



Article

Tensor Decomposition for Salient Object Detection in Images

Junxiu Zhou ^{1,*} , Yangyang Tao ² and Xian Liu ¹

¹ Department of Systems Engineering, University of Arkansas at Little Rock, Little Rock, AR 72204, USA; xxliu@ualr.edu

² Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA; ytao13@stevens.edu

* Correspondence: jxzhou@ualr.edu

Received: 23 May 2019; Accepted: 17 June 2019; Published: 19 June 2019



Abstract: The fundamental challenge of salient object detection is to find the decision boundary that separates the salient object from the background. Low-rank recovery models address this challenge by decomposing an image or image feature-based matrix into a low-rank matrix representing the image background and a sparse matrix representing salient objects. This method is simple and efficient in finding salient objects. However, it needs to convert high-dimensional feature space into a two-dimensional matrix. Therefore, it does not take full advantage of image features in discovering the salient object. In this article, we propose a tensor decomposition method which considers spatial consistency and tries to make full use of image feature information in detecting salient objects. First, we use high-dimensional image features in tensor to preserve spatial information about image features. Following this, we use a tensor low-rank and sparse model to decompose the image feature tensor into a low-rank tensor and a sparse tensor, where the low-rank tensor represents the background and the sparse tensor is used to identify the salient object. To solve the tensor low-rank and sparse model, we employed a heuristic strategy by relaxing the definition of tensor trace norm and tensor l_1 -norm. Experimental results on three saliency benchmarks demonstrate the effectiveness of the proposed tensor decomposition method.

Keywords: salient object; saliency detection; low-rank matrix recovery; tensor decomposition

1. Introduction

Salient object detection has attracted a lot of interest in the computer vision community. It allows us to quickly retrieve information from the input image by focusing on the most conspicuous object regions [1]. The fundamental challenge of salient object detection is to find the decision boundary that separates the salient object from the background.

Many saliency detection methods have been proposed to address this challenge. Depending on whether large scale training samples are used or not, existing methods broadly fall into two categories: Supervised and unsupervised. In many situations, data-driven supervised methods can provide a workable solution strategy and produce a clean saliency map. Liu et al. [2] proposed to detect salient object by formulating the detection process as a binary labeling problem. They first extracted multiple levels (locally, regionally, and globally) image features. Following this, the features were used to train a conditional random field (CRF) for detecting the saliency map. Mai et al. [3] proposed to use a data-driven based CRF approach to perform saliency aggregation in order to mitigate the performance gaps among other salient object detection methods. Jiang et al. [4] formulated the saliency detection task as a regression problem via the random forest model. First, they divided the image into multi-level regions to form multi-level image segmentation. Then each region was represented by discriminative features. Later, the features

were used to train a random forest regressor to map them to saliency values. Finally, saliency values from different regions and levels were linearly combined to achieve the final saliency map. Recently, the successful applications of deep neural networks (DNNs) in various research fields open up another promising path for accurately detecting salient objects. In [5], a multi-context deep learning method for saliency detection is proposed. The work first employed a global-context convolutional neural network (CNN) to acquire the image-level information. It then created a local-context modeling CNN to get closer-focused local information, by inputting image patches. The results of the two networks were combined to obtain the final saliency map. Li et al. [6] proposed to use a pixel-level fully CNN stream to obtain the pixel-level saliency prediction result. At the same time, they used a segment-wise spatial pooling stream to extract the segment-wise feature. Finally, they used a CRF to incorporate the results of the two streams to generate the final saliency map. Hou et al. [7] proposed a short-connection deep supervised salient object detection method. They connected the output from the other side output of hidden layers to the last pooling layer in CNN to enhance the visual distinctive region output. These data-driven supervised methods achieved remarkable detection performance in detecting the task-specified salient object. However, the detection accuracy of these supervised methods was restricted by the quality and quantity of training samples and the high computational platform required for deep learning-based methods.

On the other hand, many sophisticated approaches have been proposed to deal with the salient object detection task in an unsupervised or weakly-supervised manner. Most of them attempted to identify saliency regions in a scene via low-level features, such as color, texture, and location. Among the unsupervised saliency detection methods, the most commonly used methods are contrast-based methods [8,9] and mathematical model-based methods [10–13]. As one of the contrast-based saliency detection methods, the work in [9] proposed a contrast-filtering method. First, they split the input image into basic elements that preserve the relevant image structure. After this, they calculated two contrast metrics based on the uniqueness and spatial distribution of the elements in order to find the significance value of each element. Finally, they used high-dimensional Gaussian filters to estimate the saliency map. In [8], Cheng et al. proposed a regional contrast saliency detection method. This method used the histogram-based contrast to calculate the global color information. Then, a region-based contrast was proposed to enhance the spatial information on the final saliency map. Ishikura et al. [14] proposed an unsupervised saliency detection method based on multi-scale extrema of local perceptual color difference. However, their method was only used to estimate the locations, sizes, and saliency levels of candidate regions instead of the pixel-level saliency map.

Research has also been conducted on saliency detection methods based on mathematical models. Wang et al. [11] adopted the site entropy rate model. First, the multi-band filter response maps were calculated, and then the response map was represented by a fully-connected graph based on cortical connectivity. The site entropy rate (SER) was used to measure the average information from a node to the others in the graph. The final saliency map was calculated by using the cumulative sum of all SERs. Later, Guo et al. [12] proposed a SER based multi-scale point set saliency detection method. The most representative mathematical model for saliency detection is the low-rank matrix recovery (LRMR)-based (or low-rank and sparse matrix decomposition) model. In the LRMR model, the saliency detection task is formulated as the decomposition of a feature matrix. The model decomposes the feature matrix into a sparse subspace and a low-rank subspace [10]. It assumes that the low-rank feature space represents the redundant background image information and the salient object is represented in the sparse feature space. These unsupervised methods have the generalized ability for unseen saliency detection tasks.

The LRMR based saliency detection method is simple and powerful. However, the detection performance is usually restricted by spatial consistency and incomplete detection results. It is because LRMR loses spatial consistency during the mathematical analysis process of converting three-dimensional (3D) image feature into two-dimensional (2D) matrix, while not taking full advantage of image features. In this article, our main interest is in seeking to maintain the spatial consistency to improve the detection

accuracy. With the aid of high-dimensional tensor decomposition, we are able to solve this issue and derive better salient object detection results.

The main contribution of this article is twofold: (1) to the best of our knowledge, this is the first attempt to integrate tensor decomposition into the salient object detection task. By doing this, we can naturally represent high-dimensional features with a tensor. We then propose a series relaxation optimization processes to simplify the optimization process of tensor decomposition and solve it by using a heuristic method; (2) tensor decomposition is capable of discovering more hidden relationships of image features than LRMR in the process of capturing saliency information. Therefore, it obtains better salient object detection accuracy as shown in the experimental results section.

2. Related Works

In this section, we briefly review the work of LRMR-based salient object detection and basic knowledge about tensors.

The core of LRMR is decomposing a 2D matrix space into a sparse subspace and a low-rank subspace. Mathematically, LRMR can be formulated as [15]:

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

where $\mathbf{F} \in \mathfrak{R}^{N \times D}$ represents the input 2D image or the feature matrix. As shown in Equation (1), \mathbf{F} is decomposed into the sum of a low-rank matrix $\mathbf{L} \in \mathfrak{R}^{N \times D}$ and sparse matrix $\mathbf{S} \in \mathfrak{R}^{N \times D}$ by solving optimization problem in Equation (1). $\|\cdot\|_*$ stands for the nuclear norm which computes the sum of the singular values of a matrix. It is a convex relaxation of the matrix rank function. $\|\cdot\|_1$ represents the l_1 -norm to promote the sparsity. λ is a constant to balance the tradeoff between the low-rank term and the sparse term.

In the literature on salient object detection via LRMR, the work by Yan et al. [16] was the earliest attempt. They suggested segmenting the input image into small patches. From these image patches, they learned an over-completed dictionary. Then, the sparse coding features obtained from the learned dictionary constituted the feature space of the input image. The LRMR was performed in this feature space to generate the sparse matrix, and then it was used to produce the final saliency map. Peng et al. [1] pointed out that most LRMR-based methods did not take into consideration the inter-correlation between salient elements. Therefore, they proposed a *structured matrix decomposition* (SMD) model for saliency detection. The SMD model leverages a tree-structure during the LRMR decomposition process to smooth the detected saliency region and capture spatial relations between salient elements. However, Zheng et al. [17] mentioned that the deep structure of the tree in SMD may destroy the sparsity of LRMR decomposition. They proposed to maintain the completeness of the salient object through a coarse-to-fine integration process. Moreover, the work in [18] aimed to improve the salient performance by adding global prior as the weight of LRMR during matrix decomposition process. In summary, the salient object detection methods refer to LRMR focusing on one key objective: Properly modeling image features to separate the low-rank background from the sparse salient objects, while sustaining the completeness of the salient object. In this article, we propose to use the low-rank and sparse tensor method to achieve this objective.

A tensor is a multidimensional array [19,20]. We used bold capital fonts to represent the matrix, while a boldface script letter, e.g., \mathbf{X} , was used to denote a tensor. A vector is a first-order tensor, and a matrix is a second-order tensor. Here, the order of a tensor is the number of dimensions. Similar to matrix rows and columns, a third-order tensor has column, row, and tube fibers. A tensor can reorder its elements to form a matrix, i.e., unfolding. The mode- n unfolding of a tensor \mathbf{X} is denoted by $\mathbf{X}_{(n)}$ and arranges the mode- n fibers to be the columns of the resulting matrix [19]. For instance, a $3 \times 4 \times 5$ tensor can be arranged as a 3×20 matrix, or a 4×15 matrix, or a 5×12 matrix. For a more complete knowledge of tensor and tensor decomposition, please refer to [19,20]. Due to high-dimensional data representation and analysis capabilities, tensor and tensor decomposition have been widely

adopted in various research fields. For example, Jia et al. [21] proposed to use low-rank tensor decomposition for recognizing RGB-D action. They used the tensor Tucker decomposition to obtain tensor ranks. The obtained tensor ranks can be used to discover tensor subspace to extract useful information of RGB-D action. The work in [22] emphasized tensor-based decomposition techniques for detecting abnormalities and failures in different applications, such as environmental monitoring, video surveillance, social networks, etc. In these applications, different spatial and time series data are represented by tensor, and different techniques, such as tensor classifiers, tensor forecasting, and tensor decomposition, to consider the mutual relationship between different dimensions in the tensor. Chen and Mitra [23] proposed to use tensor Tucker decomposition-based technique to locate the source from multimodal data. Specifically, they used rank-1 tensor resulted from Tucker decomposition to compute the common signature vectors. Then, the peak locations of the signature vectors represent the coordinate of the source location. In this work, we attempted to use tensor solving salient object detection task by leveraging its high-dimensional representation ability.

We extend the matrix based low-rank and sparse decomposition into high-dimensional tensor version and apply it to the salient object detection task. The motivations for our work are: First, the high-dimensional tensor is able to express image features in their natural way. As a result, the spatial information of the image is well maintained. Second, the proposed low-rank and sparse tensor decomposition discovers useful information from high-dimensional feature analysis. Therefore, it achieves better performance in discovering salient objects than its LRMR based counterpart methods.

3. Proposed Tensor Decomposition Method for Saliency Detection

For image detection tasks, a rich image feature space is often required to capture high object variations to improve detection performance. In order to adapt LRMR to the salient object detection task, existing methods convert high dimensional feature space into 2D to accomplish model fitting process, while at the expense of spatial consistency. To sustain both the effectiveness of LRMR and the spatial consistency, we propose low-rank and sparse tensor decomposition method.

Based on Equation (1), low-rank and sparse tensor decomposition can be written as:

$$\min_{\mathbf{L}, \mathbf{S}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \quad s.t. \quad \mathbf{F} = \mathbf{L} + \mathbf{S} \quad (2)$$

Accordingly, $\mathbf{F} \in \mathfrak{R}^{M \times N \times D}$, $\mathbf{L} \in \mathfrak{R}^{M \times N \times D}$, $\mathbf{S} \in \mathfrak{R}^{M \times N \times D}$ are the input *three-dimensional* (3D) image or feature tensor, low-rank tensor, and sparse tensor, respectively.

However, unlike matrices, computing the rank of a tensor with the order greater than two is a non-deterministic polynomial-time (NP) hard problem, because the definition of tensor rank depends on the choice of the field [24]. Therefore, most works used heuristic-based strategies, such as in [21,23]. In this work, to solve Equation (2), we also employ a heuristic-based method by redefining the trace norm and l_1 -norm for the high-dimensional tensor case with the order greater than three. Based on the relationship between high-dimensional tensor and matrix, we use the following tensor trace norm [25] for a n -order tensor \mathbf{X} :

$$\|\mathbf{X}\|_* = \sum_{i=1}^3 \alpha_i \|\mathbf{X}_{(i)}\|_* \quad (3)$$

To simplify the computation here in this article and to be consistent with the above trace norm, we propose the following tensor l_1 -norm:

$$\|\mathbf{X}\|_1 = \sum_{i=1}^3 \beta_i \|\mathbf{X}_{(i)}\|_1 \quad (4)$$

where α_i 's and β_i 's are constants satisfying $\alpha_i, \beta_i \geq 0$ and $\sum_i \alpha_i = 1 = \sum_i \beta_i$. In essence, the trace norm (or l_1 -norm) of a high-dimensional tensor is a linear combination of the trace norms (or l_1 -norms) of all

matrices unfolded along each mode. Note that, the l_1 -norm definition in this article is different from the general norm definition of a tensor in [19].

Therefore, with these definitions, the optimization in Equation (2) can be written as:

$$\min_{\mathbf{L}, \mathbf{S}} \sum_{i=1}^3 \alpha_i (\|\mathbf{L}_{(i)}\|_* + \lambda \|\mathbf{S}_{(i)}\|_1) \quad s.t. \quad \mathbf{F} = \mathbf{L} + \mathbf{S} \quad (5)$$

Equation (5) with known tensor \mathbf{F} aims at finding the minimal summation value of n linear equations. To solve this, we replace it with n -sub-problem and find minimal values of each sub-problem, and then we summarize n -minimal values together to achieve the final result of Equation (5), with the aid of tensor unfolding.

$$\min_{\mathbf{L}, \mathbf{S}} \sum_{i=1}^3 \alpha_i (\|\mathbf{L}_{(i)}\|_* + \lambda \|\mathbf{S}_{(i)}\|_1) \quad s.t. \quad \sum_{i=1}^3 \alpha_i (\mathbf{L}_{(i)} + \mathbf{S}_{(i)}) = \mathbf{F}_{(i)} \quad (6)$$

To solve the non-convex sub-problem of Equation (6), the robust principal component analysis (RPCA) method [26] can be properly employed. That is, we unfolding the obtained image feature tensor $\mathbf{F} \in \mathfrak{R}^{M \times N \times D}$ into three matrices, $\mathbf{F}_{(1)} \in \mathfrak{R}^{MN \times D}$, $\mathbf{F}_{(2)} \in \mathfrak{R}^{M \times ND}$, $\mathbf{F}_{(3)} \in \mathfrak{R}^{N \times DM}$, after optimization using the RPCA method, we obtain three matrices $\mathbf{S}_{(1)}$, $\mathbf{S}_{(2)}$, $\mathbf{S}_{(3)}$ of the tensor \mathbf{S} . Therefore, our tensor decomposition turns into three LRMR optimization problems:

$$\min_{\mathbf{L}_{(i)}, \mathbf{S}_{(i)}} \|\mathbf{L}_{(i)}\|_* + \lambda \|\mathbf{S}_{(i)}\|_1 \quad s.t. \quad \mathbf{F}_{(i)} = \mathbf{L}_{(i)} + \mathbf{S}_{(i)} \quad i = 1, 2, 3 \quad (7)$$

In this article, this heuristic method is called *tensor decomposition original* (TDO). In this case, the basic LRMR model becomes the second sub-problem of our case. LRMR is the second unfolding situation of our tensor decomposition. After we obtain three sparse matrices $\mathbf{S}_{(1)}$, $\mathbf{S}_{(2)}$, $\mathbf{S}_{(3)}$, we summarize them according to weights $\alpha_1, \alpha_2, \alpha_3$ to obtain final saliency tensor \mathbf{S} . Finally, the l_1 -norm along each tube fibers, i.e., \mathbf{S}_{ij} , of tensor \mathbf{S} is used to measure the saliency of corresponding segmented region in the i -th row and j -th column of the superpixels. The final saliency map is then accordingly generated and normalized to a 0 to 1 scale image.

Following this we can then adapt Equation (5) to a salient object detection task. The overview of our framework is shown in Figure 1. Given an input image $\mathbf{I} \in \mathfrak{R}^{P \times Q \times 3}$, its low-level features including color, steerable pyramids [27], and Gabor filter [28], are calculated, forming a $P \times Q \times D$ feature tensor (in this article $D = 53$). The Gabor filter feature captures the edge, texture information of the input image. These low-level features can better represent the input image than raw pixel values. Note, here we use the same low-level features as in [1] for a fair comparison in Section 4. The feature tensor is segmented into $M \times N$ small non-overlapping regions, i.e., superpixels, using the *simple linear iterative clustering* (SLIC) method [29] to reduce the computational cost. Each small region is assigned with a D -dimensional feature vector which equals to the mean feature values over all the pixels within the region. Finally, all regions, together with their feature vectors, constitute the input tensor $\mathbf{F} \in \mathfrak{R}^{M \times N \times D}$. Then we use Equation (6) based heuristic procedures to solve this image feature-based tensor decomposition.

Similar to [1], we also integrated high-level priors into our tensor decomposition process, in order to guide the decomposition slightly incline to the most empirically potential saliency position of an image. Three types of priors including location, color, and background priors are used to generate a high-level prior weight map. Empirically, the probability of a salient object in the center of the image is higher than the boundary of the image. The location prior models this observation through a Gaussian distribution based on the distances of the pixels from the image center. Since the human eye is sensitive to red and yellow color, the color prior [10] tends to pay more attention to these color regions in an image. The background prior [30] treats image boundaries as background with high probability. In addition, the high-level prior is employed to eliminate noise saliency regions for the final saliency map.

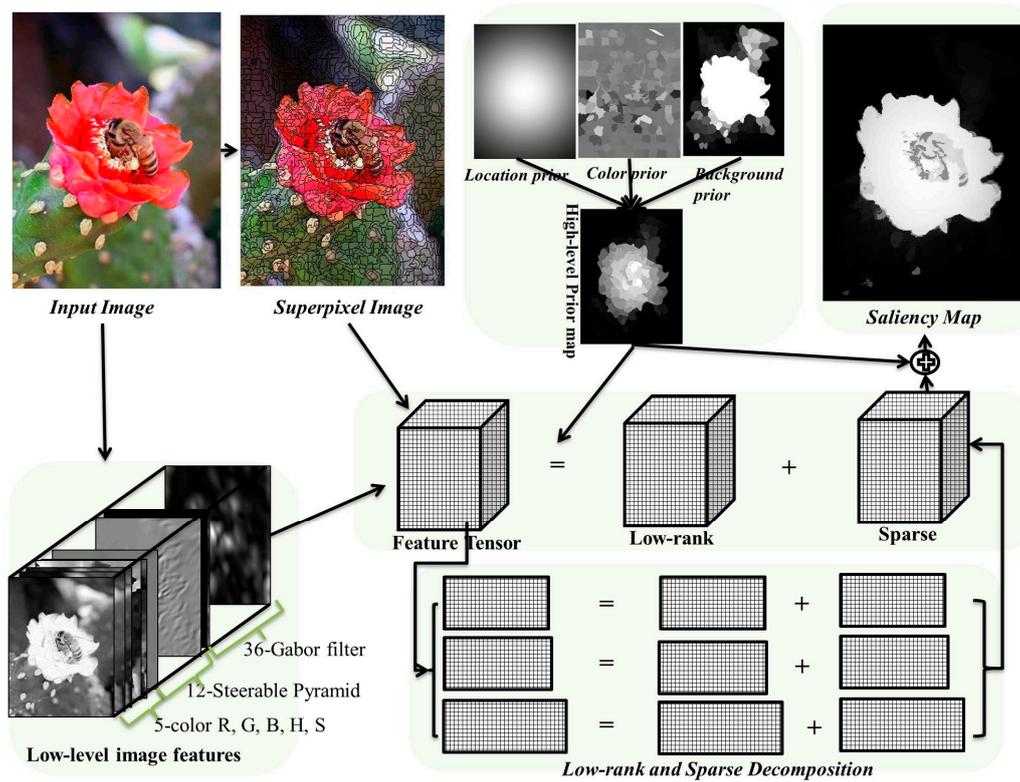


Figure 1. Framework of the tensor decomposition based salient object detection.

4. Experimental Results

To verify the effectiveness of the proposed method, we conducted experiments on three popular benchmark salient object datasets. In this article, each input image is segmented into regions containing around 10 pixels by SLIC. All experiments were conducted in MATLAB using a PC of 4.00 GHz CPU and 12.0 GB RAM. The proposed salient object detection method was tested over all of the datasets using the same settings.

4.1. Benchmark Datasets and Evaluation Criteria

Three popular benchmark saliency detection datasets were used to test the detection accuracy of the proposed method. MSRA10K dataset [8] which is widely used in salient object detection and segmentation community contains 10,000 images with pixel-level saliency masks. iCoSeg [31] dataset contains 643 images with multiple objects in each image. Extended complex scene saliency dataset (ECSSD) [32] contains 1000 structurally complex natural images.

We evaluated the performance using the area under the RO curve (AUC), mean absolute error (MAE), overlapping ratio (OR), and the F_β -measure. Specifically, MAE was used to measure the average pixel-wise absolute difference between detected saliency map and corresponding ground-truth saliency map. The OR was used to measure the importance of complete detection of salient objects. The F_β -measure is defined as

$$F_\beta = (1 + \beta^2) \cdot P \cdot R / (\beta^2 \cdot P + R) \quad (8)$$

where $\beta^2 = 0.3$ is to ensure more stress on precision than recall [33]. The weighted F_β -measure (WF) was also used to measure the importance of saliency detection performance. Precision (P) equals the percentage of salient pixel correctly assigned. Recall (R) is the ratio between the correctly detected salient pixels and true saliency pixel in the ground-truth saliency map.

4.2. Efficient of the Heuristic Tensor Decomposition Process

The proposed tensor model was divided into three unfolding matrices for optimization, to separately obtain three saliency maps. We list the individual sub-problem saliency maps in Figure 2d to f and the final saliency map after integrating solutions of these sub-problems into one tensor in Figure 2b.



Figure 2. Saliency map comparison. (a) Image; (b) tensor decomposition original (TDO) saliency map (with sub-problem solutions integrated); (c) ground-truth saliency map; (d) sub-problem 1 saliency map; (e) sub-problem 2 saliency map; (f) sub-problem 3 saliency map.

4.3. Saliency Detection Performance Comparison

We compare our proposed method (TDS and TDO with/without high-level prior as mask for the final saliency map, respectively) with six existing methods. The results are summarized in Tables 1–3 and in Figures 3 and 4. The six methods include three LRMR based methods, SMD [1], WLRR [18], SLR [34] and a further three popular methods, DRFI [4], RBD [30], HCT [35]. Note that, in Tables 1–3, the up-arrow indicates the larger value the better performance, the down-arrow indicates the smaller value the better performance. The best three results are highlighted with red, green, and blue colors, respectively.

Table 1. Result on iCoSeg dataset in terms of weighted F_{β} -measure (WF), mean absolute error (MAE), overlapping ration (OR), and area under the RO curve (AUC).

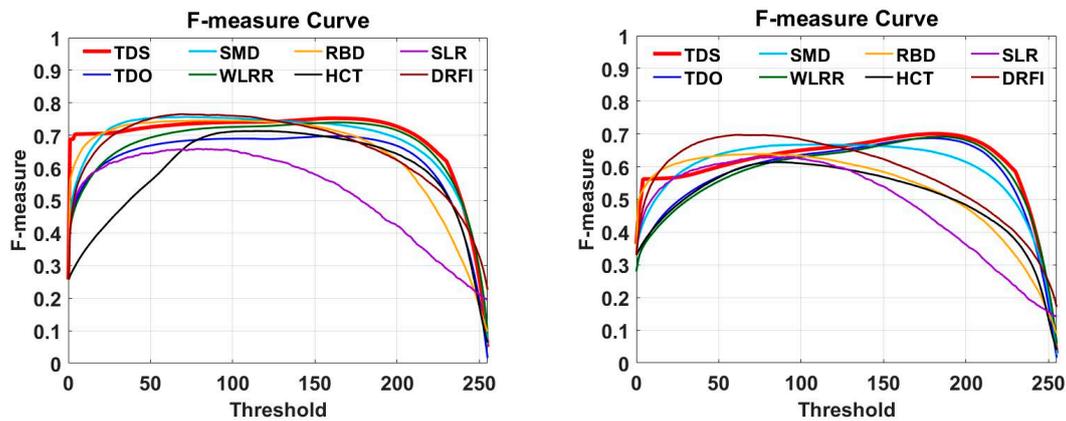
Metric	TDS	TDO	SMD	WLRR	SLR	DRFI	RBD	HCT
WF↑	0.641	0.568	0.611	0.602	0.473	0.592	0.599	0.464
OR↑	0.601	0.564	0.598	0.578	0.505	0.582	0.588	0.519
AUC↑	0.839	0.836	0.822	0.843	0.805	0.839	0.827	0.833
MAE↓	0.134	0.162	0.138	0.147	0.179	0.139	0.138	0.179

Table 2. Result on extended complex scene saliency dataset (ECSSD) dataset in terms of WF, MAE, OR, and AUC.

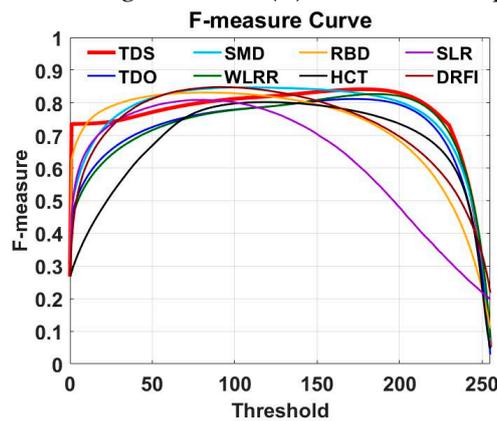
Metric	TDS	TDO	SMD	WLRR	SLR	DRFI	RBD	HCT
WF↑	0.551	0.502	0.517	0.500	0.442	0.517	0.490	0.430
OR↑	0.520	0.506	0.523	0.498	0.474	0.527	0.494	0.457
AUC↑	0.815	0.818	0.775	0.819	0.764	0.780	0.752	0.755
MAE↓	0.185	0.208	0.227	0.211	0.252	0.217	0.225	0.249

Table 3. Result on MASRA10K dataset in terms of WF, MAE, OR, and AUC.

Metric	TDS	TDO	SMD	WLRR	SLR	DRFI	RBD	HCT
WF↑	0.713	0.652	0.704	0.659	0.601	0.666	0.685	0.582
OR↑	0.706	0.690	0.741	0.696	0.692	0.723	0.716	0.674
AUC↑	0.849	0.849	0.847	0.853	0.840	0.857	0.834	0.847
MAE↓	0.108	0.131	0.104	0.130	0.141	0.114	0.108	0.143



(a) F-measure comparison on iCoSeg dataset. (b) F-measure comparison on ECSSD dataset.



(c) F-measure comparison on MASRA10K dataset.

Figure 3. Quantitative comparison on three datasets in terms of F-measure curve. (a) F-measure comparison on iCoSeg dataset; (b) F-measure comparison on ECSSD dataset; (c) F-measure comparison on MASRA10K dataset.

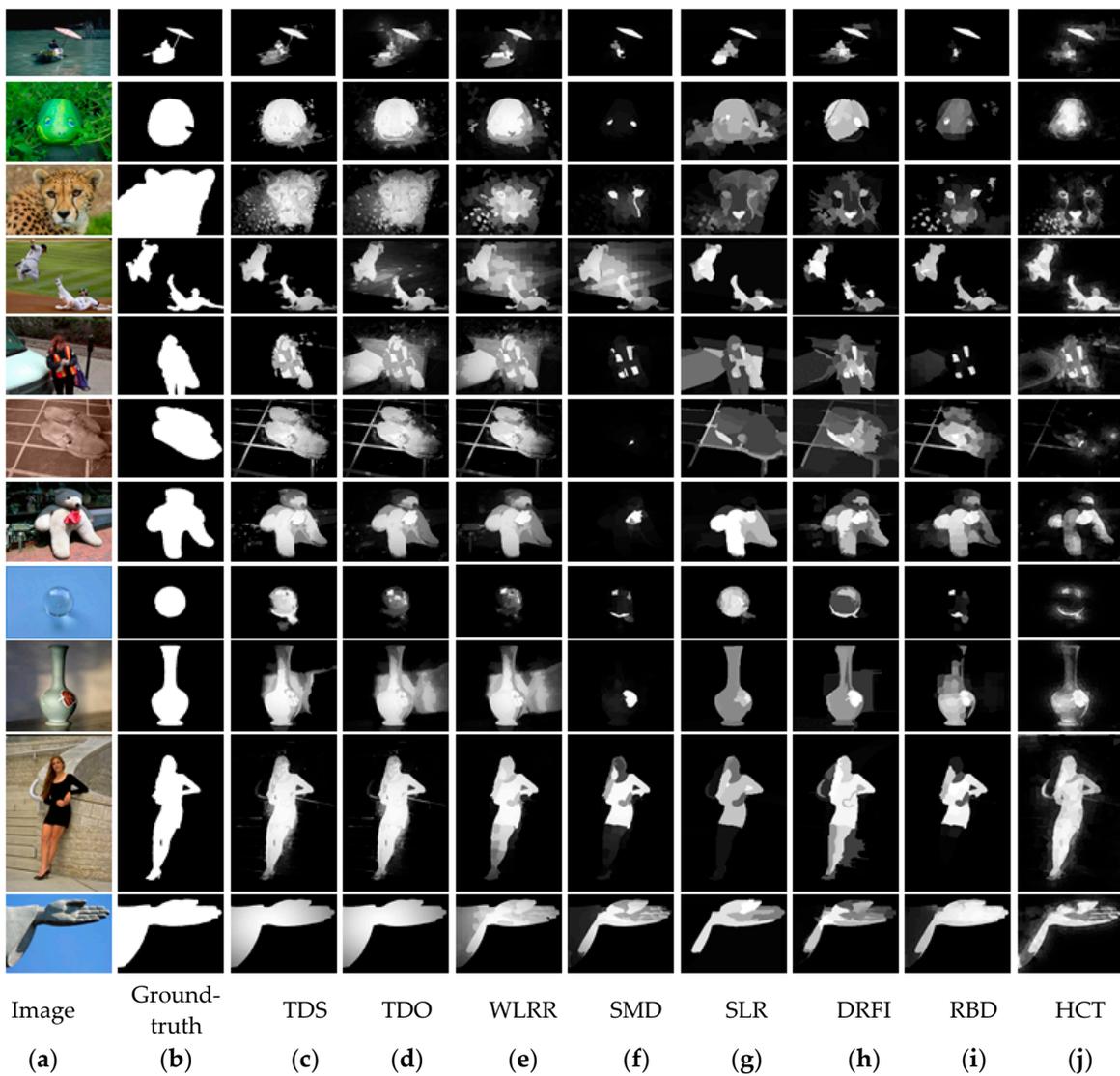


Figure 4. Visual comparison of saliency maps of different methods. (a) Image; (b) ground-truth; (c) TDS; (d) TDO; (e) WLRR; (f) SMD; (g) SLR; (h) DRFI; (i) RBD; (j) HCT.

5. Discussion

In this section, we discuss the results shown in Section 4 from the perspectives of our research motivations and future work.

In this article, we proposed a tensor decomposition method to detect the salient object from images. The task was formulated as a tensor low-rank and sparse optimization problem. We employed a heuristic method to convert the optimization problem into three sub-problems.

- The first motivation of this article is that we use a tensor to represent image features in a natural high-dimensional fashion. The tensor ensures that more dimensions can be introduced to represent information; meaning that the spatial information is maintained.

One of the drawbacks of LRMR based methods in saliency detection is that it may destroy the spatial consistency when converting high-dimensional image features into 2D matrix form. It is because the spatial adjacent image regions are separated in the feature space. The SMD method addressed this drawback by introducing a tree-structure to group spatial adjacent image regions together during the optimization process. However, in this article, we used a tensor to represent the image feature to maintain the spatial position of image regions. Although the heuristic optimization may change this

3D structure during the individual optimization process, the final step of converting the result of the optimized sub-problems into a single tensor will solve the spatial connection problem.

As can be observed from Tables 1–3 and Figure 3, the original tensor decomposition method TDO achieved a much higher performance than the LRMR-method SLR, which did not consider spatial information. However, it achieved a slightly worse performance than SMD and WLRR with their additional tree-structure and weighted prior information, respectively. The reason for this is that TDO contains more noise information during the tensor decomposition process. Therefore, we used the high-level prior information to generate a mask to reduce the saliency value in high probability background area to obtain the final saliency map, i.e., TDS. As a result, the MAE was reduced in TDS. TDS achieved slightly higher or similar performance than SMD and WLRR in terms of the metrics. This means the proposed tensor method has the ability to sustain the spatial structure.

- The second motivation for this study is that tensor decomposition is capable of discovering more saliency information.

The image feature information was identical for the LRMR-based method and TDS for the salient object detection task. The only difference was the way of representing the image feature. With TDS, the image feature was represented in 3D tensor. Our task was to employ a heuristic method to solve a tensor decomposition problem. This heuristic method converted the tensor into sub-matrices and optimized them independently. Following this, the final tensor decomposition result was obtained by combining three matrices into a unified tensor. Conventional LRMR was just one of our sub-optimization processes.

As shown in Figure 2, each of the three sub-problems extracted saliency regions of the image (Figure 2d–f). The tensor, which is obtained by integrating the three results from sub-problems, yielded the best saliency map (Figure 2b). Therefore, the sub-problems complement each other.

In Figure 4, we visualize some of the saliency map detected by the proposed method (in columns (c,d) and other methods. It can be observed that, firstly, with the high-prior map, TDS obtained cleaner saliency maps than TDO, as shown in rows 5 and 9. This verifies that TDS achieved better numerical performance than TDO, as shown in Tables 1–3. Second, compared to the LRMR-based methods SMD, SLR, TDS, and WLRR, TDS was capable of extracting more salient object information as shown in rows 3, 4, 8, and 11. In addition, saliency maps in column (c) were closer to ground-truth saliency maps in column (b) than other methods like DFI, RBD, and HCT, indicating the effectiveness of the proposed method.

Therefore, we can conclude that tensor decomposition yields more saliency information than LRMR.

- Future work

In this work, we proved that by using a tensor we can sustain spatial information and discover more information related to the saliency area in an image. Therefore, the results are mainly compared with LRMR-based methods. Improving the overall salient detection rate will be the basis of our future work. For example, we can use deep learning-based features, instead of 53-D low-level features in our framework, because deep learning-based features may have better representation ability for salient objects than low-level image features.

In addition, although we propose to use high-dimensional low-rank and sparse tensor decomposition to represent the salient detection task, the tensor decomposition is still converted to the low-rank and sparse matrix decomposition sub-problems during the optimization process. This may affect the optimization result of tensor decomposition. In the future, we will optimize the tensor decomposition from high dimensional aspects, such as using tensor Tucker decomposition, in order to improve the accuracy of the detection result.

6. Conclusions

The focus of this article is on low-rank and sparse tensor decomposition and its application in the salient object detection task. In this article, we used a tensor to represent image features in 3D space. The decomposition based on a tensor, characterized image features from multiple aspects. Therefore, for the salient object detection task, the tensor decomposition method was able to maintain spatial information and discover more saliency information. The noise region in the tensor decomposition result was eliminated by applying a simple high-level prior mask. As a result, the tensor decomposition based salient object detection obtained a better performance than its LRMR-based matrix counterparts on three benchmark saliency datasets. The experimental results also showed that tensor decomposition performed better than a supervised DRFI method.

Author Contributions: Conceptualization, visualization, J.Z.; methodology, software, validation, writing—original draft preparation, J.Z., Y.T.; writing—review and editing, J.Z., Y.T., X.L.; formal analysis, investigation, supervision, X.L.

Funding: This research received no external funding.

Acknowledgments: The authors wish to thank anonymous reviewers for valuable comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Peng, H.; Li, B.; Ling, H.; Hu, W.; Xiong, W.; Maybank, S.J. Salient object detection via structured matrix decomposition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 818–832. [[CrossRef](#)]
2. Liu, T.; Yuan, Z.; Sun, J.; Wang, J.; Zheng, N.; Tang, X.; Shum, H.-Y. Learning to detect a salient object. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 353–367.
3. Mai, L.; Niu, Y.; Liu, F. Saliency aggregation: A data-driven approach. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 1131–1138.
4. Jiang, H.; Wang, J.; Yuan, Z.; Wu, Y.; Zheng, N.; Li, S. Salient object detection: A discriminative regional feature integration approach. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 2083–2090.
5. Zhao, R.; Ouyang, W.; Li, H.; Wang, X. Saliency detection by multi-context deep learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1265–1274.
6. Li, G.; Yu, Y. Deep contrast learning for salient object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 478–487.
7. Hou, Q.; Cheng, M.-M.; Hu, X.; Borji, A.; Tu, Z.; Torr, P.H. Deeply supervised salient object detection with short connections. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3203–3212.
8. Cheng, M.-M.; Mitra, N.J.; Huang, X.; Torr, P.H.; Hu, S.-M. Global contrast based salient region detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 569–582. [[CrossRef](#)]
9. Perazzi, F.; Krähenbühl, P.; Pritch, Y.; Hornung, A. Saliency filters: Contrast based filtering for salient region detection. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 733–740.
10. Shen, X.; Wu, Y. A unified approach to salient object detection via low rank matrix recovery. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 853–860.
11. Wang, W.; Wang, Y.; Huang, Q.; Gao, W. Measuring visual saliency by site entropy rate. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2368–2375.
12. Guo, Y.; Wang, F.; Liu, P.; Xin, J.; Zheng, N. Multi-scale point set saliency detection based on site entropy rate. In Proceedings of the Pacific Rim Conference on Multimedia, Xi'an, China, 15–16 September 2016; pp. 366–375.
13. Zhang, L.; Ai, J.; Jiang, B.; Lu, H.; Li, X. Saliency detection via absorbing markov chain with learnt transition probability. *IEEE Trans. Image Process.* **2018**, *27*, 987–998. [[CrossRef](#)]

14. Ishikura, K.; Kurita, N.; Chandler, D.M.; Ohashi, G. Saliency detection based on multiscale extrema of local perceptual color differences. *IEEE Trans. Image Process.* **2018**, *27*, 703–717. [[CrossRef](#)] [[PubMed](#)]
15. Candès, E.J.; Li, X.; Ma, Y.; Wright, J. Robust principal component analysis? *J. ACM* **2011**, *58*, 11. [[CrossRef](#)]
16. Yan, J.; Zhu, M.; Liu, H.; Liu, Y. Visual saliency detection via sparsity pursuit. *IEEE Signal Process. Lett.* **2010**, *17*, 739–742. [[CrossRef](#)]
17. Zheng, Q.; Yu, S.; You, X.; Peng, Q.; Yuan, W. Coarse-to-fine salient object detection with low-rank matrix recovery. *arXiv* **2018**, arXiv:1805.07936.
18. Tang, C.; Wang, P.; Zhang, C.; Li, W. Salient object detection via weighted low rank matrix recovery. *IEEE Signal Process. Lett.* **2017**, *24*, 490–494. [[CrossRef](#)]
19. Kolda, T.G.; Bader, B.W. Tensor decompositions and applications. *SIAM Rev.* **2009**, *51*, 455–500. [[CrossRef](#)]
20. Sidiropoulos, N.D.; De Lathauwer, L.; Fu, X.; Huang, K.; Papalexakis, E.E.; Faloutsos, C. Tensor decomposition for signal processing and machine learning. *IEEE Signal Process.* **2017**, *65*, 3551–3582. [[CrossRef](#)]
21. Jia, C.; Fu, Y. Low-rank tensor subspace learning for RGB-D action recognition. *IEEE Trans. Image Process.* **2016**, *25*, 4641–4652. [[CrossRef](#)] [[PubMed](#)]
22. Fanaee-T., H.; Gama, J. Tensor-based anomaly detection: An interdisciplinary survey. *Knowl.-Based Syst.* **2016**, *98*, 130–147. [[CrossRef](#)]
23. Chen, J.; Mitra, U. A tensor decomposition technique for source localization from multimodal data. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 4074–4078.
24. Hillar, C.J.; Lim, L.-H. Most tensor problems are NP-hard. *J. ACM* **2013**, *60*, 45. [[CrossRef](#)]
25. Liu, J.; Musialski, P.; Wonka, P.; Ye, J. Tensor completion for estimating missing values in visual data. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 208–220. [[CrossRef](#)] [[PubMed](#)]
26. Wright, J.; Ganesh, A.; Rao, S.; Peng, Y.; Ma, Y. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 6 December 2010; pp. 2080–2088.
27. Simoncelli, E.P.; Freeman, W.T. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In Proceedings of the International Conference on Image Processing, Washington, DC, USA, 23–26 October 1995; pp. 444–447.
28. Feichtinger, H.G.; Strohmer, T. *Gabor Analysis and Algorithms: Theory and Applications*; Springer Science and Business Media: New York, NY, USA, 2012.
29. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [[CrossRef](#)] [[PubMed](#)]
30. Zhu, W.; Liang, S.; Wei, Y.; Sun, J. Saliency optimization from robust background detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 2814–2821.
31. Batra, D.; Kowdle, A.; Parikh, D.; Luo, J.; Chen, T. Interactively co-segmenting topically related images with intelligent scribble guidance. *Int. J. Comput. Vis.* **2011**, *93*, 273–292. [[CrossRef](#)]
32. Yan, Q.; Xu, L.; Shi, J.; Jia, J. Hierarchical saliency detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 1155–1162.
33. Achanta, R.; Hemami, S.; Estrada, F.; Süsstrunk, S. Frequency-tuned salient region detection. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2009), Miami, FL, USA, 20–25 June 2009; pp. 1597–1604.
34. Zou, W.; Kpalma, K.; Liu, Z.; Ronsin, J. Segmentation driven low-rank matrix recovery for saliency detection. In Proceedings of the 24th British Machine Vision Conference (BMVC), Bristol, UK, 9 September 2013; pp. 1–13.
35. Kim, J.; Han, D.; Tai, Y.-W.; Kim, J. Salient region detection via high-dimensional color transform. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 883–890.

