A Comparative Study of Saliency Aggregation for Salient Object Detection

Shuhan Chen[1,2](✉), Ling Zheng[1], Xuelong Hu[1], and Ping Zhou[2]

[1] College of Information Engineering, Yangzhou University, Yangzhou, China
zlapgx@163.com, xlhu@yzu.edu.cn
[2] Wanfang Electronic Technology Co., Ltd., Yangzhou, China
c.shuhan@gmail.com, zp@wfdz.com.cn

Abstract. A variety of saliency detection methods have been proposed in recently, which often complement each other. In this study, we try to improve their performances by aggregating these individual ones. First, we propose an improved Bayes aggregation method with double thresholds. Then, we compare it with five other aggregation approaches on four benchmark datasets. Experiments show that all the aggregation methods significantly outperform each individual one. Among these aggregation methods, average and Non-negative Matrix Factorization (NMF) weights perform best in terms of precision-recall curve, our Bayes is very close to them. While for mean absolute error score, NMF and our Bayes perform best. We also find that it is possible to further improve their performance by using more accurate reference map. The ideal is ground truth, of course. Our results could have an important impact for applications required robust and uniform saliency maps.

Keywords: Saliency detection · Saliency aggregation · Bayes aggregation

1 Introduction

Visual saliency has become a very active topic in computer vision, which measures the low-level stimuli that grabs viewers' attention in the early stage of human vision [1], whose research mainly contains three aspects [23]: eye fixation prediction [2], salient object detection or saliency detection [3], objectness estimation [4]. Among them, saliency detection is the most active area and a large number of approaches have been proposed in the literatures [3, 5, 6, 9–16, 20–22, 25], which is also the focus of this paper. Saliency detection, which aims to make certain regions of an image stand out from their neighbors and catch immediate attention, has attracted growing concern and made a great progress in recent five years. Efficient saliency detection makes great helpfulness to a wide range of computer vision tasks, such as object detection, segmentation, recognition, compression and so on.

Although many well performed saliency detection methods have been proposed in recent years and most of them can uniformly highlight the salient object in an image, there still exists a large margin from the ground truth, especially facing complex scenes. In addition, the performance of each method is often image-dependent, in other words, each method has its own advantages and disadvantages. More interestingly,

© Springer International Publishing Switzerland 2015
Y.-J. Zhang (Ed.): ICIG 2015, Part I, LNCS 9217, pp. 30–42, 2015.
DOI: 10.1007/978-3-319-21978-3_4

different approaches can often be complementary to each other [5]. As illustrated by Fig. 1, different saliency maps usually do not exhibit similar characteristics and each of them only works well for some images or some parts of the image, and none of them can handle all the images, such as Fig. 1(b) and (e) in the second row are complementary in measuring saliency. Thus, we can naturally ask a question: Does unity make strength?

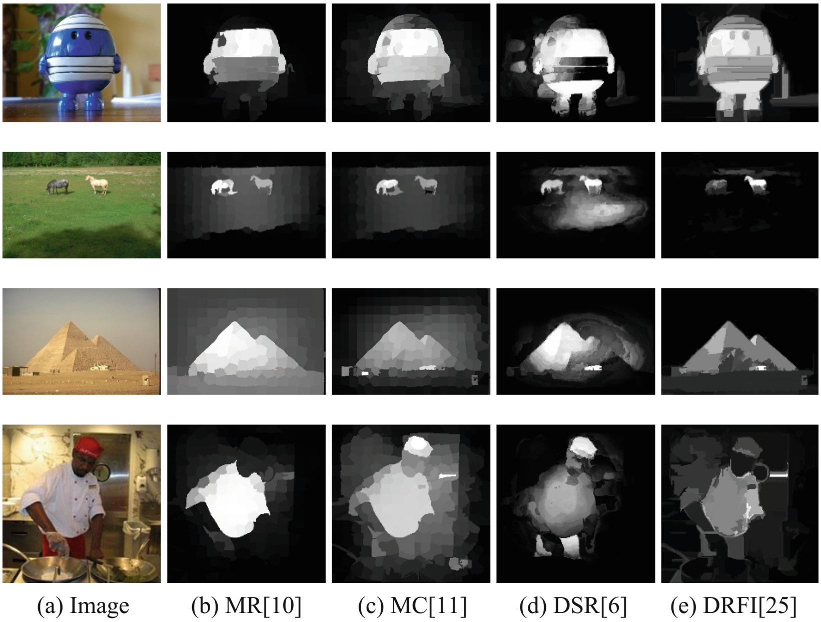


Fig. 1. Individual saliency methods often complement each other. Saliency aggregation can effectively outperform each one of them.

Only a few literatures have been explored on saliency aggregation. In [6], Li et al. proposed a saliency detection method whose final step is Bayes integration of two saliency maps generated by dense and sparse reconstruction errors respectively. Focusing on eye fixation prediction, Le Meur et al. [7] made a detailed comparison of various aggregation methods including unsupervised and learning-based schemes, in which got the following conclusions: a simple average of the top two saliency maps significantly outperforms each individual one, and considering more saliency maps tends to decrease the performance. Similar to this work, Borji et al. [8] proposed two combining strategies: Naive Bayesian evidence accumulation and linear summation, which also demonstrated aggregation results working better than individual ones. In [5], Conditional Random Field (CRF) was used for saliency aggregation. However, CRF is time consuming due to its training and inference steps.

The main drawback of the above mentioned studies concerns the choice of the tested individual methods, which have poor performance compared with the state-of-the-art due to the quick development of this area. Therefore, there are still a big room can be explored and improved. Different with the aforementioned studies, we select top five state-of-the-art methods [6, 10, 11, 13, 25] for aggregation and make comparison on four benchmark datasets, and try to make the detected salient object more uniform and as close as possible to the ground truth. In our study, we try to investigate the following issues: whether we could improve the saliency detection quality by aggregating various saliency maps or not, whether considering more saliency maps decreases the performance or not, whether linear average combination outperforms the others or not. Our results could have an important impact for applications required robust and uniform saliency maps.

The rest of the paper is organized as follows. Section 2 presents the different methods used for saliency aggregation, including our improved Bayes aggregation. Section 3 shows the performance of the recent saliency approaches, taken alone, and the performance of the aggregation schemes. Finally, conclusions and future work are listed in Sect. 4.

2 Saliency Aggregation

In this section, we introduce different aggregation approaches including previous works and our proposed method. Among the previous works, we mainly focus the ones with high performance. Before that, we make some definition of the symbols which will be used in the following subsections. Let $\{S_i||1 \leq i \leq m\}$ be the saliency maps generated by different saliency detection algorithms on a given image I, whose saliency value in each map is normalized to [0, 1], G is the corresponding ground truth. Each element $S_i\,(z)$ in a saliency map denotes the saliency value at pixel z. Our goal is to aggregate these m saliency maps into a final saliency map which can outperform each individual one.

2.1 Linear Aggregation

The simplest aggregation scheme is linear summation which is defined as below:

$$S = \sum_{i=1}^{m} w_i \times S_i \tag{1}$$

where w_i is the weighting coefficient, the sum of it is equal to 1 and $w_i \geq 0$.

Based on this function, we can design various aggregation schemes only by varying the weighting coefficients. The most used is average weights which is uniform and spatial invariant, $w_i = 1/m$. We call it AVG for short, which is verified by Borji [8] and Le Meur [7] and can produce satisfied aggregation results.

Linear weights can also be computed by minimizing the residual between different saliency maps and their corresponding ground truths.

$$W = \arg\min \left\| G - \sum_{i=1}^{m} w_i S_i \right\|^2 \tag{2}$$

In which W is the vector of weights with m dimensions.

Here, we summarize three different methods to compute W. The first one is the classical least-squares method (LS). However, the LS weights do not sum to one and can be positive or negative. Thus, we can add constraints to ease the interpretation of the computed weights. But it will also reduce the solution space. The second one is to make the weights have to sum to one, which moves the LS problem onto the Locally Linear Embedding (LLE) [18]. The final one is to make the weights not only to sum to one but also to be positive, which is similar to the problem of Non-negative Matrix Factorization (NMF) [19].

Then, the only problem is how to produce a reference map S_p to instead of the ground truth G. Here, we simply compute it by linear summation of input saliency maps.

$$S_p = \frac{1}{m}\sum_{i=1}^{m} S_i \tag{3}$$

2.2 Nonlinear Aggregation

A nonlinear combination is also tested by Le Meur [7] which is defined as:

$$S_{MED}(z) = \text{median}\{S_1(z), \cdots, S_i(z), \cdots, S_m(z)\}, m \geq 3 \tag{4}$$

Max and min can also be applied in the above function. However, they are significantly worse than the median weight (MED) in our experiment. Thus, they are not selected for comparison.

2.3 Bayes Aggregation

In [8], various saliency models are combined by a Naive Bayesian evidence accumulation. It is simplified to a multiplication case for simplicity, where the posterior probability is replaced by the saliency value of different saliency model at each pixel. However, its performance is very poor. While in [6], Bayes aggregation is proposed for two saliency maps. In this paper, we improve it to fit the case of multiple saliency maps.

Given m saliency maps, we select S_p as the prior and use each individual one S_i to compute the likelihood. Then, S_p is thresholded to obtain its background and foreground regions described by B_p and F_p respectively. In each region, likelihoods are computed by comparing S_i and S_p in terms of the background and foreground bins at pixel z:

$$P\left(S_i(z)|B_p\right) = \frac{N_{b_{B_p(S_i(z))}}}{N_{B_p}}, P\left(S_i(z)|F_p\right) = \frac{N_{b_{F_p(S_i(z))}}}{N_{F_p}} \tag{5}$$

where N_{B_p} denotes the number of pixels in the background region and $N_{b_{B_p(S_i(z))}}$ is the number of pixels whose saliency value fall into the background bin $b_{B_p(S_i(z))}$, while N_{F_p} and $N_{b_{F_p(S_i(z))}}$ are denoted for foreground region.

Consequently, the posterior probability is computed as below:

$$P\left(F_p|S_i(z)\right) = \frac{S_p(z)P\left(S_i(z)|F_p\right)}{S_p(z)P\left(S_i(z)|F_p\right) + \left(1 - S_p(z)\right)P\left(S_i(z)|B_p\right)} \tag{6}$$

Then, combined saliency map can be generated by the summation of these posterior probabilities:

$$S_{Bayes}(z) = \frac{1}{m}\sum_{i=1}^{m} P\left(F_p|S_i(z)\right) \tag{7}$$

To improve performance, double thresholds are used for the binaryzation of S_p. The detail is described in Sect. 3.3. Thus, we can get two aggregated saliency maps which are used to produce the final Bayes aggregation result.

In our experiment, we find that the Bayes aggregation outperforms most of the aggregation methods and each individual saliency model in mean absolute error (MAE) score with a large margin. Based on this observation, we introduce it into the input saliency maps to further improve performance. Then we have m + 1 input saliency maps in total, which are used for linear and nonlinear aggregation mentioned above. While for nonlinear aggregation (MED), Bayes result is only introduced when the number of the input saliency maps is even. Therefore, the new reference map S'_p is replaced by

$$S'_p = \frac{1}{m+1}\left(S_{Bayes} + \sum_{i=1}^{m} S_i\right) \tag{8}$$

3 Performance Evaluation

In our study, top five state-of-the-art salient object detection methods are selected for aggregation as reported in [17], including DSR [6], MR [10], MC [11], RBD [13], DRFI [25], whose codes or results can be acquired from the authors' personal websites. There are six aggregation schemes tested which are denoted as: AVG, MED, LS, NMF, LLE, and Bayes.

3.1 Datasets

For fair comparison, it is necessary to test over different datasets so as to draw objective conclusions. To this end, four widely used benchmark datasets are selected including: ASD [9], SOD [24], ECSSD [15], and DUT-OMRON [10].

ASD includes 1000 images selected from the MSRA database. Most images in it have only one salient object and there are usually strong contrast between objects and backgrounds. SOD is based on the Berkeley segmentation dataset. ECSSD consists of a large number of semantically meaningful but structurally complex natural images. DUT-OMRON contains 5,172 carefully labeled images.

3.2 Evaluation Measures

Precision-Recall (PR) curve and MAE are evaluated in our experiments. PR curve: Given corresponding masks, the precision and recall rate are defined as bellows:

$$\text{Precision} = \frac{|M \cap G|}{|M|}, \text{Recall} = \frac{|M \cap G|}{|G|} \tag{9}$$

where M is the binary object mask generated by thresholding corresponding saliency map and G is the corresponding ground truth. A fixed threshold changing from 0 to 255 is used for thresholding. On each threshold, a pair of precision/recall scores are computed, and are finally combined to form a PR curve to describe the performance at different situations.

MAE: PR curve does not consider the true negative saliency assignments. For a more comprehensive comparison, MAE is further introduced to evaluate the performance between the saliency map S and the ground truth G, which is defined as:

$$\text{MAE} = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} |S(i,j) - G(i,j)| \tag{10}$$

where W and H are the width and the height of the saliency map, respectively. Lower MAE value indicates better performance. This measure is also found complementary to PR curves [13, 22]. As described in [17], we draw our conclusions mainly based on PR curves, and also report MAE scores for comprehensive comparisons and for facilitating specific application requirements.

3.3 Quantitative Comparison

Validation of Double Thresholds. In our study, the high threshold TH is generated by Otsu and the low one TL is set to $0.5TH$. The results are shown in Fig. 2. Note that scales are different for different figures to improve the clarity of the plot, similarly hereinafter. We can see that our double thresholds can effectively improve the performance of Bayes aggregation.

Influence of S_p. In this comparison, we try to find out that whether different S_p can influence the performance of saliency aggregation. Three different reference maps are tested, which are S_p, S'_p, and G. The results are shown in Fig. 3, in which top two saliency models are used for average in each dataset. Using G as the reference map is as

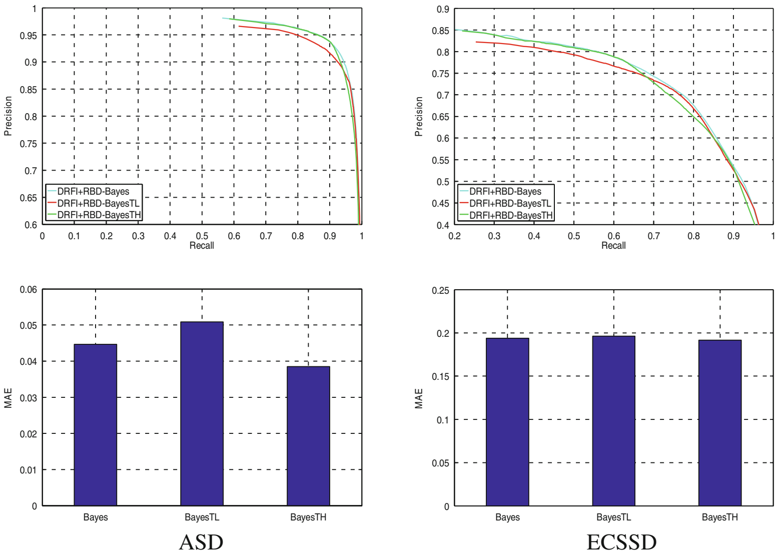


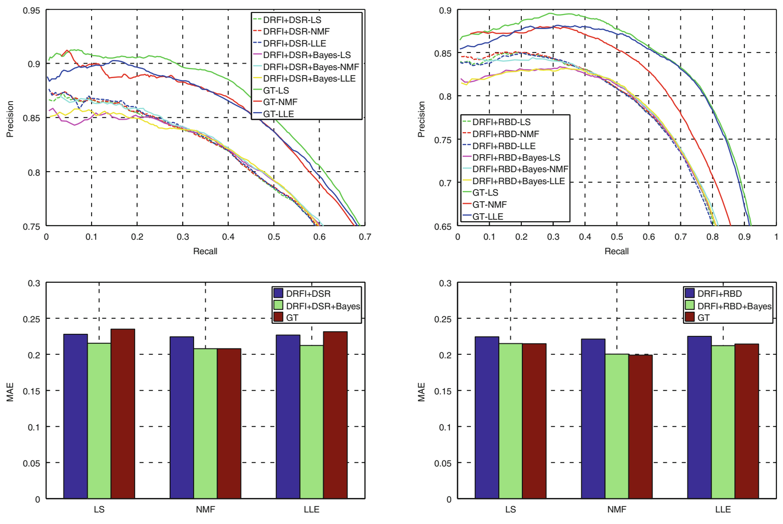
Fig. 2. PR curves (left) and MAE (right) results of double thresholds in Bayes aggregation.

expected to outperform the others with large margins in PR curves. Their MAE scores are almost the same. Thus, we should try to make the reference map as accurate as *G*, which is our future work.

With or Without Bayes Aggregation as Input. In this comparison, we try to find out that whether introducing Bayes aggregation result into the input saliency maps can improve performance or not. As can be seen from Fig. 4, their performance is significantly improved by combining Bayes aggregation result into the input saliency maps except AVG aggregation which has almost no change in terms of PR curve. While for MAE score, it can be improved with big margins in all the datasets.

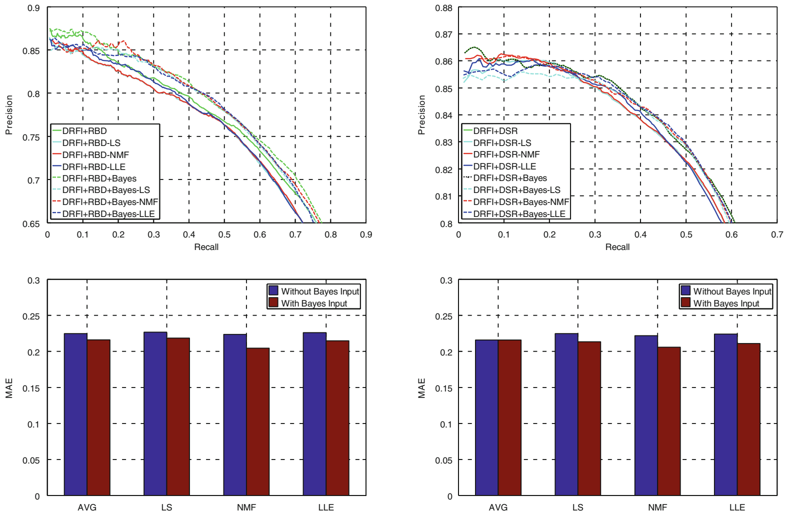
AVG Aggregation of Top Two or More. We first examine whether considering more saliency maps decreases the performance or not. We choose AVG aggregation for comparison. Figure 5 shows all of the saliency aggregation methods by AVG significantly outperform each individual one. However, it is hard to say which one is better between AVG aggregation of top two and more. Their performances are almost the same both in PR curve and MAE score. Thus, we only select the top two for aggregation in the other comparisons for efficiency.

Aggregation Performance Comparison. Finally, we try to find out which aggregation method performs best, which is the main purpose in our study. Here, we compare six aggregation methods mentioned above using top two individual models for each dataset. As can be seen in Fig. 6, in terms of PR curve, AVG performances best in ASD dataset, while in the other datasets, AVG and NMF achieve almost the same



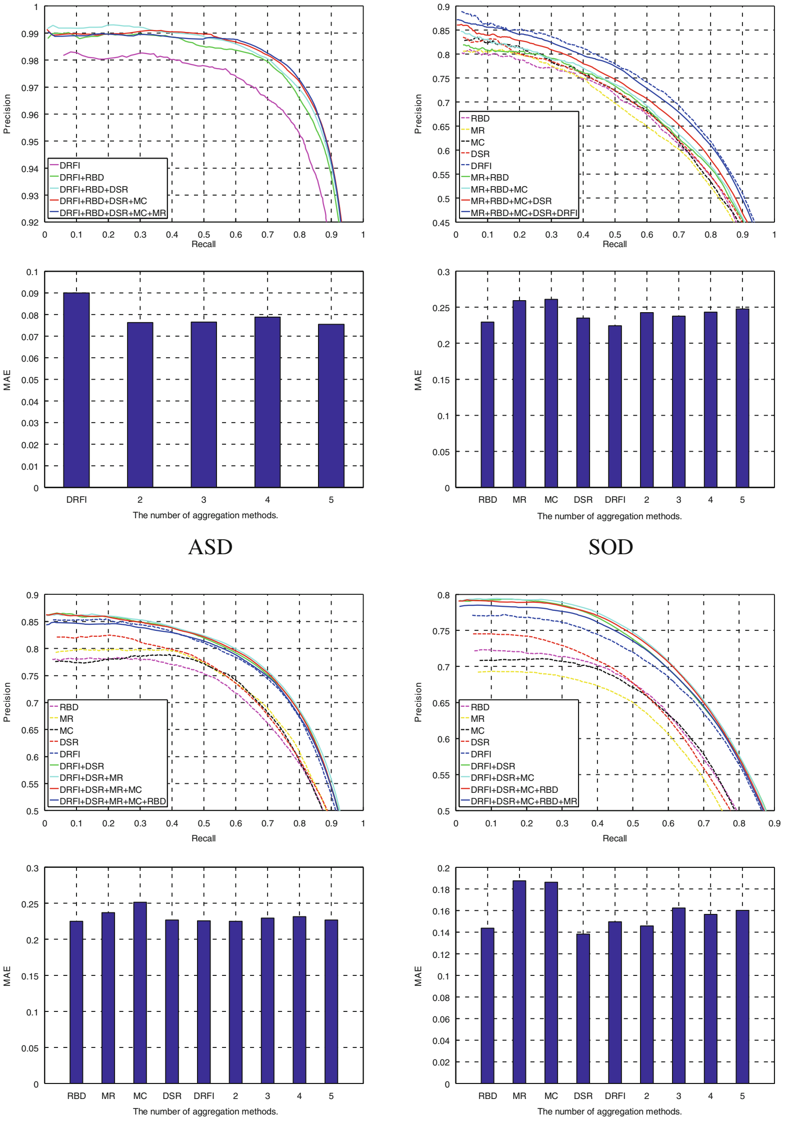
SOD ECSSD

Fig. 3. Comparison with different reference maps.



SOD ECSSD

Fig. 4. Comparison between with and without Bayes result as input.



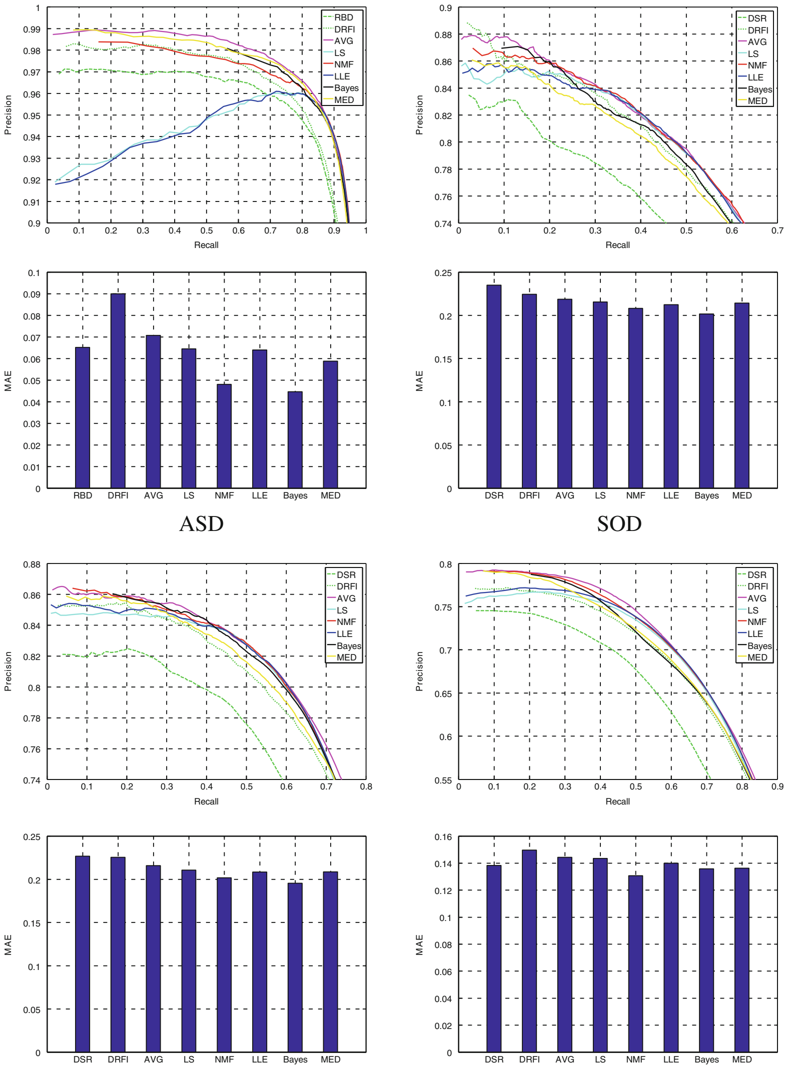
ASD

SOD

ECSSD

DUT-OMRON

Fig. 5. Comparison of AVG aggregation with different input numbers.



ASD

SOD

ECSSD

DUT-OMRON

Fig. 6. Comparison of various aggregation methods on different datasets.

performance. With respect to the MAE score, NMF and our improved Bayes outperform all the others on all datasets. In addition, our improved Bayes also achieve close performance to them in PR curve. Some representative aggregation results are shown in Fig. 7.

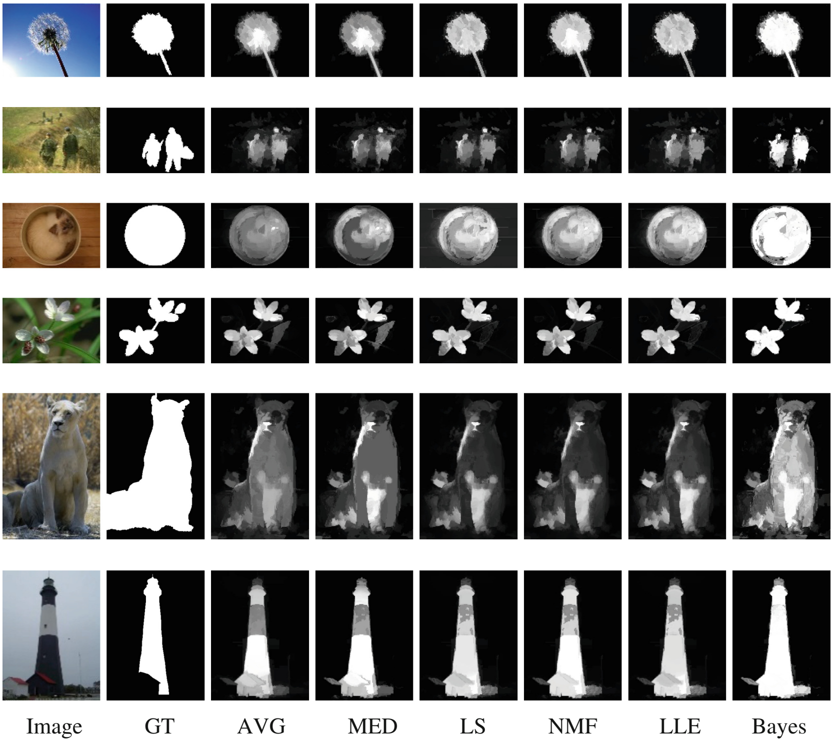


Fig. 7. Saliency aggregation examples.

4 Conclusions and Future Work

In this paper, we make a comparison of different saliency aggregation methods and also propose an improved Bayes aggregation approach. By detailed experiments and analysis, we can draw the following conclusions: (1) It is hard to say which one is better between AVG aggregation of top two and more. Thus, AVG aggregation of top two is the best choice for efficiency. (2) AVG and NMF usually get best performance in PR curve, and NMF and Bayes outperform the others in MAE score. In addition, Bayes is very close to AVG and NMF in PR curve. Lower MAE means that the saliency value is very close to the ground truth. Thus, our improved Bayes aggregation is a good choice for object segmentation task. (3) Introducing Bayes result into the input saliency maps for the other aggregations can significantly improve their performance. Double thresholds can further improve the performance of Bayes aggregation.

The key problem of saliency aggregation mentioned in this paper is how to generate the reference map S_p. Here, we only simply generate it by linear summation of all input saliency maps. In the future, we will try to improve it by using good object segmentation methods, such as SaliencyCut [3]. It will also be required to make a more comprehensive study with more saliency methods on more benchmark datasets.

Acknowledgements. The authors would like to express their sincere thanks to the anonymous reviewers for their invaluable suggestions and comments to improve this paper. This work was supported by "Lv Yang Jin Feng" Program of Yangzhou, science and technology planning project of General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China (No. 2014QK260).

References

1. Itti, L., Koch, C.: Computational modeling of visual attention. Nat. Rev. Neurosci. **2**, 194–203 (2001)
2. Hou, X., Zhang, L.: Saliency detection: a spectral residual approach. In: CVPR (2007)
3. Cheng, M.-M., Zhang, G.-X., Mitra, N.J., Huang, X., Hu, S.-M.: Global contrast based salient region detection. In: CVPR (2011)
4. Cheng, M.-M., Zhang, Z., Lin, W.-Y., Torr, P.: BING: binarized normed gradients for objectness estimation at 300fps. In: CVPR (2014)
5. Mai, L., Niu, Y., Liu, F.: Saliency aggregation: a data-driven approach. In: CVPR (2013)
6. Li, X., Lu, H., Zhang, L., Ruan, X., Yang, M.-H.: Saliency detection via dense and sparse reconstruction. In: ICCV (2013)
7. Le Meur, O., Liu, Z.: Saliency aggregation: does unity make strength? In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) ACCV 2014. LNCS, vol. 9006, pp. 18–32. Springer, Heidelberg (2015)
8. Borji, A., Sihite, D.N., Itti, L.: Salient object detection: a benchmark. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part II. LNCS, vol. 7573, pp. 414–429. Springer, Heidelberg (2012)
9. Achantay, R., Hemamiz, S., Estraday, F., Susstrunk, S.: Frequency-tuned salient region detection. In: CVPR (2009)
10. Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M-H.: Saliency detection via graph-based manifold ranking. In: CVPR (2013)
11. Jiang, B., Zhang, L., Lu, H., Yang, M.-H., Yang, C.: Saliency detection via absorbing Markov chain. In: ICCV (2013)
12. Jiang, H., Wang, J., Yuan, Z., Wu, Y., Zheng, N., Li, S.: Salient object detection: a discriminative regional feature integration approach. In: CVPR (2013)
13. Zhu, W., Liang, S., Wei, Y., Sun, J.: Saliency optimization from robust background detection. In: CVPR (2014)
14. Cheng, M.-M., Mitra, N.J., Huang, X., Torr, P.H.S., Hi, S.-M.: Global contrast based salient region detection. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 569–582 (2015)
15. Yan, Q., Xu, L., Shi, J., Jia, J.: Hierarchical saliency detection. In: CVPR (2013)
16. Liu, T., Sun, J., Zheng, N.-N., Tang, X., Shum, H.-Y.: Learning to detect a salient object. In: CVPR (2007)
17. Borji, A., Cheng, M.-M., Jiang, H., Li, J.: Salient object detection: a benchmark. arXiv eprint (2015)

18. Roweis, S., Saul, L.: Nonlinear dimensionality reduction by locally linear embedding. Science **5500**, 2323–2326 (2000)
19. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: NIPS (2000)
20. Liu, R., Cao, J., Lin, Z., Shan, S.: Adaptive partial differential equation learning for visual saliency detection. In: CVPR (2014)
21. Chen, S.-H., Shi, W.-R., Zhang, W.-J.: Visual saliency detection via multiple background estimation and spatial distribution. Optik **125**(1), 569–574 (2014)
22. Perazzi, F., Krahenbuhl, P., Pritch, Y., Hornung, A.: Saliency filters: contrast based filtering for salient region detection. In: CVPR (2012)
23. Borji, A., Cheng, M.-M., Jiang, H., Li, J.: Salient object detection: a survey. arXiv eprint (2014)
24. Movahedi, V., Elder, J. H.: Design and perceptual validation of performance measures for salient object segmentation. In: POCV, pp. 49–56 (2010)
25. Jiang, H., Yuan, Z., Cheng, M.-M., Gong, Y., Zheng, N., Wang, J.: Salient object detection: a discriminative regional feature integration approach. arXiv (2014)