Comparing Salient Object Detection Results without Ground Truth

Long Mai and Feng Liu

Department of Computer Science, Portland State University, USA

Abstract. A wide variety of methods have been developed to approach the problem of salient object detection. The performance of these methods is often image-dependent. This paper aims to develop a method that is able to select for an input image the best salient object detection result from many results produced by different methods. This is a challenging task as different salient object detection results need to be compared without any ground truth. This paper addresses this challenge by designing a range of features to measure the quality of salient object detection results. These features are then used in various machine learning algorithms to rank different salient object detection results. Our experiments show that our method is promising for ranking salient object detection results and our method is also able to pick the best salient object detection result such that the overall salient object detection performance is better than each individual method.

1 Introduction

Visual saliency measures the low-level stimuli that grabs viewers' attention in the early stage of human vision [21]. It has been used as an alternative to semantic image understanding in a range of applications in computer vision, computer graphics, and multimedia, such as object detection [22], adaptive image compression [9], and content-aware image manipulation [35, 42]. There is now a rich literature on visual saliency analysis [1,4,8,10,13–16,18,19,22,23,26,30,31,33,34, 36–41, 43, 44, 50–54, 56, 57]. Many of these methods aim to detect salient objects from an input image as recently surveyed in [2], which is also the focus of this paper.

While the research on salient object detection has been progressing quickly and has achieved good results statistically on public benchmarks [7, 24, 25, 28, 29, 33, 37, 40, 50, 53, 54], each individual method has its own advantages and disadvantages. As shown in Fig. 1, each method can produce good results for some images but none of them can outperform the other methods for all the images. For a specific input image, it is often useful to select the best salient object detection result from many results created by different methods. This is a challenging task as the quality of different salient object detection results need to be compared without knowing the ground truth.

Our problem is relevant to the research on non-reference image quality assessment, which estimates the quality of an image without a ground-truth one [45,49].

D. Fleet et al. (Eds.): ECCV 2014, Part III, LNCS 8691, pp. 76–91, 2014.

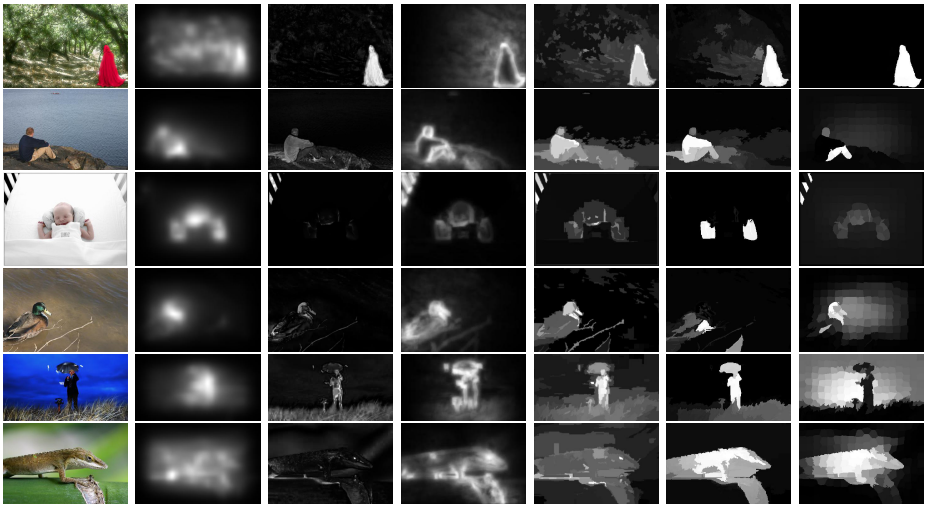


Image GBVS [16] FT [1] CA [14] GC [8] HS [53] GBMR [54]

Fig. 1. Saliency detection examples. Different saliency detection methods have their own advantages and disadvantages. Each method can produce good results on some images but none of them can outperform the others on all the cases.

The non-reference image quality assessment methods mostly detect and measure image artifacts, such as those from compression. They cannot be used in our problem.

This paper addresses the problem of comparing salient object detection results without ground truth using a data-driven approach. We study what makes a good salient object detection result and design a range of features to measure the quality of salient object detection. We then use a learning-to-rank method to rank salient object detection results of the same input image. Specifically, we first train a binary classifier to compare the quality between every two detection results and then aggregate these pair-wise comparison results to rank all the detection results.

To the best of our knowledge, this paper provides the first method that is able to rank the quality of salient object detection results without any ground truth. While this paper does not provide a new salient object detection method per se, we provide a way to better leverage the vast amount of detection methods provided by the community. As shown in our experiments, our method can reliably select for each input image the best detection result or the good results from a range of different methods.

2 Feature Design

In this section, we discuss what makes a good salient object detection result. According to previous research as well as our observation, we measure the quality

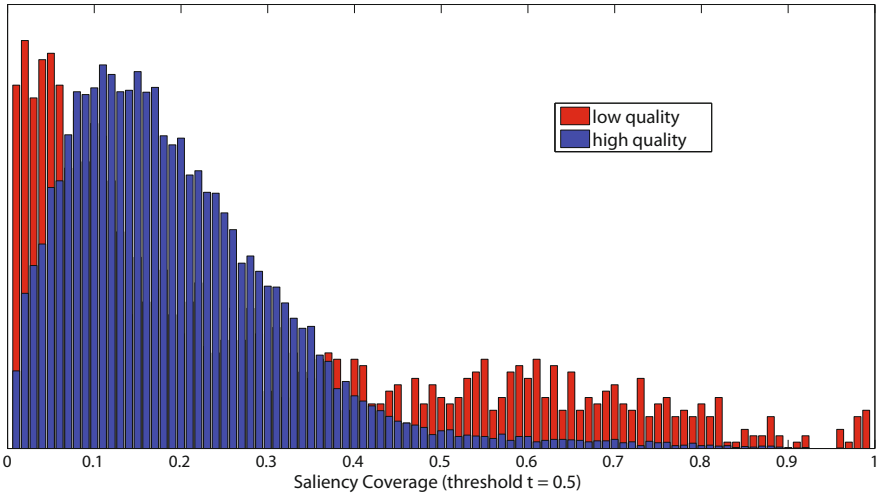


Fig. 2. Saliency coverage distribution for low-quality and high-quality salient object detection results

of salient object detection through the analysis of the saliency map itself as well as the interaction between the saliency map and the original image. Below we describe how we design features to capture the aspects of a good saliency map in detail.

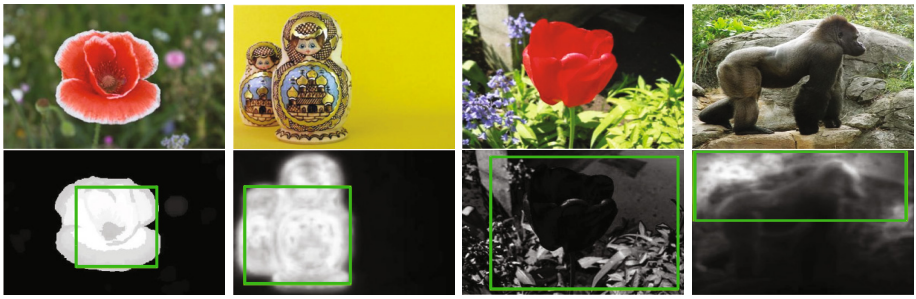
Saliency Coverage. The sizes of salient objects in natural images often fall into a certain range. Therefore, when a saliency map has its salient pixels covering an abnormally large or small area in the image, it is unlikely to be a good map. We design the saliency coverage feature f_C to encode this knowledge.

Given a saliency map M (whose value is scaled to the range $[0, 1]$), we binarize it using a threshold value $t \in [0, 1]$ and compute the saliency coverage value as

$$f_C(M) = \frac{1}{n} \sum_{i \in M} \delta\{M(i) > t\} \tag{1}$$

where n denotes the number of pixels in M. $\delta\{.\}$ is an indicator function whose value is 1 if its argument is true and 0 otherwise.

To examine the ability of this feature in discriminating between high- and low-quality saliency maps, we conduct a small experiment. We randomly collect 3,000 images from the salient object detection benchmark THUS-10000 [7]. For each image, we apply five different saliency detection methods GC [8], HC [8], LC [55], GBMR [54], and HS [53] and create 15,000 saliency maps in total. We then form two groups of images out of those saliency maps according to their AUC score computed using the ground truth. The high-quality group contains only good saliency maps with AUC score at least 0.9. The low-quality group contains only the results with AUC score below 0.6.



$f_{CP} = 0.75$ $f_{CP} = 0.73$ $f_{CP} = 0.19$ $f_{CP} = 0.23$

Fig. 3. Saliency map compactness feature. The green windows contain 75 % of total saliency values. A good saliency map tends to have a high average saliency value inside the enclosing window.

For each map from these two groups, we compute the saliency coverage value (with threshold $t = 0.5$). We show the normalized histograms on the saliency coverage value over each of these two groups in Fig. 2. This figure demonstrates that the high-quality and low-quality saliency maps have clearly different distributions, which shows the capability of the saliency coverage feature in discriminating the high- and low-quality saliency maps.

To make our system robust against the choice of the threshold t, we use 10 different threshold values in the range between 0 and 1. For each threshold value, we compute a corresponding saliency coverage value. Finally, we concatenate the 10 saliency coverage values into a feature vector.

Saliency Map Compactness. Previous research on saliency analysis shows that a good saliency map should concentrate its salient pixels in a compact region in the image [7, 16, 26]. To encode this observation, we estimate how dense the salient pixel distribution is in the most salient area predicted by the saliency map. Specifically, we first compute the minimal image window W_p covering the proportion p of the total saliency in the saliency map M. We then compute the compactness value f_{CP} as the average saliency value inside W_p. In computing W_p, we resize the saliency map to 100 x 100 to reduce the search space. We then use the integral image method [48] to speed up the computation of the total saliency in a candidate window in each search step, which involves only 3 additions and subtractions in total. For $p = 0.5$, W_p can typically be found in 0.02 seconds on a desktop machine with an i7 3.40 GHz CPU.

As demonstrated in Fig. 3, good saliency maps tend to have a high average enclosed saliency value. To make our method robust against the selection of the parameter value for p, we compute the compactness values using multiple values for p and concatenate them all into a single feature vector. In this paper, we use three p values, namely 0.25, 0.5, and 0.75.

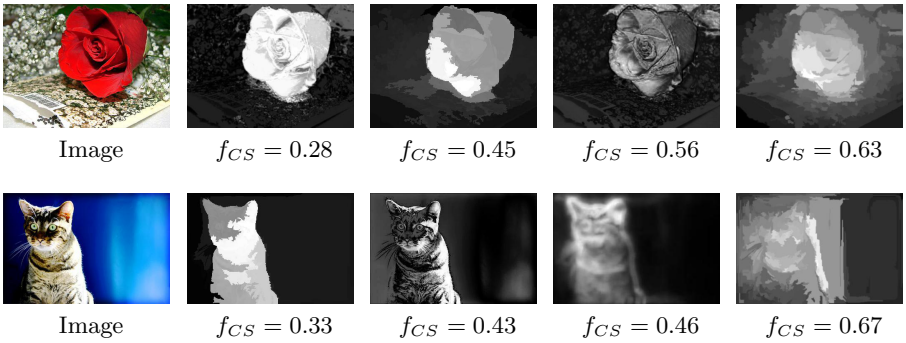


Fig. 4. Color separation feature. Good saliency maps can well separate the color distributions of salient regions and background regions. They thus give small f_{CS} values.

Saliency Histogram. The distribution of saliency values in the saliency map can also be an indicator of its quality. For example, in salient object detection, a good saliency map should well separate salient objects from the image background. This favors saliency histograms with concentrated peaks at two ends. On the other hand, a saliency map where most of the pixels have middle-range saliency values is unlikely to be a good map as it is mostly fuzzy. To capture this observation, we compute the normalized 20-bin histogram over saliency values in the saliency map and use it as our feature f_H.

Color Separation. Many of the previous saliency analysis methods use a global contrast assumption [1,8,14]. This assumption suggests that salient regions and background regions tend to be different from each other. We design the feature f_{CS} to score a saliency map according to the separation between the color distributions of salient regions and non-salient regions based on the saliency map.

We use color histograms to model the color distributions. As saliency often provides a soft assignment of image pixels into salient and background regions, we adopt a weighting-based technique to incorporate the saliency values into the computation of the color histograms. In this paper, we use the histograms with 16 bins per-channel in the RGB color space.

Let b_i denote the color range of the i^{th} bin, the salient region's normalized color histogram h_s can be computed as

$$h_s(i) = \frac{\sum_{p \in I} M(p) \delta\{I(p) \in b_i\}}{\sum_{p \in I} M(p)} \tag{2}$$

where $M(p)$ and $I(p)$ denote the saliency value and color at pixel p. δ is the indicator function defined previously. Similarly, we compute the normalized background color histogram h_g as

$$h_g(i) = \frac{\sum_{p \in I} (1 - M(p)) \delta\{I_p \in b_i\}}{\sum_{p \in I} (1 - M(p))} \tag{3}$$

We estimate the color separation feature f_{CS} as the intersection between the histograms for the salient and background region as follows,

$$f_{CS}(M) = \frac{1}{n_h} \sum_{i=1}^{n_h} \min(h_s(i), h_g(i)) \tag{4}$$

where n_h denotes the number of bins in the histograms. Fig. 4 compares different saliency maps generated from the same images. The examples show that good saliency maps tend to have smaller feature values (i.e. the salient and background color distributions are better separated) than those less accurate ones.

Segmentation Quality. Salient object detection should give a good object segmentation result. We design the feature f_{NC} to measure the quality of a saliency map by assessing the segmentation result it induces.

We first obtain the induced segmentation by binarizing the saliency map with a threshold value t, which partitions an image into the salient region S_t and background region B_t. According to the normalized-cut image segmentation method [46], a good segmentation result should maximize the intra-region similarity while minimizing the inter-region similarity. The estimation for the segmentation quality can then be computed using the normalized-cut energy function.

$$f_{NC}(S_t, B_t) = \frac{\sum_{i \in S_t, j \in B_t, j \in N(i)} w_{ij}}{\sum_{i \in S_t, j \in N(i)} w_{ij}} + \frac{\sum_{i \in S_t, j \in B_t, j \in N(i)} w_{ij}}{\sum_{i \in B_t, j \in N(i)} w_{ij}} \tag{5}$$

where $N(i)$ denotes the set of pixels neighboring to i. w_{ij} represents the color similarity between the neighboring pixels i and j. Following [31], we compute w_{ij} as

$$w_{ij} = exp(-\eta ||I_i - I_j||) \tag{6}$$

where I_i and I_j represent the RGB color values at the pixels i and j, respectively. $||.||$ denotes the L_2 norm. η is set as $\eta = 2 < ||I_i - I_j||^2 >^{-1}$ [31], where $< . >$ denotes the expectation operator. For the sake of robustness, we compute the feature values using three different values for the threshold t, namely 0.5, 0.75, and 0.95, and concatenate them into the feature vector.

Boundary Quality. Good saliency maps should provide accurate and well defined object boundaries. Therefore, the object boundary reflected in the saliency map M should align well with strong edges in the input image I. We design the boundary quality feature f_B to encode this observation. Specifically, given the input image I and its saliency map M, we compute the boundary map B_M from M and measure how well it correlates with the strong edge map E_I of the image I.

We generate E_I using the structured-forests edge detection method [11]. We then compute the saliency boundary map B_M from M as

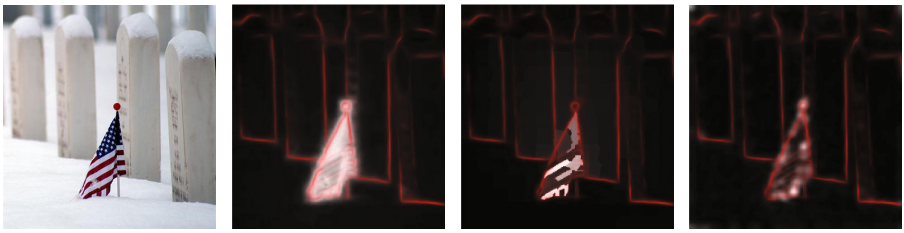


Image $f_B = 0.6$ $f_B = 0.21$ $f_B = 0.15$

Fig. 5. Boundary-quality feature. The strong edge maps have been overlayed in red onto the saliency maps. The middle saliency map has low f_B value because its boundary does not correlate well with strong image edges, while the right-most saliency map has low f_B value because it fails to provide a well-defined boundary.

$$B_M(p) = \frac{w_p |M(p_1) - M(p_2)|}{\sum_{p \in M} w_p} \tag{7}$$
$$w_p = M(p) \max(|M(p) - M(p_1)|, |M(p) - M(p_2)|)$$

where p_1 and p_2 denote two neighboring pixels of p. Those two neighboring pixels are taken orthogonal to edge direction at p. w_p computes the saliency-weighted edge magnitude in the saliency map.

The factor $|M(p_1) - M(p_2)|$ in Equation 7 weighs the boundary map computation based on how well the edge point separates the salient region and the background region. A large value of the factor indicates that the boundary is well defined.

The boundary-quality feature f_B measures the correlation between B_M and E_I using the dot product between two edge maps.

$$f_B(M, I) = \sum_{p \in I} B_M(p) E_I(p) \tag{8}$$

As shown in Fig. 5, better saliency maps tend to have higher boundary-quality values. The saliency map in the third column fails because a portion of its predicted boundary has a low edge value, which results in the low correlation between B_M and E_I. In the right most saliency map, as most of the edge points do not separate salient region and background region, the boundary map B_M was suppressed, which suggests the low-quality boundary and thus leads to a small boundary-quality value.

As edge detection often depends on image scale, we compute the above boundary-quality value at each of four different image scales, namely 0.25, 0.5, 0.75, and 1.0 and concatenate them into the feature vector.

3 Salient Object Detection Ranking

Given an image I and its k saliency maps $\{M_i\}_{i=1..k}$, our goal is to rank these maps according to their quality without any ground truth. In this paper, we

adopt the pairwise-based learning-to-rank methodology [5, 12, 20]. Specifically, we compare every pairs among the k saliency maps and then aggregate these pairwise comparison results to produce the final overall ranking.

Before describing our method, let us describe how we obtain the labeled trainining data. For each image in the dataset, we use k different salient object detection methods to generate k saliency maps. We then use the salient object mask provided with each image to measure the quality of each saliency map. In this paper, we use the popular Area Under the ROC Curve (AUC score) as the objective quality measurement for saliency maps. We obtain the ground-truth ranking for each image by ranking its k saliency maps according to their AUC scores.

We now elaborate our pairwise-based ranking method. Given two saliency maps M_i and M_j generated for the same image, we model the pairwise preference (quality comparison) as the probability that M_i has higher quality than M_j

$$P_{M_i,M_j} = P(M_i \succ M_j | f_{M_i}, f_{M_j}) \tag{9}$$

where $M_i \succ M_j$ means that M_i has higher quality than M_j. f_M denotes the feature vector extracted from the map M using the feature extraction method described in Section 2.

To obtain the pairwise preference model P_{M_i,M_j}, we consider it as a binary classifier. For a given pair of saliency maps M_i and M_j, the feature vector $f_{M_iM_j}$ is created by concatenating f_{M_i} and f_{M_j}. We train a binary classifier C that takes $f_{M_iM_j}$ as input and outputs the preference label 1 if $M_i \succ M_j$, and 0 otherwise. The output of C can be used as an estimation for P_{M_i,M_j}. To train the classifier, we obtain the preference labels for every pairs of saliency maps from each image in the dataset using its ground-truth map ranking computed previously.

In this paper, we experimented with three binary classification methods, including Random Forest Classifier (RFC) [3], Support Vector Machine (SVM) [47], and Multi-Layer Perceptron (MLP) [17]. Specifically, for SVM, we use the RBF-kernel probabilistic SVM implemented in the LIBSVM package [6]. For RFC, we use the Random Forest implementation for MATLAB from Jaiantilal et al.[1]. For MLP, we use the MATLAB Neural Network Toolbox implementation[2] to train an MLP network with one hidden layer. All the models' hyperparameters are selected automatically via cross validation.

Ranking Prediction. Once the pairwise preference model has been trained, it can be used for ranking salient object detection results on new images. Specifically, given an image I and a set $\{M_i\}_{i=1..k}$ of k saliency maps generated from I, we compute a relative score for each saliency map M_i as

$$r(M_i) = \sum_{j=1..k, j\neq i} P_{M_i,M_j} \tag{10}$$

[1] `https://code.google.com/p/randomforest-matlab/`
[2] `http://www.mathworks.com/help/nnet/index.html`

The overall ranking for every M_i's can then be obtained by sorting to their scores.

4 Experiments

We experiment with our method on the public salient object detection benchmark THUS-10000 [7]. This dataset contains 10,000 images. Each image is associated with a manually segmented salient object mask. For each of the experiments in this section, we randomly select from the dataset 2,000 images for training our ranking model, and use the remaining 8,000 images for testing. For all experiments, we repeat this random partition 10 times and report the average results.

In our experiments, we use ten state-of-the-art salient object detection methods, including GBVS [16], IT [23], FT [1], GC [8], HC [8], LC [55], SR [19], CA [14], GBMR [54], HS [53]. For each image in the dataset, we use the codes published by the authors of those methods to generate ten saliency maps. We then perform the ranking on the resulting saliency maps for each image.

4.1 Ranking Accuracy

We examine the quality of the our ranking results by assessing how well they correlate with the ground-truth ranking. To the best of our knowledge, our work is the first to approach the problem of ranking salient object detection results. To serve as the baseline for our comparison, we consider the ranking strategy that assigns a fix ranking to every input image. That fix ranking is computed based on the performance of each salient object detection method over the whole training set. In particular, we compare our method against the following two baseline ranking methods.

Mean-AUC-Based Ranking (MAR) : In this baseline ranking method, the mean AUC score over the whole training set is computed for each method. The MAR ranking is obtained by sorting the methods according to their mean AUC scores. The resulted ranking is applied for all testing images.

Voting-Based Ranking (VBR) : In this method, each image in the training set casts a vote for the saliency method that works best for it (according to the AUC score). The resulted ranking is then obtained by sorting the methods according to their number of votes over the whole training set.

We implement both methods and compare their performance to that from our image-specific ranking method.

Correlation with Ground-Truth Ranking. To evaluate how well our ranking results agree with the ground-truth ranking, we compute their rank correlation. In particular, we experiment with two rank correlation metrics.

Table 1. Rank correlation

	Kendall τ correlation	Weighted Kendall τ correlation
MAR Baseline Ranking	0.49	0.74
VBR Baseline Ranking	0.48	0.73
Our Method (RFC)	0.62	0.81
Our Method (MLP)	0.64	0.83
Our Method (SVM)	**0.65**	**0.83**

Kendall τ Rank Correlation : The Kendall τ rank correlation [27] is one of the most well known method for comparing ranking results [32]. Given a set of elements $S = \{s_i, i = 1..n\}$ and two ranking functions r_1 and r_2, the Kendall τ rank correlation is computed as

$$\tau(r_1, r_2) = 1 - \frac{2\sum_{i,j} \delta\{r_1(i,j) \neq r_2(i,j)\}}{n(n-1)} \tag{11}$$

where δ denotes the indicator function. $r(i,j)$ outputs 1 if the ranking function r gives i the higher rank than j, and 0 otherwise. This metric penalizes a pair of elements if their relative orders given by the two ranking functions disagree.

Weighted Kendall τ Rank Correlation : Inspired by [32], we also experiment with the weighted Kendall τ rank correlation metric

$$\tau(r_1, r_2) = 1 - \frac{2\sum_{i,j} w_{ij}\delta\{r_1(i,j) \neq r_2(i,j)\}}{n(n-1)} \tag{12}$$

where the weight w_{ij} is defined as

$$w_{ij} = max\{AUC(i), AUC(j)\}|AUC(i) - AUC(j)| \tag{13}$$

Intuitively, the weighted Kendall τ rank correlation reduces the penalty on the discordant pairs of maps when their AUC scores are close to each other. At the same time, it emphasizes the penalty on the pairs containing high-quality maps.

Table 1 shows the average rank correlation on the test data. The results show that the saliency map ranking from our method has significantly higher correlation with the ground-truth ranking than those from the baseline methods.

Rank-*n* Accuracy. In this experiment, we examine the effectiveness of our ranking results for the task of retrieving the best saliency map for a given image. For evaluation, we measure the rank-n accuracy.

Given the ranking results for all testing images, the rank-n accuracy is computed as the percentage of the test data for which the actual best method is ranked within the top n positions.

Fig. 6 shows rank-1, rank-2, rank-3, and rank-4 accuracy from our methods, as well as those from the baseline ranking methods. From the figure, we can see

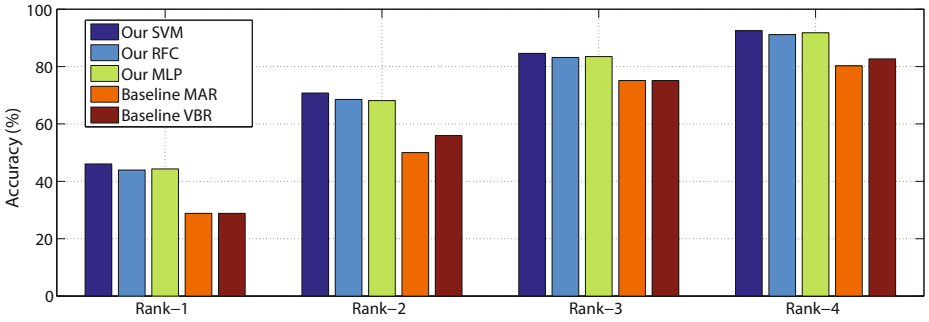


Fig. 6. Rank-n accuracy. The ranking predicted by our model is significantly better than the baseline ranking in selecting the best salient object detection results.

that the ranking results from our methods can provide significantly better best map prediction accuracy than those from the baseline ranking methods.

Discussion. To further evaluate the robustness of our saliency map ranking method, we perform an additional test. For each image in the dataset, we randomly select the ground-truth mask from another image to use as its additional saliency map. This map can be considered a "noisy" individual map as it is a good map on its own but is likely inaccurate with respect to the input image. With the resulting new set of individual maps, we redo all the experiments described previously in this section.

Examining the results, we observe that the ranking accuracy is almost unaffected by the inclusion of the additional map. Specifically, we obtain the rank-1 accuracy of 44.8%, the average Kendall τ rank correlation of 0.66, and the weighted Kendall τ rank correlation of 0.83. This result shows that our method is robust against the inclusion of such a "noisy" individual map.

4.2 Saliency Map Selection Quality

We now further evaluate the ability of our method in selecting the best saliency map given an input image. While rank-1 assessment can give the absolute accuracy, it gives little insight for the cases where the system fails to predict the true best map. In practice, there are scenarios where an incorrectly selected saliency map is still useful as long as its quality is close to that of the true best map.

To take that into account, we consider an alternative method to evaluate the best-map selection quality. In particular, a best map selection result is considered correct if the AUC score of the selected saliency map differs from that of the true best map by no more than a tolerance value ϵ. The accuracy of a saliency map selection method can then be measured as the percentage of all testing images for which the map selection are correct according to this condition. We call this the tolerance-based accuracy. For comparison, we consider the baseline best-map selection methods that always pick a single saliency detection algorithm to use for all images. In particular, we experiment with three methods that always

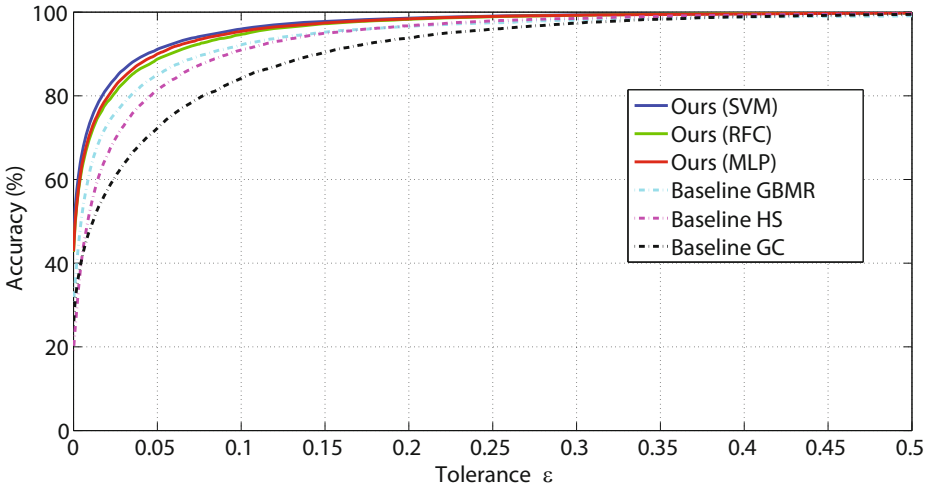


Fig. 7. Tolerance-based best map selection accuracy. At the high tolerance values, our results produce noticeably smaller number of errors than selecting any single method to use for all images. Note that the larger the tolerance value is, the more severe an error would be.

select the GBMR, HS, and GC saliency map, respectively. We choose those three methods for our comparison as they have the highest mean AUC scores over the whole dataset.

Table 2 shows the tolerance-based accuracy at $\epsilon = 0.02$ (i.e. two percent of the AUC range). This result shows that for 82% of all testing images our method can select the good saliency maps that are close to the ground-truth best maps.

Table 2. Tolerance-based accuracy at $\epsilon = 0.02$

	Our SVM	Our RF	Our MLP	Baseline GBMR
Accuracy (%)	**82**	79.5	80	72

We note that the higher the tolerance threshold is, the more severe an incorrect map selection would be. We provide in Fig. 7 the curve representing the tolerance-based best-map selection accuracy measured over the wide range of tolerance values ϵ. The curves show that our methods make noticeably smaller number of errors with small values of ϵ compared to the baseline methods.

4.3 Salient Object Detection Performance

Although our main goal in this paper is not to develop a new salient object detection method, it is interesting to investigate how our method can be used to improve the overall performance of individual methods. For each testing image,

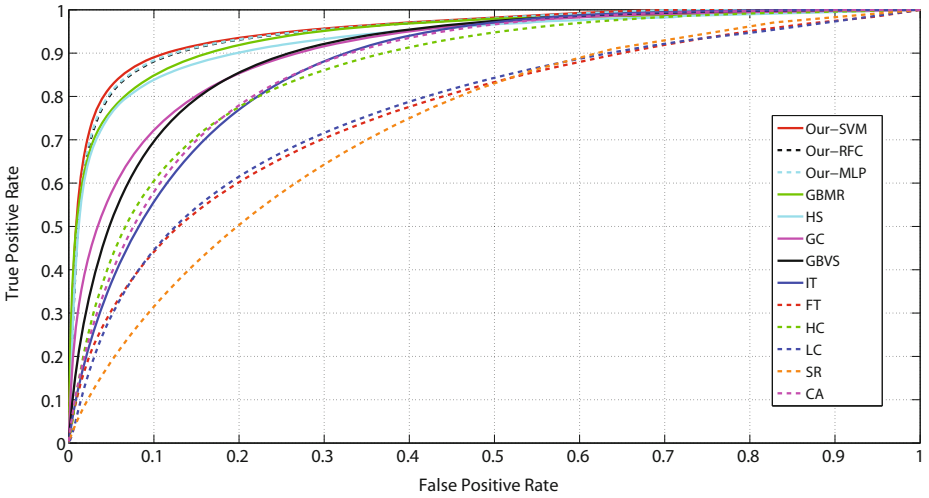


Fig. 8. ROC curves for different salient object detection methods

we use the saliency map ranking obtained for that image to select the best saliency map and use it as our final salient object detection result.

Fig. 8 shows the ROC curves measured from the saliency maps generated using different methods. The figure shows that using the saliency maps selected by our methods, we can improve the ROC curves over each individual method. This confirms the observation that although some methods clearly dominate the others over a large number of images, selecting the image-dependent best saliency map for each specific image is promising in pushing forward the salient object detection performance.

Limitation. One limitation of our method is that it requires to run all the individual salient object detectors for each input image. Some of the individual detectors are slow, which makes our method computationally expensive in terms of running time. This limits the applicability of our method compared to some fast individual methods. However, we note that in practice we can pick the best one from only a few fast and statistically best-performed detectors instead of using all available detectors. This will significantly speed up our method as our method typically takes only 20 seconds to execute both the feature extraction step and the saliency map ranking step.

5 Conclusion

In this paper, we develop a data-driven approach for comparing salient object detection results without knowing the ground truth. We designed a wide range of dedicated features to capture the saliency map quality. Those features are used in our learning-to-rank framework to produce the saliency map ranking for

each input image. Experiments on the large salient object detection benchmark show that our method can produce ranking results that correlate well with the ground-truth ranking. Our method can be used to adaptively select good saliency maps for each input image, which can improve the overall salient object detection performance.

Acknowledgement. This work was supported in part by NSF grants IIS-1321119, CNS-1205746, and CNS-1218589.

References

1. Achanta, R., Hemami, S., Estrada, F., Süsstrunk, S.: Frequency-tuned salient region detection. In: IEEE Conference on Computer Vision and Pattern Recognition (2009)
2. 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)
3. Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)
4. Bulling, A., Alt, F., Schmidt, A.: Increasing the security of gaze-based cued-recall graphical passwords using saliency masks. In: SIGCHI International Conference on Human Factors in Computing Systems (2012)
5. Cao, Z., Qin, T., Liu, T.Y., Tsai, M.F., Li, H.: Learning to rank: From pairwise approach to listwise approach. In: International Conference on Machine Learning (2007)
6. Chang, C.C., Lin, C.J.: Libsvm: A library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, 27:1–27:27 (2011)
7. Cheng, M.M., Warrell, J., Lin, W.Y., Zheng, S., Vineet, V., Crook, N.: Efficient salient region detection with soft image abstraction. In: IEEE International Conference on Computer Vision (2013)
8. Cheng, M., Zhang, G., Mitra, N.J., Huang, X., Hu, S.: Global contrast based salient region detection. In: IEEE CVPR (2011)
9. Christopoulos, C., Skodras, A., Ebrahimi, T.: The jpeg2000 still image coding system: An overview. IEEE Trans. on Consumer Electronics 46(4), 1103–1127 (2000)
10. Cohen, Y., Basri, R.: Inferring region salience from binary and gray-level images. Pattern Recognition 36, 2349–2362 (2003)
11. Dollár, P., Zitnick, C.L.: Structured forests for fast edge detection. In: IEEE International Conference on Computer Vision (2013)
12. Fürnkranz, J., Hüllermeier, E.: Pairwise preference learning and ranking. In: Lavrač, N., Gamberger, D., Todorovski, L., Blockeel, H. (eds.) ECML 2003. LNCS (LNAI), vol. 2837, pp. 145–156. Springer, Heidelberg (2003)
13. Gao, D., Vasconcelos, N.: Bottom-up saliency is a discriminant process. In: IEEE International Conference on Computer Vision (2007)
14. Goferman, S., Zelnik-manor, L., Tal, A.: Context-aware saliency detection. In: IEEE Conference on Computer Vision and Pattern Recognition (2010)
15. Hall, D., Leibe, B., Schiele, B.: Saliency of interest points under scale changes. In: British Machine Vision Conference (2002)
16. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: NIPS, pp. 545–552 (2006)

17. Haykin, S.: Neural Networks: A Comprehensive Foundation, 2nd edn. Prentice Hall PTR, Upper Saddle River (1998)
18. Holtzman-gazit, M., Zelnik-manor, L., Yavneh, I.: Salient edges: A multi scale approach. In: European Conference on Computer Vision Workshop on Vision for Cognitive Tasks (2010)
19. Hou, X., Zhang, L.: Saliency detection: A spectral residual approach. In: IEEE Conference on Computer Vision and Pattern Recognition (2007)
20. Hüllermeier, E., Fürnkranz, J., Cheng, W., Brinker, K.: Label ranking by learning pairwise preferences. Artificial Intelligence 172(16-17), 1897–1916 (2008)
21. Itti, L., Koch, C.: Computational modeling of visual attention. Nature Reviews Neuroscience 2, 194–203 (2001)
22. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research 40, 1489–1506 (2000)
23. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. 20, 1254–1259 (1998)
24. Jiang, B., Zhang, L., Lu, H., Yang, C., Yang, M.H.: Saliency detection via absorbing markov chain. In: IEEE International Conference on Computer Vision (2013)
25. Jiang, Z., Davis, L.S.: Submodular Salient Region Detection. In: IEEE Conference on Computer Vision and Pattern Recognition (2013)
26. Judd, T., Ehinger, K., Durand, F., Torralba, A.: Learning to predict where humans look. In: IEEE International Conference on Computer Vision (2009)
27. Kendall, M.G.: A new measure of rank correlation. Biometrika 30(1/2), 81–93 (1938)
28. Li, X., Li, Y., Shen, C., Dick, A., Hengel, A.V.D.: Contextual hypergraph modeling for salient object detection. In: IEEE International Conference on Computer Vision (2013)
29. Li, X., Lu, H., Zhang, L., Ruan, X., Yang, M.H.: Saliency detection via dense and sparse reconstruction. In: IEEE International Conference on Computer Vision (2013)
30. Liu, F., Gleicher, M.: Region enhanced scale-invariant saliency detection. In: 2006 IEEE International Conference on Multimedia and Expo (2006)
31. Liu, T., Sun, J., Zheng, N.N., Tang, X., Shum, H.Y.: Learning to detect a salient object. In: IEEE Conference on Computer Vision and Pattern Recognition (2007)
32. Liu, Y., Wang, J., Cho, S., Finkelstein, A., Rusinkiewicz, S.: A no-reference metric for evaluating the quality of motion deblurring. ACM Trans. Graph. 32(6), 175:1–175:12 (2013), http://doi.acm.org/10.1145/2508363.2508391
33. Mai, L., Niu, Y., Liu, F.: Saliency aggregation: A data-driven approach. In: IEEE Conference on Computer Vision and Pattern Recognition (2013)
34. Marchesotti, L., Cifarelli, C., Csurka, G.: A framework for visual saliency detection with applications to image thumbnailing. In: IEEE International Conference on Computer Vision (2009)
35. Margolin, R., Zelnik-Manor, L., Tal, A.: Saliency for image manipulation. The Visual Computer, 1–12 (2012)
36. Ming, Y., Li, H., He, X.: Winding number for region-boundary consistent salient contour extraction. In: IEEE Conference on Computer Vision and Pattern Recognition (2013)
37. Niu, Y., Geng, Y., Li, X., Liu, F.: Leveraging stereopsis for saliency analysis. In: IEEE Conference on Computer Vision and Pattern Recognition (2012)
38. Oikonomopoulos, A., Patras, I., Pantic, M.: Spatiotemporal salient points for visual recognition of human actions. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 36(3), 710–719 (2005)

39. Parikh, D., Zitnick, C.L., Chen, T.: Determining patch saliency using low-level context. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 446–459. Springer, Heidelberg (2008)
40. Perazzi, F., Krhenbl, P., Pritch, Y., Hornung, A.: Saliency filters: Contrast based filtering for salient region detection. In: IEEE Conference on Computer Vision and Pattern Recognition (2012)
41. Ramanathan, S., Katti, H., Sebe, N., Kankanhalli, M., Chua, T.-S.: An eye fixation database for saliency detection in images. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 30–43. Springer, Heidelberg (2010)
42. Rubinstein, M., Gutierrez, D., Sorkine, O., Shamir, A.: A comparative study of image retargeting. ACM Trans. Graph. 29, 160:1–160:10 (2010)
43. Rudoy, D., Goldman, D.B., Shechtman, E., Zelnik-Manor, L.: Learning video saliency from human gaze using candidate selection. In: IEEE Conference on Computer Vision and Pattern Recognition (2013)
44. Rutishauser, U., Walther, D., Koch, C., Perona, P.: Is bottom-up attention useful for object recognition (2004)
45. Sheikh, H., Bovik, A., Cormack, L.: No-reference quality assessment using natural scene statistics: Jpeg2000. IEEE Transactions on Image Processing 14(11), 1918–1927 (2005)
46. Shi, J., Malik, J.: Normalized cuts and image segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition (1997)
47. Vapnik, V.N.: The nature of statistical learning theory. Springer-Verlag New York, Inc., New York (1995)
48. Viola, P.A., Jones, M.J.: Rapid object detection using a boosted cascade of simple features. In: CVPR, vol. (1), pp. 511–518 (2001)
49. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing 13(4), 600–612 (2004)
50. Wei, Y., Wen, F., Zhu, W., Sun, J.: Geodesic saliency using background priors. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 29–42. Springer, Heidelberg (2012)
51. Wu, C., Frahm, J.-M., Pollefeys, M.: Detecting large repetitive structures with salient boundaries. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part II. LNCS, vol. 6312, pp. 142–155. Springer, Heidelberg (2010)
52. Xie, Y., Lu, H., Yang, M.H.: Bayesian saliency via low and mid level cues. IEEE Transactions on Image Processing 22(5), 1689–1698 (2013)
53. Yan, Q., Xu, L., Shi, J., Jia, J.: Hierarchical saliency detection. In: IEEE Conference on Computer Vision and Pattern Recognition (2013)
54. Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M.H.: Saliency detection via graph-based manifold ranking. In: IEEE Conference on Computer Vision and Pattern Recognition (2013)
55. Zhai, Y., Shah, M.: Visual attention detection in video sequences using spatiotemporal cues. In: ACM Multimedia (2006)
56. Zhao, R., Ouyang, W., Wang, X.: Person re-identification by salience matching. In: IEEE International Conference on Computer Vision (2013)
57. Zhao, R., Ouyang, W., Wang, X.: Unsupervised salience learning for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition (2013)