = \sum_{i=1}^{n} \bar{\omega}_{ij}$.

To show the spatial regularization term more clearly, denoted the adjacent relationship as $\omega$, we illustrate a simple case in Fig. 3. Specifically, $W$ is induced as follows:

\[
\omega = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}, \quad \bar{\omega} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix},
\]

\[
D = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad W = \begin{pmatrix} -2 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}.
\]

Therefore, $ZW = \begin{bmatrix} z_2 + z_3 - 2z_4, z_4 - z_2, z_4 - z_3, 0 \end{bmatrix}$, which is devised to enforce the neighbors more probably to be segmented into the same superpixel.

**Locality-Constrained Subspace Clustering Model:** Integrated with the spatial regularization in (2), we propose a locality-constrained subspace model for image superpixel segmentation:

\[
\begin{align*}
\min_{Z} & \quad \frac{1}{2} \|X - XZ\|_F^2 + \lambda_1 \|Z\|_1 + \lambda_2 \|ZW\|_{2,1}, \\
\text{s.t.} & \quad \text{diag}(Z) = 0_n.
\end{align*} \tag{4}
\]

where $\lambda_1$ and $\lambda_2$ are the trade-off parameters to balance the weights of global regularization and spatial regularization, respectively.

**Algorithm 1** Locality-Constrained Subspace Clustering Superpixel

**Input:** An image $I$, desired number of superpixels $K$, initial number of pixel units $n$  
**Output:** $K$ superpixels

1. Initially segment the image $I$ into $n$ raw pixel units by K-means.
2. Extract features of the raw pixel units as input data set $X = \{x_1, x_2, \ldots, x_n\}$ (each column is the feature vector of the corresponding pixel unit).
3. Construct matrices $\bar{\omega}, D$ and $W$.
4. Solve the coding matrix $Z$ according to Algorithm 2.
5. Build the affinity graph $G$ with respect to $Z$ according to (5).
6. Segment $G$ with Ncut to generate $K$ clusters.
7. Merge isolated pixels with the nearest neighbor clusters and update $K$.

After working out the coding matrix $Z$ (see the detailed optimization algorithm in Section III-C), an affinity graph $G$ can be constructed for segmenting the coding matrix $Z$ and determining the subspace clusters. Similar to [34], the affinity graph $G$ is defined as:

\[
G = |Z^T| + |Z| - 2. \tag{5}\]

This symmetrization transformation aims to assure the two connected points are in each other’s sparse representation, to achieve the ideal correspondence of each element to the points from the same subspace. Then, Ncut [39] is adopted to generate the superpixel clusters with $G$.

Moreover, in image superpixel segmentation, there may exist some isolated pixels not belonging to any clusters after the clustering step. Generally, such pixels are enforced to be assigned to the nearest cluster. The complete procedure of the proposed method is summarized in Algorithm 1. We show the result using our locality-constrained subspace clustering method in Fig. 2(b). It is clear that compared with other methods, only our method preserves the boundaries of the spots. More experimental results are presented in Section IV.

**C. Optimization**

To solve the objective function in (4), we adopt alternating direction method of multipliers (ADMM) [45] to devise an optimization algorithm. Two auxiliary variables $U$ and $V = UW$ are induced to separate the terms of variable $Z$. Hence, (4) is equivalent to be written as:

\[
\min_{Z,U,V} \frac{1}{2} \|X - XU\|_F^2 + \lambda_1 \|Z\|_1 + \lambda_2 \|V\|_2, \tag{6}\]

s.t. \(\text{diag}(Z) = 0_n, U = Z, V = UW\).

The augmented Lagrangian function of (6) is:

\[
\mathcal{L}(Z,U,V) = \frac{1}{2} \|X - XU\|_F^2 + \lambda_1 \|Z\|_1 + \lambda_2 \|V\|_2, \tag{7}\]

\[
+ \langle \Theta, Z - U \rangle + \frac{\rho}{2} \|Z - U\|_F^2 + \langle \Xi, V - UW \rangle + \frac{\rho}{2} \|V - UW\|_F^2.
\]

Fig. 3. A simple case of adjacent matrix. If two units are neighbors, the corresponding element is labeled as “1” in the adjacent matrix; otherwise, the element is labeled as “0”.
where $\Theta \in \mathbb{R}^{n \times n}$ and $\Xi \in \mathbb{R}^{n \times n}$ are the Lagrangian multiplier matrices, $\alpha_1$ and $\alpha_2$ are the regularization parameters.

We can solve for $U$, $V$, and $Z$ alternatively by minimizing $\mathcal{L}$ when fixing the other two.

1) Update $Z$ when fixing $U$ and $V$: The problem in (7) becomes

$$\min_z \lambda_1 \|Z\|_1 + \frac{\alpha_1}{2} \left\| Z - \left( U - \frac{\Theta}{\alpha_1} \right) \right\|^2_F.$$  

(8)

Since $\mathbb{R}$ is element-wise to each $Z_{ij}$, by soft thresholding operator [46], we can get a closed-form solution:

$$Z = \text{sign} \left( U - \frac{\Theta}{\alpha_1} \right) \max \left( \left| U - \frac{\Theta}{\alpha_1} \right| - \frac{\lambda_1}{\alpha_1} \right).$$  

(9)

2) Update $U$ when fixing $Z$ and $V$:

$$\min_u \frac{1}{2} \left\| X - XU \right\|^2_F + \langle \Theta, Z - U \rangle + \frac{\alpha_2}{2} \left\| Z - U \right\|^2_F + \langle \Xi, V - UW \rangle + \frac{\alpha_2}{2} \left\| V - UW \right\|^2_F.$$  

(10)

By setting the derivative with respect to $U$ to zero, we can obtain the following equation:

$$(X^T X + \alpha_1 I) U + \alpha_2 VW^T = X^T X + \alpha_2 VW^T + \alpha_1 Z + \Theta + \Xi W T.$$  

(11)

(11) is a standard Sylvester equation [47], which can be linearized and solved by:

$$[I \otimes (X^T X + \alpha_1 I) + \alpha_2 W V^T \otimes I] \text{vec}(U) = \text{vec}(X^T X + \alpha_2 W V^T + \alpha_1 Z + \Theta + \Xi W T),$$  

(12)

where $\otimes$ is the tensor product and $\text{vec}(\cdot)$ returns a column vector by stacking the columns of a matrix.

3) Update $V$ when fixing $Z$ and $U$:

$$\min_v \lambda_2 \|V\|_{2,1} + \langle \Xi, V - UW \rangle + \frac{\alpha_2}{2} \left\| V - UW \right\|^2_F.$$  

(13)

(13) is equivalent to

$$\min_v \lambda_2 \|V\|_{2,1} + \frac{\alpha_2}{2} \left\| V - \left( UW - \frac{1}{\alpha_2} \Xi \right) \right\|^2_F.$$  

(14)

Let $N = UW - \frac{1}{\alpha_2} \Xi$. (14) can be solved by a closed-form solution [48]:

$$V(:,i) = \begin{cases} \frac{\|N(:,i)\| - \frac{\lambda_2}{\alpha_2}}{\|N(:,i)\|}, & \text{if } \|N(:,i)\| > \frac{\lambda_2}{\alpha_2}, \\ 0, & \text{otherwise}, \end{cases}$$  

(15)

where $V(:,i)$ denotes the $i$th column of $V$.

4) Update $\Theta$, $\Xi$, $\alpha_1$ and $\alpha_2$ by:

$$\Theta = \Theta^{old} + \alpha_1 (Z - U),$$  

(16)

$$\Xi = \Xi^{old} + \alpha_2 (V - UW),$$  

(17)

$$\alpha_1 = \min(\rho \alpha_1^{old}, \alpha_1^{max}), \quad \alpha_2 = \min(\rho \alpha_2^{old}, \alpha_2^{max}),$$  

(18)

where $\Theta \in \mathbb{R}^{n \times n}$ and $\Xi \in \mathbb{R}^{n \times n}$ are initialized as all zeros matrix, $\rho = 1.1$, $\alpha_1$ and $\alpha_2$ are both initialized as 0.01, $\alpha_1^{max} = 10^6$, $\alpha_2^{max} = 10^8$.

The complete optimization algorithm is summarized in Algorithm 2.

**Algorithm 2** Optimization to (4)

**Input**: $X$, $W$, maximum iterations $\text{iter}_{\text{max}}$, parameters $\lambda_1$, $\lambda_2$, $i = 0$

**Output**: $Z$

1: while not converged and $i \leq \text{iter}_{\text{max}}$ do
2: $Z_{i+1} = \arg \min_Z \mathcal{L}(Z_i, U_i, V_i)$; (referred to (9))
3: $U_{i+1} = \arg \min_U \mathcal{L}(Z_i, U_i, V_i)$; (referred to (12))
4: $V_{i+1} = \arg \min_V \mathcal{L}(Z_i, U_i, V_i)$; (referred to (15))
5: Update $\Theta$, $\Xi$, $\alpha_1$ and $\alpha_2$ according to (16)-(18);
6: $i = i + 1$;
7: end while

**D. Discussion**

In the proposed model, three blocks of variables $Z$, $V$, and $U$ need to be optimized in eq. (6). In each iteration, with the help of the ADMM, the objective function can converge to a local optimum with the optimized variables. Though there is no general convergence proof for the ADMM based algorithm with more than two blocks of variables [49], to the best of our knowledge, we observe that our algorithm empirically converges well and produces satisfactory performance on real-world data sets, contributing to the fact that the proposed model is convex, and each subproblem has a closed form solution.

The convergence curve is illustrated in Fig. 4. The algorithm can be updated iteratively until it the stopping conditions are satisfied, i.e., the iteration error is reduced to a stable value ($10^{-3}$ in our experiments) after several iterations or the algorithm exceeds the maximum iteration. We empirically find that 20 iterations are sufficient for most images, and thus we set the maximum iteration number to 20 for efficiency.

The computational complexity of the proposed method is dominated by the initialization by K-means and the optimization process. The computational complexity of K-means is $O(N)$, where $N$ is the total number of pixels in the input image. That of the optimization process is $O(k^3)$, where $k$ is the initial number of pixel units, which is far less than the total pixel number $N$. Hence, the time complexity for the proposed method is $O(N + k^3)$.

**IV. EXPERIMENTAL RESULTS AND DISCUSSIONS**

To show the merits qualitatively and quantitatively, we compare the proposed method with some state-of-the-art methods,
produce raw pixel units to keep the dimension of input data at 154401. To speed up the computation, we use K-means to produce segments from the ground truth boundaries. A high percentage of the most labeled pixels of superpixel segments are preserved more accurately than other methods. Ncut, LRW and TPS cannot catch most of the boundaries in the doggies’ boundaries, and the most boundaries of noses. The reason is that in our method, each superpixel is regarded as a subspace containing pixels with similar intrinsic attributes, such as similar texture, boundary, color and coherent spatial information. By this way, the pixels in the unit containing independent semantic information are more probably segmented into one superpixel, no matter how small the unit area is.

It is clear that other methods can also capture more image boundaries like ours if the desired superpixel number \( K \) increases heavily. In Fig. 7, we investigate the minimum desired superpixel number \( K \) for accurate spot segmentation in the butterfly image. The proposed method can segment the spot when \( K = 300 \), while that of LRW, SLIC, and LSC are 1600, 1800, and 1600, respectively. This shows that our method can achieve more details and semantic information with less superpixels compared with other methods.

Subsequently, the proposed method is compared with three recent deep learning based methods, and Fig. 8 illustrates some visual comparison.
Fig. 6. Visual comparison with some state-of-the-art methods. The first, third, and fifth columns are the image superpixel maps. The second, fourth, and sixth columns are the corresponding magnified regions.

Fig. 7. We compare the minimum initial settings of the desired superpixel number $K$ for spot segmentation in the butterfly image. The proposed method can segment the spot when $K = 300$, while LRW, SLIC, and LSC are 1600, 1800, and 1600, respectively.

visual results. It can be seen that our method can preserve more detailed boundaries than the deep-learning based methods. In the first image, only SEAL and our method segment the logo of the airplane. In the second image, both the two athletes’ number “1257” and “282” are mostly recognizable in the proposed method, while in FCN, SSN, and SEAL, the number “1257” are unrecognizable. So our method is comparable with deep learning based methods, and even has superiority in detail preservation. It is also worth pointing out that our method is unsupervised, which does not need any ground truth information, while the compared deep-learning based methods all need some labeled samples that are costly to acquire.
Fig. 8. Visual comparison with three recent deep learning based methods, FCN, SSN and SEAL. (a) is the original image and corresponding magnified regions. (b), (c) (d) and (e) are the superpixel label maps of FCN, SSN, SEAL, and the proposed method, respectively.

D. Quantitative Comparison

Fig. 9 shows the quantitative comparisons with some state-of-the-art methods. Fig. 9 (a) and Fig. 9 (b) are the BR and USE metric results, which show that our method can largely improve the performance and outperforms all the other compared methods. Taking 300 superpixels for example, our method’s minimum percentage gain (computed with the highest score of the compared methods) of USE is 62.5%, while that of BR is 6.7%. Fig. 9 (c), Fig. 9 (d), and Fig. 9 (e) are the ASA scores on BSDS500, DAVIS, Cityscapes datasets, respectively. In terms of ASA, the proposed method is competitive and improves the performance especially when the number of superpixels is small. Fig. 9 (g), and Fig. 9 (h) are the BR and USE score on NYUv2 dataset. The performance of our method also achieves the best performance in this dataset. The minimum and maximum percentage gains of BR are 3.60% and 41.15%, respectively. The minimum and maximum percentage gains of USE are 3.43% and 40.22%, respectively. The proposed method can largely improve the segmentation quality.

Fig. 9 (i) shows the BR scores of our method and deep learning based methods FCN, SEAL and SSN on the 200 images of test in BSDS500. Compared with the deep learning based methods, the BR score of our method is higher than SEAL and FCN, and comparable with SSN. Moreover, our method performs better than SSN when the number of superpixels is less than 300.

Our method is implemented in Matlab and runs on a PC with a 3.6 GHZ Intel Core i7-6850K CPU and 16GB RAM. Fig. 9 (j) shows the trade-off between efficiency and performance compared with some state-of-the-art methods when desired superpixel number is 100 and 600. According to the BR
In conclusion, the quantitative results show that the proposed method performs better than the compared state-of-the-art methods by standard evaluation criteria.

scores, the proposed method can achieve the best segmentation performance, slightly slower than SLIC and LSC, but faster than others.

In conclusion, the quantitative results show that the proposed method performs better than the compared state-of-the-art methods by standard evaluation criteria.
TABLE I

| $K$ | Metric | $E + GR$ | $E + SR$ | $E + GR + SR$ |
|-----|--------|----------|----------|---------------|
| 100 | BR     | 0.9188   | 0.9204   | 0.9355        |
|     | USE    | 0.1024   | 0.0934   | 0.0884        |
|     | ASA    | 0.9470   | 0.9528   | 0.9550        |
| 200 | BR     | 0.9256   | 0.9276   | 0.9473        |
|     | USE    | 0.0884   | 0.0854   | 0.0827        |
|     | ASA    | 0.9504   | 0.9532   | 0.9580        |
| 300 | BR     | 0.9492   | 0.9522   | 0.9602        |
|     | USE    | 0.0843   | 0.0805   | 0.0746        |
|     | ASA    | 0.9523   | 0.9567   | 0.9622        |
| 400 | BR     | 0.9544   | 0.9578   | 0.9646        |
|     | USE    | 0.0789   | 0.0780   | 0.0714        |
|     | ASA    | 0.9557   | 0.9577   | 0.9640        |

E. Ablation Study

To verify the effectiveness of the global regularization ($GR$) and spatial regularization ($SR$) in the proposed model, we evaluate our method on BSDS500 by varying the trade-off parameters $\lambda_1$ (the weight of global term in (4)) and $\lambda_2$ (the weight of spatial term in (4)). Table I lists the results of our model without the $GR$ or $SR$ respectively with different $K$. $E$ is the error term in (4). It can be seen both the $SR$ and $GR$ are important to the proposed model, since $E + GR + SR$ always performs the best. Moreover, $E + SR$ performs better than $E + GR$, indicating the importance of the spatial prior term.

Moreover, Fig. 10 illustrates the BR, USE, and ASA scores with different $\lambda_1$ and $\lambda_2$. From Fig. 10 we can see that excessively increasing the weight of $\lambda_2$ reduces the performance of BR. When $\lambda_1 = 10^{-6}$ and $\lambda_2 = 10^{-4}$, all the metrics achieve the highest score. Therefore, we choose these two values for the parameters. For DAVIS and Cityscapes datasets, we adopted the same parameter values and achieved excellent results.

V. Conclusion

In this paper, we formulated image superpixel segmentation as a subspace clustering problem based on the assumption that an input image is constituted of several subspaces, and pixels with similar intrinsic attributes lie in the same subspace. We designed a novel local regularization term to enforce the spatial correlation, and the proposed spatially constrained subspace clustering model is able to generate content-aware superpixels with more detailed boundaries. The proposed method has been compared with some state-of-the-art methods, and the experiments on different standard datasets demonstrate that the proposed method achieves superior performance both quantitatively and qualitatively.

REFERENCES

[1] H. Fu, X. Cao, D. Tang, Y. Han, and D. Xu, “Regularity preserved superpixels and supervoxels,” IEEE Trans. Multimedia, vol. 16, no. 4, pp. 1165–1175, 2014.
[2] D. Tang, H. Fu, and X. Cao, “Topology preserved regular superpixel,” in Proc. ICME, 2012, pp. 765–768.
[3] R. Cong, J. Lei, H. Fu, W. Lin, Q. Huang, X. Cao, and C. Hou, “An iterative co-saliency framework for RGBD images,” IEEE Trans. Cybern., vol. 49, no. 1, pp. 233–246, January 2019.
[4] R. Cong, J. Lei, H. Fu, Q. Huang, X. Cao, and C. Hou, “Co-saliency detection for RGBD images based on multi-constraint feature matching and cross label propagation,” IEEE Trans. Image Process., vol. 27, no. 2, pp. 568–579, 2018.
[5] J. Zhang, J. Chen, Q. Wang, and S. Chen, “Spatiotemporal saliency detection based on maximum consistency supervoxels merging for video analysis,” IEEE Trans. Ind. Inform., vol. 16, no. 1, pp. 606–614, 2020.
[6] R. Cong, J. Lei, H. Fu, M.-M. Cheng, W. Lin, and Q. Huang, “Review of visual saliency detection with comprehensive information,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 10, pp. 2941–2959, 2019.
[7] R. Cong, J. Lei, H. Fu, J. Hou, Q. Huang, and S. Kwong, “Going from RGB to RGBD saliency: A depth-guided transformation model,” IEEE Transactions on Cybernetics, vol. 50, no. 8, pp. 3627–3639, 2020.
[8] Z. Chen, R. Cong, Q. Xu, and Q. Huang, “Dpanet: Depth potentiality-aware gated attention network for rgb-d salient object detection,” IEEE Transactions on Image Processing, 2020.
[9] Q. Zhang, R. Cong, C. Li, M.-M. Cheng, Y. Fang, X. Cao, Y. Zhao, and S. Kwong, “ Dense attention fluid network for salient object detection in optical remote sensing images,” IEEE Transactions on Image Processing, 2020.
[10] R. Cong, J. Lei, H. Fu, F. Porikli, Q. Huang, and C. Hou, “Video saliency detection via sparsity-based reconstruction and propagation,” IEEE Transactions on Image Processing, vol. 28, no. 10, pp. 4819–4831, 2019.
[11] C. Li, R. Cong, J. Hou, S. Zhang, Y. Qian, and S. Kwong, “Nested network with two-stream pyramid for salient object detection in optical remote sensing images,” IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 11, pp. 9156–9166, 2019.
[12] R. Cong, J. Lei, H. Fu, Q. Huang, X. Cao, and N. Ling, “HSCS: Hierarchical sparsity based co-saliency detection for RGBD images,” IEEE Transactions on Multimedia, vol. 21, no. 7, pp. 1660–1671, 2019.
[13] R. Cong, J. Lei, C. Zhang, Q. Huang, X. Cao, and C. Hou, “Saliency detection for stereoscopic images based on depth confidence analysis and multiple cues fusion,” IEEE Signal Processing Letters, vol. 23, no. 6, pp. 819–823, 2016.
[14] C. Li, R. Cong, S. Kwong, J. Hou, H. Fu, G. Zhu, D. Zhang, and Q. Huang, “ASIF-Net: Attention steered interweave fusion network for RGB-D salient object detection,” IEEE Transactions on Cybernetics, vol. PP, no. 99, pp. 1–13, 2020.
[15] Z. Chen, Q. Xu, R. Cong, and Q. Huang, “Global context-aware progressive aggregation network for salient object detection,” in AAAI, 2020, pp. 10.599–10.606.
[16] Y. Zhang, L. Li, R. Cong, X. Guo, H. Xu, and J. Zhang, “Co-saliency detection via hierarchical consistency measure,” in Proc. ICME, 2018, pp. 1–6.
[17] Q. Zhang, R. Cong, J. Hou, C. Li, and Y. Zhao, “CoADNet: Collaborative aggregation-and-distribution networks for co-salient object detection,” in Proc. NeurIPS, 2020.
[18] C. Li, R. Cong, Y. Piao, Q. Xu, and C. C. Loy, “RGB-D salient object detection with cross-modality modulation and selection,” in Proc. ECCV, 2020, pp. 1–17.
[19] F. Li, R. Cong, H. Bai, and Y. He, “RGB-D salient object detection with cross-modality modulation and selection,” in Proc. IJCAI, 2020, pp. 534–543.
[20] C. Guo, C. Li, J. Guo, R. Cong, H. Fu, and P. Han, “Hierarchical features driven residual learning for depth map super-resolution,” IEEE Transactions on Image Processing, vol. 28, no. 5, pp. 2545–2557, 2019.
[21] N. Lv, C. Chen, T. Qiu, and A. K. Sangiah, “Deep learning and superpixel feature extraction based on contractive autoencoder for change detection in sar images,” IEEE Trans. Ind. Inform., vol. 14, no. 12, pp. 5530–5538, Dec 2018.
[22] X. Zhou, Y. Wang, Q. Zhu, J. Mao, C. Xiao, X. Lu, and H. Zhang, “A surface defect detection framework for glass bottle bottom using visual attention model and wavelet transform,” IEEE Trans. Ind. Inform., vol. 14, no. 12, pp. 5530–5538, Dec 2018.
[23] C. Li, R. Cong, Y. Pang, and B. Wang, “Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior,” IEEE Trans. Image Process., vol. 25, no. 12, pp. 5664–5677, 2016.
