Visual Salience Learning via Low Rank Matrix Recovery

Junxia Li, Jundi Ding, Jian Yang

School of Computer Science and Engineering, Nanjing University of Science and Technology

Abstract. Detection of salient object regions is useful for many vision tasks. Recently, a variety of saliency detection models have been proposed. They often behave differently over an individual image, and these saliency detection results often complement each other. To make full use of the advantages of the existing saliency detection methods, in this paper, we propose a salience learning model which combines various saliency detection methods such that the aggregation result outperforms each individual one. In our model, we first obtain several saliency maps by different saliency detection methods. The background regions of each saliency map usually lie in a low-dimensional subspace as most of them tend to have lower salience values, while the object regions that deviating from this subspace can be considered as sparse noises. So, an individual saliency map can be represented as a low rank matrix plus a sparse matrix. We aim at learning a unified sparse matrix that represents the salient regions using these obtained individual saliency maps. The sparse matrix can be inferred by conducting low rank matrix recovery using the robust principal component analysis technique. Experiments show that our model consistently outperforms each individual saliency detection approach and state-of-the-art methods.

1 Introduction

Humans have a remarkable ability to effortlessly judge the importance of image pixels or regions in real time, and pay more attention to those important and informative parts. Detection of such salient pixels or regions of an image automatically is an active research area in recent decades. There are two major research directions of visual attention modeling, including eye fixation prediction and salient object detection. The former is to identify a few human fixation locations on natural images, which is helpful for many high-level vision tasks. The later focuses on uniformly highlighting entire salient object regions, thus benefiting wide applications in computer vision like: salient object segmentation [1, 3, 2], object based image retrieval [4], content-aware image resizing [5], automatic image cropping [6], and adaptive image compression [7]. In this paper, we focus on the salient object detection.

Although a rich literature has been appeared on image saliency analysis [8–13, 15–20], a few commonly noticeable and critically influencing issues still endure.

They are related to complexity of patterns and behave differently in natural images, since different saliency models are based on variety of theories and techniques. The saliency maps obtained by different methods vary remarkably from each other. Each of them has its advantages and disadvantages. Fig. 1 shows a few results from several previous representative salient object detection methods. We can clearly see that each method just works well for some parts of an image. As shown in Fig. 1(b-c), some objects boundaries are well-defined, but most of the salient regions do not stand out. Differently, the detection results shown in Fig. 1(d-f) well highlight most of the object regions, but some background regions are also detected as salient regions shown in white. As shown in the 'Board' image of Fig. 1(d), the trees around the board also stand out simultaneously with the salient board regions. More interestingly, it is observed that these results can complement each other in general. This is mainly because each saliency detection method often captures some aspects of the visual information from different perspectives (*e.g.* local/ global contrast, sparsity or spatial distribution). *This motivates us to combine different saliency maps to get better results. Specifically, for a given image, we can first obtain several saliency maps by different saliency detection methods, and then try to find a way to utilize the advantages of these methods and meanwhile suppress the disadvantages of them, aiming to effectively combine these saliency maps.*

(a) Image (b) CA (c) FT (d) RC (e) LR (f) PCAS (g) Ours

Fig. 1. Visual salience learning. Individual salience maps, such as CA[10], FT[9], RC[11], LR[15], PCAS[20] often complement each other. Visual salience learning can effectively combine their results and perform better than each of them.

Although there are several approaches attempting to integrate different saliency maps to detect saliency [12][13], they might not make use of the advantages and disadvantages of these saliency maps for each input image. A. Borji et al. [12] presented a simple combination model for saliency detection using pre-defined functions. It takes each individual model all equal in the integration process. However, this strategy may not fully capture the advantages of each individual saliency detection method. L. Mai et al. [13] proposed an approach for saliency aggregation using a conditional random field framework. Sometimes, this method

can better determine the contribution for each input saliency map in the aggregation, as it considers the performance gaps among different saliency analysis methods. Unfortunately, the cross-information among individual saliency maps is not well utilized in the aggregation process. It is often difficult for such models that are use linear or nonlinear fusion strategy to produce reliable results.

For the saliency maps obtained by various saliency detection methods, if we first segment the original image into many homogeneous sub-regions, the background regions of each saliency map usually lie in a low-dimensional subspace since most of them tend to have the lower salience values shown in black, while the object regions that deviating from this subspace can be represented by a sparse matrix. Inspired by this, we are dedicated to learning a unified sparse matrix that represents the salient regions using these obtained individual saliency maps. And our saliency detection can be cast as a sparse matrix pursuit problem. Our idea differs from the previous saliency detection models which also based on the theory of low-rank-sparsity matrix decomposition [21][22][15] in its motivation. In [21][22][15], the saliency map is inferred by integrating multiple types of features of the given image. However, in their models, for many complex natural images, the assumptions that the background matrix has a low rank and the salient regions correspond to a sparse matrix may not hold.

In our method, we first decompose a given image into many homogeneous sub-regions by an image segmentation technology, each of which is called a super-pixel. Then we conduct various saliency detection methods and obtain the same number of saliency maps. Each saliency map is represented by a vector, where each element of the vector corresponds to the mean of the saliency value of that super-pixel. Arranging these vectors forms a combinational matrix representation of these saliency maps. The sparse matrix indicating the salient regions can be well inferred by conducting low-rank matrix recovery using the robust principal component analysis (RPCA) technique [23]. Since the cross-information among individual saliency maps has been well considered, such a sparsity pursuit scheme can produce more accurate and reliable results than the saliency aggregation models of using simple linear or nonlinear fusing strategy with fixed coefficients [12][13], and also can outperform the performance of each individual saliency detection method.

Compared with existing methods, the contributions of our method mainly include:

- Our proposed approach considers the cross-information among individual saliency detection methods, it performs better than these methods which combine saliency maps through weighted averaging.

- In our proposed approach, the contribution of each saliency map is not equally constant, but learned adaptive to each image.

- Our proposed approach treats the saliency detection as a sparsity pursuit problem based on the theory of low rank matrix recovery. It provides an interesting perspective for visual saliency learning framework.

2 Related Work

To serve as the baseline of our approach, a set of saliency detection methods need to be chosen to produce individual saliency maps. Today, there are many saliency analysis methods based on variety of techniques with interesting performance. In the following we give a review of saliency detection methods that are related to our approach.

The major difference among these saliency detection approaches is the strategy for measure saliency. In recent years, a growing number of saliency detection methods have been proposed. Here, we classify existing saliency object detection models into three categories: contrast based methods, spatial distribution based methods and sparsity based methods. Note that, these three kinds of methods are not completely disjoint, they are interspersed with each other to some extent.

As a pioneer, Itti et al. [24] use center-surround differences across multi-scale image features to detect image saliency. Hereafter, many contrast based models have been proposed to extend this approach, including the fuzzy growth model (MZ) by Ma and Zhang [25], and graph based visual saliency model (GB) by Harel et al. [26]. Later, a method presented by Achanta et al. (AC) [8] which determine salient regions using low level features of color and luminance. Achanta et al. [9] implemented a frequency-tuned method (FT) to define pixel saliency based on the color difference from the average color of entire image. Recently, Margolin et al. [20] use PCA (PCAS) to represent the set of patches of an image and use this representation to determine patch distinctness. Cheng et al. [11] consider the global region contrast differences with respect to the whole image and spatial coherence across the regions to define saliency map (RC). However, since spatial distribution among patches is not formulated, RC cannot well handle images with cluttered and textured backgrounds. To deal with the images which contain small-scale structures, Yan et al. [17] presented a hierarchical saliency model (HS) that infers saliency values from multiple image layers. These contrast based methods have their difficulty in distinguishing among similar saliency cues (e.g. color, structure) in both foreground and background regions. Besides, they generally fail when the images are with large-scale objects.

Spatial distribution based methods are generally built on two common priors which come from the basic rule of photographic composition. The first one is the object prior which considers that salient regions are likely to appear at the center of an image. The second one is the background prior which assumes that the image boundary is mostly background. Based on these two priors there are many saliency detection models have been presented, e.g, graph-based manifold ranking model [33], absorbing markov chain model [18] and dense and sparse reconstruction model [16]. Spatial distribution based models have achieved success in many images, but still have certain limitations. Typically, if the assumption of the object prior or background prior is not hold, it nevertheless provides useful visual information which can be utilized to guide the salience detection.

Sparsity based models are performed under the assumption that in a certain feature space the salient region is sparse compared with the background regions. SR [27] proposed by Hou and Zhang is a typical sparsity based model

which measures saliency via spectral residual in the frequency domain. Later, several low-rank-sparsity matrix decomposition theory based models have been proposed, which infer saliency map by integrating multiple types of features of the given image [21, 22, 15, 19].

Although still each year many new saliency models are introduced, there is a large gap between models and human performance in detecting salient object regions in free-viewing of natural scenes. It is nice to see that these models often vary over an individual image and complement each other. In our model, we choose newly published methods HS [17], PCAS [20], GC [32], DS [16], AMC [18], GBMR [33] and SLR [19] to produce individual saliency maps for our saliency aggregation. Note that as more and more saliency detection models have been developed recently, more existing and forthcoming saliency models can also used in our individual initial saliency maps production.

It should be noted that the idea of employing the low-rank matrix recovery for saliency detection is not new [22][15]. In [22], an image is partitioned into non-overlapping patches of size $p \times q$ pixels, each of which is represented by a feature vector. These feature vectors are then arranged to form a multiple-feature matrix for low rank matrix recovery. However, when the object of the image is not small enough, the noises expected to indicate saliency will no longer be sparse. This violates the underlying assumption of the model. Different from this method, the approach proposed in [15] represents the image in another way. It incorporated image segmentation into the saliency detection and partitioned an image into many small regions after multi-scale feature extraction. To ensure the validity of the low rank recovery model, they modulated the image features with a learnt transform matrix. However, the learnt transform matrix is somewhat biased toward the training data set, it suffers from limited adaptability.

Differing from these approaches, we select various saliency maps to form a matrix for low rank matrix recovery. After the over-segmentation, the input image is divided into many regions. Thus, for each saliency map, the corresponding background regions can be represented by a low rank matrix, and the salient regions can be indicated by a sparse matrix. We use the matrix combined by various saliency maps to conduct the low rank recovery. By this way, our model can adaptively make use of the advantages of individual saliency maps and yield a satisfactory result even without higher-level prior.

3 Salience Learning by Low Rank Affinity Pursuit

Our method starts from running a set of d saliency detection methods on a given image $\mathbf{I}$, and produces d saliency maps, $\{S_k||1 \leq k \leq d\}$, one for each method. Each element $S_k(p)$ in a saliency map is the salience value of the pixel p. In each saliency map, the values of pixels are represented in gray and normalized in the range $[0, 1]$. Our goal is to take these d salience maps as original input and learn a final salience map S. In this section, we describe details of our learning model of salience maps aggregation.

3.1 Problem Formulation

In this subsection, we will describe the formulation of our problem in detail. And Fig.2 gives an illustration for easy understanding of our procedure.

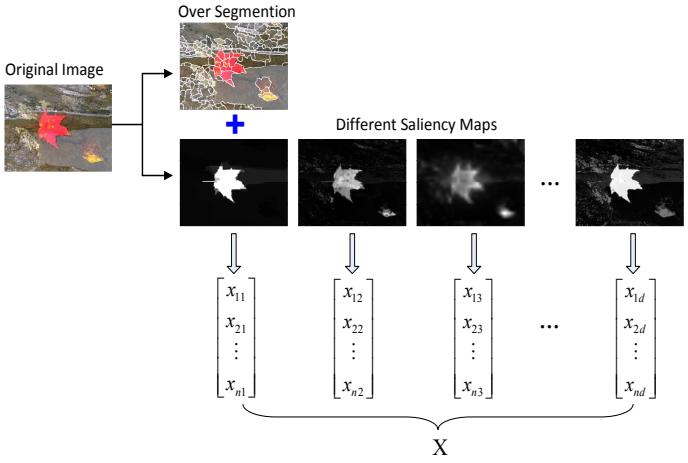


Fig. 2. Illustration of our problem formulation.

In order to get an effective representation for a saliency map, we use super-pixels other than image pixels or patches with the same size as basic image elements. We choose mean-shift clustering [28] to produce our required image segmentation. Each sub-region is called a super-pixel. With suitable parameters selection, the image background also contains multiple super-pixels even if it is visually homogeneous. See the first row of Fig.2 for an example, the original is divided into a set of super-pixels $\{P_i\}_{i=1,\ldots,n}$, where n is the number of super-pixels. Combining the over-segmentation result, saliency map S_k can be represented by a n-dimensional vector $X_k = [x_{1k}, x_{2k}, \ldots, x_{nk}]^T$, the $i-th$ element of the vector corresponds to the mean of the saliency values of super-pixel P_i. By arranging these vectors into a matrix, we get the combinational matrix representation of all the saliency maps $\mathbf{X} = [X_1, X_2, \ldots, X_d]$. $\mathbf{X} \in R^{n \times d}$, where each X_i corresponds to the $i-th$ saliency map, d denotes the number of saliency maps. Then, our goal is to find an assignment function $S(P_i) \in [0, 1]$. Function $S(P_i)$ is referred to as the final saliency map, where the value of $S(P_i)$ represents the probability of super-pixel P_i belonging to the salient objects.

3.2 Saliency Aggregation using Robust PCA

Our task described by the above formulation is to find a criterion for measuring and detecting the final saliency which can effectively utilize the advantages of individual saliency methods and meanwhile suppress the disadvantages of them.

For the previous saliency detection methods, the term "sparsity" is an important perspective for detecting salient objects. It is in essence similar to the contrast (another point of view for saliency), since the pixels or regions different from their surroundings usually receive higher response on contrast-based features (*e.g.*, color or texture). From this perspective, the salient regions are different from the background regions are mostly sparse. Then, an image can be considered as a combination of a background part with a salient part. In other words, the image $\mathbf{I}$ can be decomposed into two parts in a certain feature space:

$$\emptyset(\mathbf{I}) = \mathbf{A} + \mathbf{E}, \tag{1}$$

where $\emptyset$ is a feature transformation, $\mathbf{A}$ and $\mathbf{E}$ denote matrices of the background part and the salient part, respectively. Although most of the existing saliency detection methods are based on the contrast or sparsity criterion, they often perform differently and complement each other (see Fig.1 for example). Therefore, we design a 'powerful alliances' strategy to achieve a satisfactory saliency aggregation.

Since each saliency detection method is a nonlinear transformation from the given image to the saliency map, then $\mathbf{X}$ can be treated as a feature representation of the image $\mathbf{I}$ in the saliency feature space (to distinguish this new feature space from the traditional feature space, *e.g.*, RGB color space, we called the feature space composed by various saliency maps as saliency feature space). Naturally, Eq.1 can be rewritten as:

$$\mathbf{X} = \mathbf{A} + \mathbf{E}, \tag{2}$$

Eq.2 is a severely under-constrained problem. Theoretically speaking, it is almost impossible to find $\mathbf{A}$ and $\mathbf{E}$ without any restrictions information. In other words, without imposing any restrictions to Eq. 2, there are an infinite number of solutions with regard to $\mathbf{A}$ and $\mathbf{E}$. To seek a suitable solution that is benefit for our saliency detection, some criteria for characterizing matrices $\mathbf{A}$ and $\mathbf{E}$ are needed. To this end, we here consider two basic principles. On the one hand, the background regions usually lie in a low dimensional subspace so that they can be represented as a low rank matrix. This suggests that matrix $\mathbf{A}$ may have the property of low rankness. On the other hand, in a saliency map, only a small portion of super-pixels are salient regions even when the object size is large, since salient objects usually have characteristic and spatial coherence. So, we can regard the salient regions that are deviate from the low dimensional subspace as noises or errors, *i.e.*, matrix $\mathbf{E}$ is sparse. The relation between low-rank-sparsity and saliency is consistent with the fact that only the distinctive sensory information is selected for further processing in a human vision system. In summary, we incorporate two criteria to solve the Eq. 2, *i.e.*, the low rank constraint for the background regions and the sparsity constraint for the salient regions. Therefore, the saliency detection can be cast as a sparse matrix recovery problem. Fortunately, the recently established robust principle component analysis (RPCA) technique [23] may fit well to the saliency detection problem. For matrix $\mathbf{X} = [X_1, X_2, \ldots, X_d]$ with each column representing a

corresponding saliency detection method, RPCA is appropriate to efficiently and exactly recover the sparse matrix $\mathbf{E}$ by solving a tractable optimization problem:

$$\begin{aligned} \min_{\mathbf{A},\mathbf{E}} ||\mathbf{A}||_* + \lambda||\mathbf{E}||_1 \\ s.t. \mathbf{X} = \mathbf{A} + \mathbf{E} \end{aligned} \tag{3}$$

where $||\cdot||_*$ denotes the nuclear norm (sum of the singular values of a matrix), $||\cdot||_1$ is the ℓ_1-norm and parameter $\lambda > 0$ balances the effects between rank and sparsity. In our implementation, we set $\lambda = 0.06$. Note that, problem (3) is a convex optimization problem. Recent theoretic analysis indicates that there are various algorithms can be used to recover the sparse matrix $\mathbf{E}$ in high probability [23][29]. Here, we apply the exact ALM method [29] to extract the sparse matrix $\mathbf{E}$.

As the minimization of the ℓ_1-norm encourages the columns of $\mathbf{E}$ to be zero, *i.e.*, the columns of $\mathbf{E}$ are sparse, it fits well to our visual saliency learning problem. Naturally, the sparse matrix $\mathbf{E}$ measures the contribution of each individual saliency method and detects satisfactory visual saliency. For a row corresponding to the $i-th$ super-pixel, larger element implies that the corresponding saliency detection method made a greater contribution in terms of aggregation. That is, in the saliency aggregation process, our model learned a set of combinational coefficients for each super-pixel rather than simply combining the saliency responses through weighted averaging.

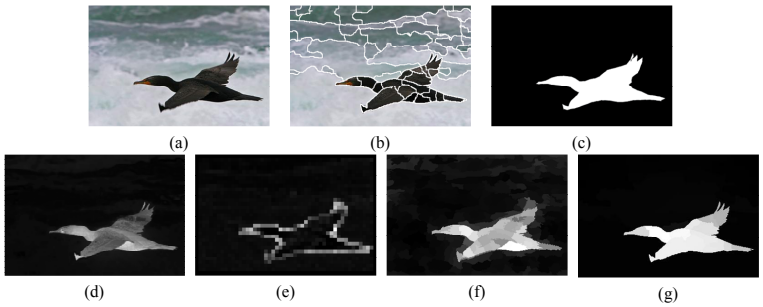


Fig. 3. Illustration on our saliency learning. (a) input image, (b) over-segmentation result by mean-shift, (c) ground truth, (d) saliency aggregation result using pre-defined function [12], (e) sparse coding result [21], which only highlights some edges, (f) detected saliency by LR[15] without high-level prior interaction, better than (e), but has some high saliency values in the background, (g) saliency by our model, which is better than others.

Let $\mathbf{E}^*$ be the optimal solution (with respect to $\mathbf{E}$) to optimization problem (3). To obtain a saliency value for the super-pixel P_i, we only need a following

step to quantify the response of the obtained sparse matrix $\mathbf{E}^*$:

$$S(P_i) = ||\mathbf{E}^*(i,:)||_1 = \sum_{j=1}^{d} |\mathbf{E}^*(i,j)| \tag{4}$$

Here, $||\mathbf{E}^*(i,:)||_1$ is the ℓ_1-norm of the $i-th$ row of $\mathbf{E}^*$. Larger (smaller) magnitude of $S(P_i)$ implies that the super-pixel P_i is more salient (non-salient), then we assign a higher (lower) value to it. In this way, the final saliency map is accordingly generated. Fig. 3(g) presents our saliency aggregation result and shows that our model outperforms others.

4 Experiments

4.1 Experimental Setup

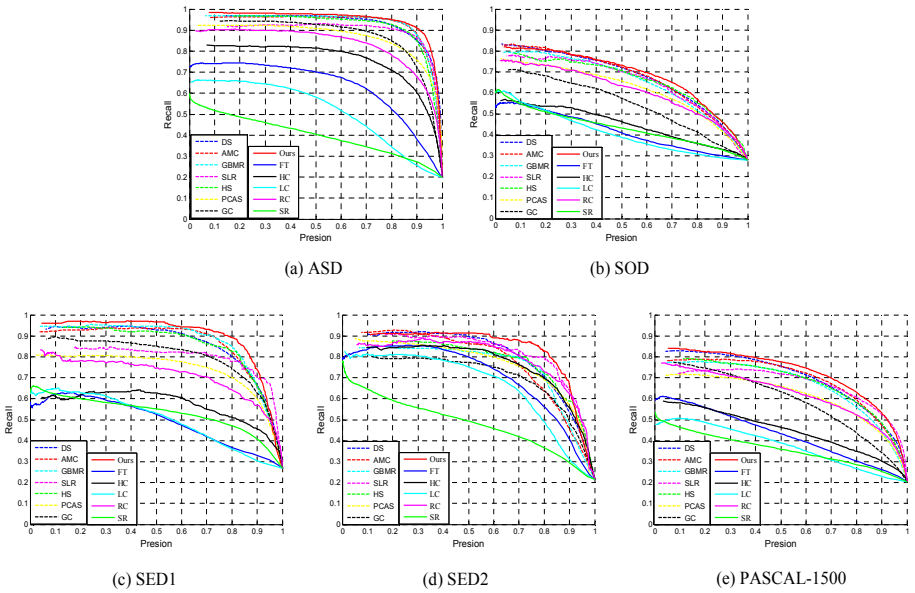
Data Sets and Evaluation Metrics In order to comprehensively evaluate the performance of our proposed method, we conducted extensive experiments on four benchmark datasets. The first one is ASD dataset provided by Achanta et al. [9], which contains 1000 images with accurate human-marked labels for salient objects. The second one is the SED1 dataset [30] with 100 images of a single salient object. The third one is the SED2 dataset [30] which contains 100 images of two salient objects. Both SED1 and SED2 are with ground truths labeled by three different users. The SOD dataset [31] contains 300 images, which is based on Berkeley segmentation dataset. The fifth one is the PASCAL-1500 dataset [19] in which many images contain multiple objects with various locations and scales, and highly cluttered background. This dataset is first used for image segmentation evaluation [34]. Both SOD dataset and PASCAL-1500 dataset are the most challenging ones for saliency detection.

To evaluate the quantitative performance of different methods, in this paper we adopt three metrics including two popular ones precision recall (PR) curve and F-measure, and the newly presented VOC score [35]. Specifically, the PR curve is obtained by binarizing the saliency map according to different fixed thresholds ranging from 0 to 255. At each value of the threshold, the precision and recall are computed. The F-measure is the overall performance measurement computed by the weighted harmonic of precision and recall, which defined as $F = ((\beta^2+1)P*R)/(\beta^2 P + R)$ (P=precision, R=recall). Here, precision and recall are obtained by binarizing the saliency map by an adaptive threshold that is twice the overall mean saliency value of the entire image. We set $\beta^2 = 0.3$ which is the same as in [9][11][15]. The VOC score is defined as $VOC = (S \cap G)/(S \cup G)$, where S is the object segmentation result obtained by binarizing the saliency map using the same adaptive threshold as in the computation of F-measure, and G is the ground-truth.

4.2 Quantitative Evaluations

We first compared the performance of our method with seven used individual saliency detection methods, i.e., DS [16], AMC [18], GBMR [33], SLR [19], HS

[17], PCAS [20] and GC [32]. In addition, we also compared our method with other five classical approaches, including FT [9], HC [11], LC [36] RC [11] and SR [27]. Most of them were presented recently or have a high citation rate. Note that, SLR is the extension of the LR [15] which also uses idea of employing the low-rank matrix recovery for saliency detection, so here we are no longer report the results of LR. And, theoretically, a comparison with the related approach, namely the saliency aggregation method proposed by L. Mai et al. [13] is necessary. However, we could not find the authors' implementation.



(a) ASD (b) SOD

(c) SED1 (d) SED2 (e) PASCAL-1500

Fig. 4. Precision recall curves of all the twelve methods on the five datasets. Clearly, our approach obtains a better PR performance than the other approaches.

Fig. 4 shows the PR carves with fixed thresholds of all the above methods on the five datasets, which demonstrate that our method achieves the consistent and favorable performance against the other competing approaches. From the Fig. 5, it is observed that our proposed approach achieves the best F-measure performance. Further, Fig. 6 presents the corresponding VOC scores of all the twelve approaches. It can be seen from the bar graphs, our model obtains the highest VOC scores over the five datasets. Overall, our approach consistency outperforms each individual saliency detection approach and state-of-the-art methods.

4.3 Visual Comparison

Fig. 7 presents some results of our method with the seven selected individual saliency detection methods. All of these images are from the four datasets ASD,

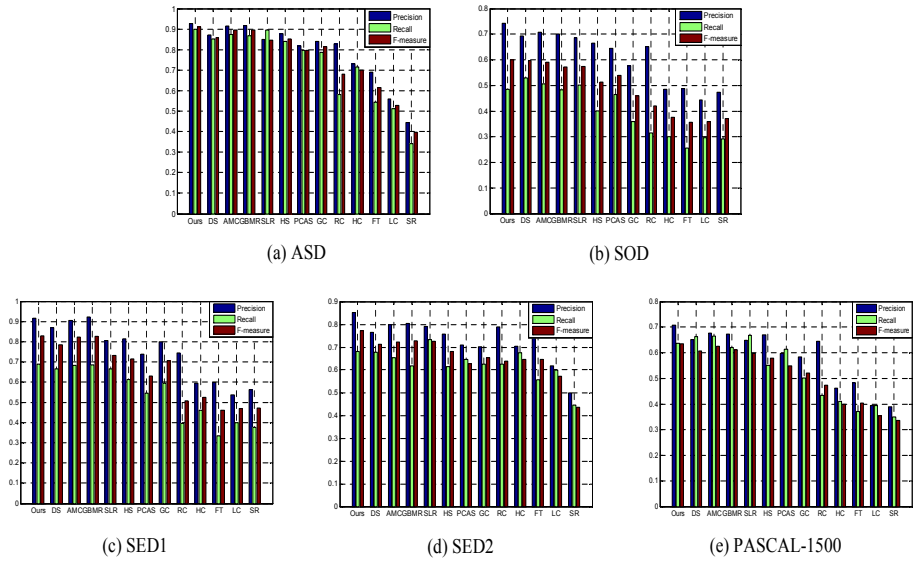


Fig. 5. Average precision, recall and F-measure of all the twelve approaches over the five datasets. Our method achieves the best precision, recall and F-measure.

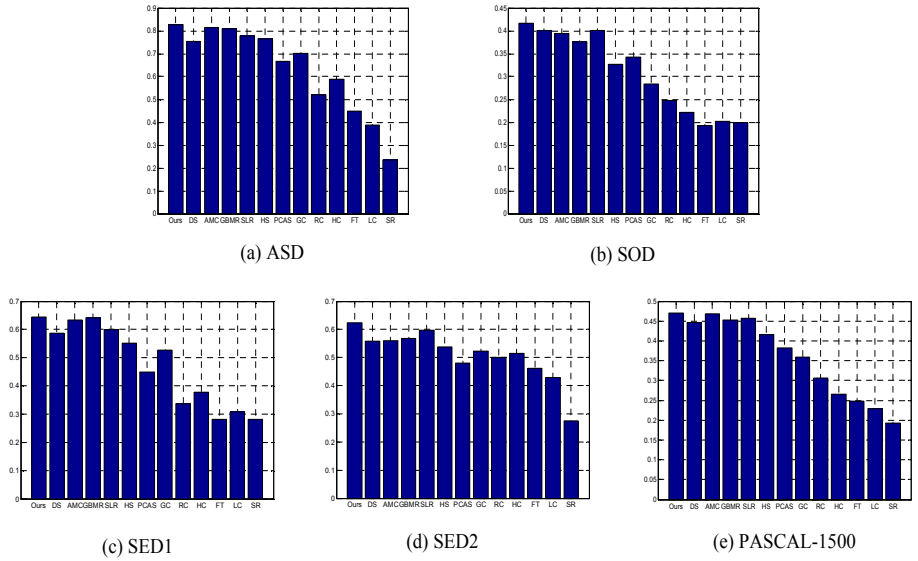


Fig. 6. Quantitative VOC performance of all the twelve methods on the five datasets. Our approach achieves the highest VOC scores over the five datasets.

SOD, SED1, SED2. Visually, these images are simple, however, most of them are with cluttered object or background. In the individual saliency maps, some salient regions are not pop not from the background, especially for the images which come from the SED2 dataset. And, the results shown in Fig. 7 (h), some parts of background pixels or regions also stand out simultaneously with the object regions. Compared with the ground truth shown in Fig. 7 (j), we can clearly see that our approach consistently outperforms each individual saliency detection method. This confirms that our model can effectively integrate the results of these methods. More comparison results on the PASCAL-1500 dataset are shown in Fig. 8.

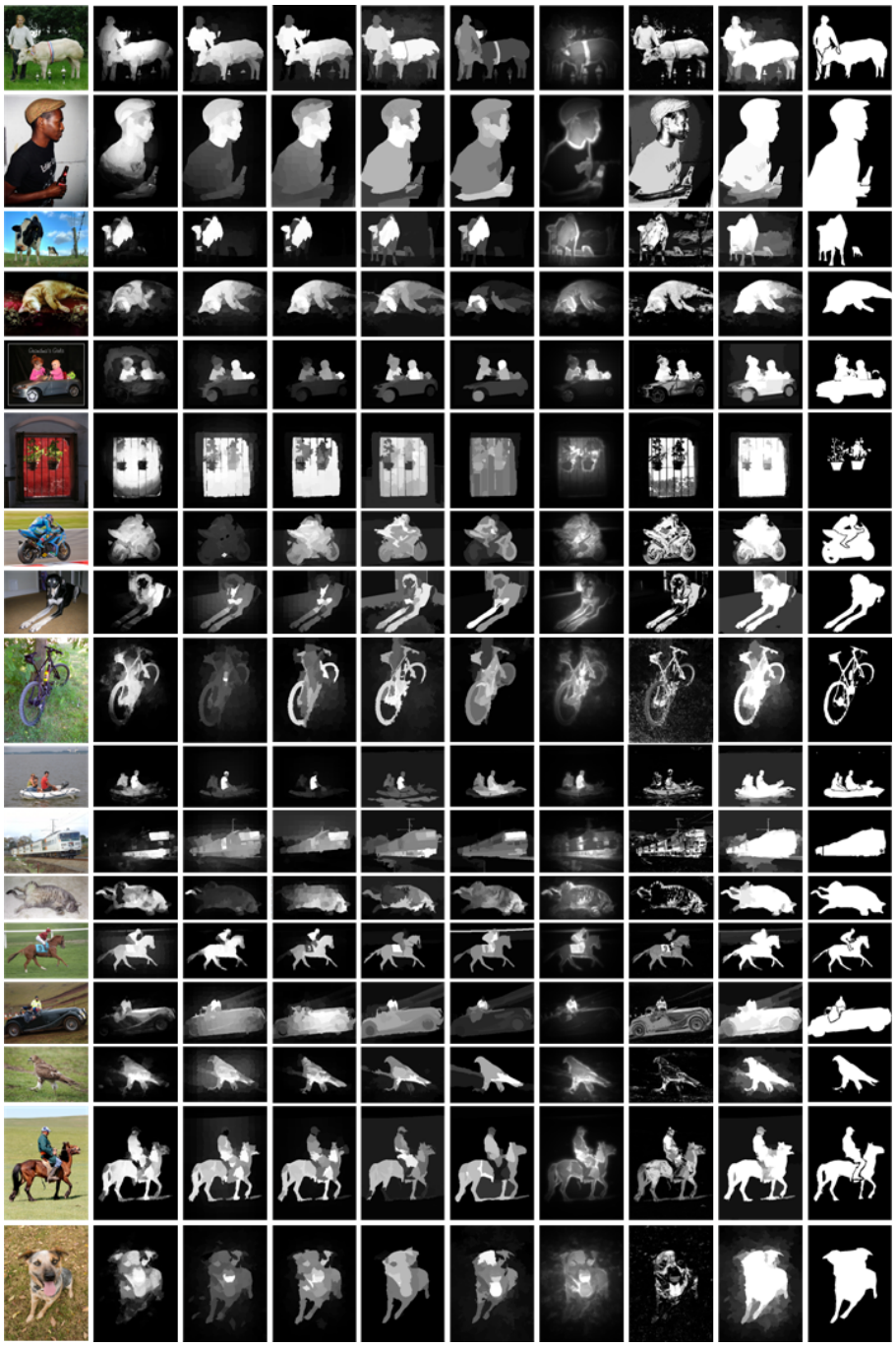
(a) Image (b) DS (c) AMC (d) GBMR (e) SLR (f) HS (g) PCAS (h) GC (i) Ours (j) GT

Fig. 7. Examples of saliency learning on the four datasets: ASD (rows 1-2), SOD (rows 3-4), SED1 (rows 5-6), and SED2 (rows 7-8). Given an input image (a), we conduct some saliency detection methods and obtain corresponding individual saliency maps (b)-(h), (i) shows the saliency map produced by our model. Compared with the ground truth (j), our model achieved best performance visually.

In addition, some saliency maps of the evaluated methods are shown in Fig.9. We note that the proposed method can uniformly highlight the salient regions and preserve object boundaries well than the other approaches.

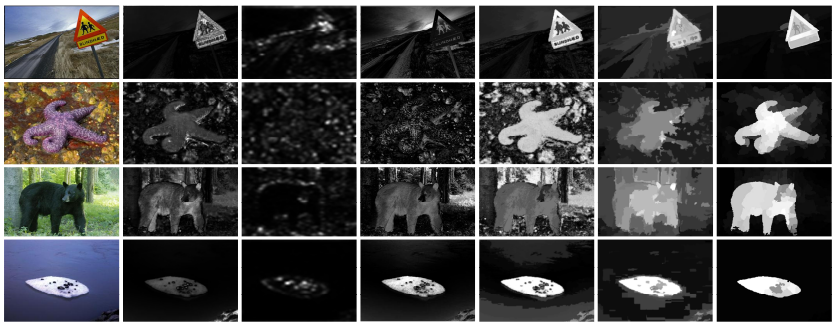
4.4 Limitations

Up until now, we have evaluated the effectiveness of our model which can consistently improve the performance of each individual saliency detection method.



(a) Image (b) DS (c) AMC (d) GBMR (e) HS (f) SLR (g) PCAS (h) GC (i) Ours (j) GT

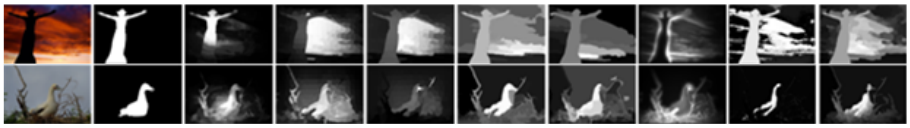
Fig. 8. More comparison results on the PASCAL-1500 dataset. Compared with the ground truth (j), our model achieved best performance visually.



(a) Input (b) FT (c) SR (d) LC (e) HC (f) RC (g) Ours

Fig. 9. Visual comparisons between five classical methods and our proposed approach.

However, some difficult images are still challenging for our model as well as other state-of-the-art saliency models. As shown in Fig. 10, most of the saliency models fail to identify the salient regions for an image. In this case, the advantages of some saliency maps are not learned into the low rank matrix recovery based framework to help the final saliency combination.



(a) Image (b) GT (c) DS (d) AMC (e) GBMR (f) HS (g) SLR (h) PCAS (i) GC (j) Ours

Fig. 10. Failure cases. The advantages of some saliency maps are not learned into our framework to help the final saliency combination.

5 Conclusions and Future Work

In this paper, we propose a saliency learning model for saliency aggregation that integrates saliency maps produced by multiple saliency detection methods. Specifically, we cast the saliency detection as a sparsity pursuit problem. Our method provides a robust way to combine individual saliency detection methods into a more powerful one. Experimental results prove that our proposed approach performs better than the individual saliency detection methods and outperforms the state-of-the-art approaches.

With the development of saliency detection technology, we believe that our method can benefit from the forthcoming saliency methods. In the future, we plan to investigate a strategy to learn which method can improve the performance of our aggregation model. Furthermore, we will focus on how to overcome the failure of our method and how to improve its speed.

References

1. B. C. Ko and J. Y. Nam: Object-of-interest image segmentation based on human attention and semantic region clustering. In: JOptSoc Am **23** (2006) 2462–2470
2. J. Li, R. Ma and J. Ding: Saliency-seeded region merging: automatic object segmentation. In: ACPR (2011) 691–695
3. J. Han, K. Ngan, M. Li, and H. Zhang: Unsupervised extraction of visual attention objects in color images. In: IEEE Transactions on Circuits and Systems for Video Technology **16** (2006) 141–145
4. T. Chen, M. Cheng, P. Tan, A. Shamir, and S. Hu: Sketch2photo: Internet image montage. In: ACM Transactions on Graphics **28** (2009) 1–10
5. S. Avidan and A. Shamir: Seam carving for content-aware image resizing. In: ACM Transactions on Graphics **26** (2007)
6. A. Santella, M. Agrawala, D. Decarlo, D. Salesin, and M. Cohen: Gaze-based interaction for semi-automatic photo cropping. In: Proc. Conf. Human Factors in Computing Systems (2006) 771–780
7. C. Christopoulos, A. Skodras, and T. Ebrahimi: The JPEG2000 still image coding system: an overview. In: IEEE Trans. Consumer Elec. **46** (2002) 1103–1127
8. R. Achanta, F. J. Estrada, P. Wils, and S. Susstrunk: Salient region detection and segmentation. In: ICVS (2008) 66–75
9. R. Achanta, S. Hemami, F. J. Estrada, and S. Susstrunk: Frequency-tuned salient region detection. In: CVPR (2009) 1597–1604
10. S. Goferman, L. Zelnik-Manor, and A. Tal: Context-aware saliency detection. In: CVPR (2010) 2376–2383
11. M. Cheng, G. Zhang, N. J. Mitra, X. Huang, and S. Hu: Global contrast based salient region detection. In: CVPR (2010) 409–416
12. A. Borji, D. N. Sihite, and L. Itti: Salient object detection: A benchmark. In: ECCV (2012)
13. L. Mai, Y. Niu, and F. Liu: Saliency aggregation: a data-driven approach. In: CVPR (2013) 4321–4328
14. Alpher, A.: Advances in Frobnication. J. of Foo **23** (2006) 234–778
15. X. Shen and Y. Wu: A unified approach to salient object detection via low rank matrix recovery. In: CVPR (2012) 853–860
16. X. Li, H. Lu, L. Zhang, X. Ruan, and M. Yang: Saliency detection via dense and sparse reconstruction. In: ICCV (2013) 2976–2983
17. Q. Yan, L. Xu, J. Shi, and J. Jia: Hierarchical saliency detection. In: CVPR (2013) 1155–1162
18. B. Jiang, L. Zhang, H. Lu and M. Yang: Saliency detection via absorbing markov chain. In: ICCV (2013) 1665–1672
19. W. Zou, K. Kpalma, Z. Liu, and J. Ronsin: Segmentation driven low-rank matrix recovery for saliency detection. In: BMVC (2013) 1–13
20. R. Margolin, A. Tal, L. Manor: What makes a patch distinct?. In: CVPR (2013) 1139–1146
21. J. Yan, M. Zhu, H. Liu, and Y. Liu: Visual saliency detection via sparsity pursuit. In: IEEE Signal Process. Lett. **17** (2010) 739–742
22. C. Lang, G. Liu, J. Yu, and S. Yan: Saliency detection by multitask sparsity pursuit. In: IEEE Trans. on Image Processing **21** (2012) 1327–1338
23. J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma: Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In: NIPS (2009)

24. L. Itti, C. Koch, and E. Niebur: A model of saliency-based visual attention for rapid scene analysis. In: IEEE TPAMI **20** (1998) 1254–1259
25. Y. Ma and H. Zhang: Contrast-based image attention analysis by using fuzzy growing. In: ACM Multimedia (2003) 374–381
26. J. Harel, C. Koch, and P. Perona: Graph-based visual saliency. In: NIPS (2006) 545–552
27. X. Hou and L. Zhang: Saliency detection: a spectral residual approach. In: CVPR (2007) 1–8
28. D. Comaniciu and P. Meer: Mean shift: A robust approach toward feature space analysis. J. of Foo **24** (2002) 603–619
29. Z. Lin, M. Chen, L. Wu, and Y. Ma: The augmented lagrange multiplier method for extract recovery of corrupted low rank matrices UIUC. In: Tech. Rep. UILU-ENG-09-2215 (2009)
30. S. Alpert, M. Galun, R. Basri and A. Brandt: Image segmentation by probabilistic bottom-up aggregation and cue integration. In: CVPR (2007) 1–8
31. V.Movahedi and J. Elder: Design and perceptual validation of performance measures for salient object segmentation. In: CVPRW (2010) 49–56
32. M. Cheng, J. Warrell, W. Lin, S. Zheng, V. Vineet, and N. Crook: Efficient salient region detection with soft image abstraction. In: ICCV (2013) 1529–1536
33. C. Yang, L. Zhang, H. Lu, X. Ruan, and M. Yang: Saliency detection via graph-based manifold ranking. In: CVPR (2013) 3166–3173
34. M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn and A. Zisserman: The pascal visual object classes (VOC) challenge. In: IJCV **88** (2010) 303–338
35. A. Rosenfeld and D. Weinshall: Extracting foreground masks towards object recognition. In: ICCV (2011) 1371–1378
36. Y. Zhai and M. Shah: Visual attentation detection in video sequences using spatiotemporal cues. In: ACM Multimedia (2006) 815–824