

Saliency detection on sampled images for tag ranking

Jingfan Guo · Tongwei Ren ✉ · Lei Huang · Jia Bei

Received: date / Accepted: date

Abstract Image saliency contributes to rank the unordered tags extracted from social media, but the existing saliency detection methods can hardly efficiently handle massive images in tag ranking. In this paper, we focus on improving the efficiency of saliency detection methods by applying them on the sampled images with suitable resolutions. We extensively investigate the influence of image resolution to saliency detection performance of the typical methods, and summarize a sampling strategy for different categories of salient object detection methods. Furthermore, we validate the effectiveness of the sampling strategy by applying the salient object detection methods on the sampled images with the selected resolutions in tag ranking. The experimental results show that sampling can significantly improve the efficiency of the existing salient object detection methods without obvious loss in effectiveness.

Keywords Saliency detection · tag ranking · image annotation · sampling

1 Introduction

The explosive growth of web images requires effective retrieval technology for acquiring the desirable images on the Internet [1–4]. Due to the existence of

Jingfan Guo
State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
E-mail: guojf@smail.nju.edu.cn

Tongwei Ren ✉
State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
E-mail: rentw@nju.edu.cn

Lei Huang
Software Institute, Nanjing University, Nanjing, China
E-mail: leihuang@nju.edu.cn

Jia Bei
Software Institute, Nanjing University, Nanjing, China
E-mail: beijia@nju.edu.cn

semantic gap, content-based image retrieval remains a challenging problem though it has been widely investigated on various features [5–8]. Instead, retrieving images by their tags provides an alternative solution, which can index the images by tags and retrieve them by a text query [9,10]. Considering the intensive labor cost in manual labeling, automatic image annotation has attracted great attention from numerous researchers on multimedia and computer vision [11]. One solution of automatic image annotation is directly assigning tags to images according to their content by the pre-trained classifiers [12–16]. However, these methods usually require sufficient training data for each tag to obtain satisfactory performance. Another solution is extracting the image tags from users’ interaction, such as image title, tag and description, which are ubiquitous in social media in the Web 2.0 era [17,18]. The extracted tags demonstrate user-perceived visual semantics of image content, but they usually suffer the problem of low quality [19]. For example, the tags provided by users in social media are generally unordered, which cannot emphasize the important content in images. Therefore, series of techniques are proposed for tag ranking to improve the effectiveness of image tags in retrieval [20].

Two strategies are commonly used in tag ranking, tag relevance ranking and tag saliency ranking. Tag relevance ranking measures the relevance between a tag and an image based on whether the tag is relevant to the images similar to the given image in visual representation [21]. The performance of such methods is sensitive to the scale of image data set, i.e., the performance will obviously descend when the image data set is small, and the similarity measurement of image visual presentation is also challenging, e.g., two images with the same dominant object but different background may be measured with low similarity. In contrast, tag saliency ranking determines the tag order of a given image according to the saliency of the corresponding image regions of the tags [22]. Image saliency denotes the degree of image regions attracting human attention [23–25], which has been used in numerous multimedia applications, such as image editing [26,27], object classification [28,29] and surveillance analysis [30,31]. Tag saliency ranking reduces the requirement of a large-scale and well-tagged image data set, and determines the tag order of a given image based on its own content.

Obviously, detecting image saliency play a key role in tag saliency ranking. The existing saliency detection methods mainly include two categories, fixation prediction and salient object detection. Compared to fixation prediction [23], salient object detection aims to provide complete salient objects with consistent saliency value within each object [32], which is more suitable for tag saliency ranking. In the past few years, amounts of salient object detection methods have been proposed. Most methods focuses on improving the effectiveness of saliency detection results, including some saliency detection methods specifically proposed for tag ranking [33,34], but the efficiency improvement of salient object detection is seldom concerned.

In the existing salient object detection methods, one common strategy for efficiency improvement is using super-pixel instead of pixel as the basic

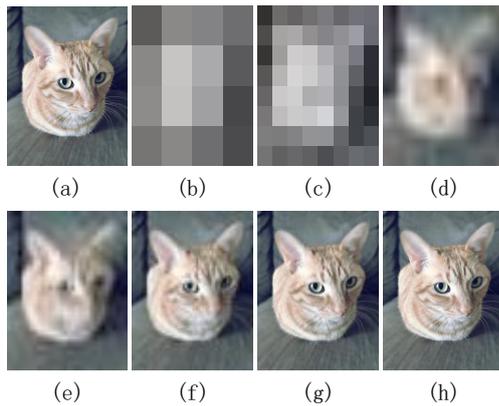


Fig. 1 Salient object is recognizable in diverse resolutions. (a) Original image in the resolution of $2,000 \times 1,500$. (b) - (h) Sampled images with 4×4 , 8×8 , ..., 256×256 resolutions represented with the original size and aspect ratio.

processing unit in saliency calculation, which can obviously reduce the number of processing units from more than half million to only several hundreds. It has been used in a large number of saliency detection methods [35–37]. Some other strategies are also used to improve the efficiency of salient object detection. For example, Zhang *et al.* [38] propose *FastMBD*, an approximate iterative algorithm for the Minimum Barrier Distance transform which takes advantage of the raster scanning technique, to efficiently detect salient object in pixel-level. Cheng *et al.* [32] uses color quantization for each region to speed up histogram calculation and comparison. Based on the above efforts, the time cost of current salient object detection methods can be reduced to less than 0.1 second per image [32] on public data sets.

However, the existing salient object detection methods still suffer the efficiency problems when using them in real world applications. Firstly, the images in such applications usually have much higher resolution than those in public data sets. The resolutions of images captured by cameras are usually more than ten million pixels, but those in public data sets are only around one hundred thousand pixels, in which the former is about one hundred times larger than the later. Even using super-pixel as the processing unit in saliency calculation, the time costs of super-pixel generation on the images with such distinct resolutions are quite different. Secondly, for massive images are required to be process in social media computing, even slight reduction of time cost in handling each image is sufficiently meaningful, which can accelerate the processing procedure and diminish the requirement of computational platform.

Figure 1 shows an example of resizing an image into diverse resolutions. It is obvious that the region of salient object can be briefly identified when resizing the image into quite low resolutions, e.g., 16×16 , for the dominant color and structure of the image are preserved in sampling. Based on the above observation, we explore the relationship between image resolution and

the performance of the existing salient object detection methods. Hou *et al.* propose a similar observation that 64 pixels of the image width is a good estimation of the scale of normal visual conditions [39], but they only concern their proposed method without a comprehensive study of the existing salient object detection methods with various categories. In this paper, we utilize a public data set *MSRA10K* and construct a high resolution data set *HR100* in our experiments, and resize the images into diverse resolutions. We firstly classify the existing salient object detection methods into three categories and select two typical methods for each category, and then investigate the performance of these selected methods on the images with diverse resolutions to summarize a sampling strategy for different categories of salient object detection methods. To the summarized sampling strategy, we validate its effectiveness by applying the salient object detection methods on the sampled images with the selected resolutions in tag ranking.

Some preliminary results of our method were presented in [40]. In this paper, we additionally analyze the application of saliency detection efficiency improvement in tag ranking, and briefly survey the typical tag ranking methods. We also present more details of the investigation of the influence of image resolution to saliency detection performance. Moreover, we validate the summarized sampling strategy in tag ranking on a subset of a public data set *NUS-WIDE*.

Our major contribution can be summarized as follows:

- We extensively investigate the influence of image resolution to saliency detection performance of the typical saliency detection methods of different categories, and summarize a sampling strategy for these salient object detection methods.
- We validate the effectiveness of the sampling strategy by applying the salient object detection methods on the sampled images with the selected resolutions in tag ranking.
- We construct a high resolution image data set *HR100* with manually labeled salient objects, which is used to evaluate the efficiency of the existing saliency detection methods on diverse image resolutions.

The rest of the paper is organized as follows. We briefly review the typical methods in tag ranking and saliency detection in Section 2. Then, we introduce the data sets used in our experiments in Section 3. The experiments and analysis of the influence of image resolution to saliency detection performance are presented in Section 4, and the validation of the sampling strategy in tag ranking is shown in Section 5. Finally, we conclude our work in Section 6.

2 Related Work

2.1 Tag ranking

Current tag ranking methods mainly use two strategies, tag relevance ranking and tag saliency ranking. The former determines the tag order by referring the

images with similar visual representation, and the latter determines the order of tags according to the saliency of their corresponding image regions.

Tag relevance ranking. Tag relevance ranking measures the relevance between a tag and an image based on whether the tag is relevant to the images similar to the given image in visual representation. Li *et al.* [21] learn tag relevance by a neighbor voting algorithm, in which the relevance between a tag and a given image is obtained by k nearest neighbor of the image based on visual similarity. Liu *et al.* [20] initialize the relevances between tags and images with a probabilistic algorithm and refine them by random walk on tag graph. Zhuang *et al.* [41] exploit the correlations between tags and images with a two-view learning methods using both textual and visual content. Tang *et al.* [42] combine saliency detection in tag relevance ranking by considering the relationships among images based on both the whole image and the salient regions and estimating the relevance of each tag with regard to a given image on both image level and region level. To obtain good performance, tag relevance ranking methods require large-scale data sets and well-defined visual representation similarity.

Tag saliency ranking. Tag saliency ranking determines the tag order of a given image according to the saliency of the corresponding regions of the tags with in the given image. Feng *et al.* [22] firstly propose the concept of tag saliency ranking and implement it with an improved multi-instance learning algorithm to reassign the tags to image regions. Wang *et al.* [33] iteratively boost saliency detection and tag ranking instead of detecting image saliency only using visual content. Feng *et al.* [43] combine tag relevance ranking and tag saliency ranking into an unified framework by pre-classified with a linear SVM, in which the images with salient objects are processed with tag saliency ranking and other images are processed with tag relevance ranking. Cao *et al.* [34] extend tag saliency ranking to stereo images by improving region segmentation, saliency detection and multi-instance learning. Tag saliency ranking methods are usually dependent on the performance of saliency detection and assignment of tags to image regions.

2.2 Salient object detection

According to the difference of processing unit, the existing salient object detection methods can be roughly classified into two categories, pixel-level methods and region-level methods. The former directly operates on original image pixels without any abstraction. On the contrary, the latter segments the input image into regions and treats these regions as the basic processing units in saliency detection.

Pixel-level salient object detection. Pixel-level saliency detection methods are widely investigated for the full control over each pixel. Achanta *et al.* [44] model pixel-level saliency as Euclidean distance between color of each pixel and average color of entire image on $L^*a^*b^*$ color space. In [45], saliency is computed based on the idea of maximal symmetric surround. Zhang *et*

al. [38] use image boundary connectivity cue to calculate pixel-level saliency in a highly efficient way. However, pixel-level methods have to tightly restrict the computation complexity for each pixel considering the massive pixels in one image. As a consequence, the above pixel-level methods could only extract simplex features from each pixel.

Region-level salient object detection. Super-pixel is a powerful technique to generate regions for it is able to retain the intrinsic structure of images [46]. Cheng *et al.* [32] model saliency value of each super-pixel as spatially weighted color contrast to other super-pixels. Jiang *et al.* [36] calculate saliency value of each super-pixel by measuring absorbing time in a Markov chain. Yang *et al.* [35] present a graph-based manifold ranking approach to measure saliency of super-pixels. Zhu *et al.* [37] propose an optimization framework to integrate multiple cues, including boundary connectivity and color contrast, to produce saliency maps. Ju *et al.* [47] use the anisotropic center-surround approach to model salient object in depth images. Guo *et al.* [48] propose a evolution strategy to detect salient object in RGB-D images. These region-based methods usually treat each region as an operation unit, and extract simple features [35, 36], such as average color and region center location, or complex features [32, 37], such as color histogram and boundary connectivity, from each unit.

3 Data Sets

Two data sets are used in the study of the influence of image resolution on salient object detection, including *MSRA10K* and *HR100*. *MSRA10K* consists of 10,000 images with the pixel-wise ground truths manually labeled by multiple participants, which is a large-scale and widely used public data set for salient object detection [32]. Nevertheless, the resolutions of the images in *MSRA10K* are around 300×400 , which are quite lower than the images captured by smart phones and cameras in daily life. It leads to the fact that experiments on *MSRA10K* cannot show the efficiency improvement on the sampled images than on the original images. Hence, we construct a high resolution data set *HR100*. It consists of 100 high resolution images from the Internet, and the larger one between width and height of each image is up to 2,000 pixels. To construct the data set, five participants are invited to label the regions of the most attractive object in each image by *Adobe Photoshop*, and the pixels labeled by more than three participants are considered within the salient objects. Fig. 2 shows the examples of the images in *MSRA10K* and *HR100* with their resolutions and the corresponding ground truths of salient objects, in which the top two rows are the images in *MSRA10K* and the bottom two rows are the images in *HR100*.

In the validation of the effectiveness of the sampling strategy in tag ranking, we use a public data set *NUS-WIDE* [49]. *NUS-WIDE* includes 269,648 images with 5,018 tags crawled from *Flickr*, and it also provides the manually labeled ground truths with 81 concepts. In our experiments, we construct a subset

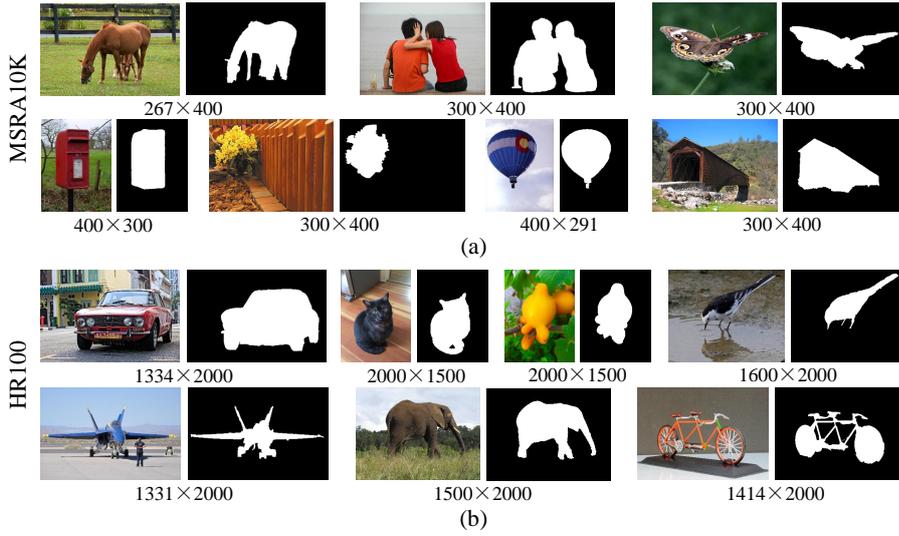


Fig. 2 Examples of the images with the corresponding resolutions and manually labeled salient objects in *MSRA10K* and *HR100*. (a) *MSRA10K*. (b) *HR100*.

of *NUS-WIDE* by randomly selecting 20,000 images. In the subset of *NUS-WIDE*, all of 81 concepts and 4,997 of 5,018 tags appear. For the tags are directly crawled from *Flickr*, the problem of tag omission is serious, which is out of the scope of this paper. So we complement the tag list of each image by merging its concepts, i.e., the combination of the concepts and tags of each image is used as its new tag list, and only consider tag ranking in our experiments. We count up the appearance times of each tag, and only retain the tags if their appearance times are no less than a threshold, which equals 10 in our experiments. Finally, all 81 concepts and 3058 tags are retained, and each image has 2.40 concepts and 10.02 tags on average. Fig. 3 shows some examples of the images with the corresponding tags and manually labeled concepts in the subset of *NUS-WIDE*.

image	concept	tag	image	concept	tag	image	concept	tag
	clouds, sky, lake, water, reflection,	vacation, lake, mountains, water, sunrise, island, quality, glacier, clouds, reflection, sky		buildings, house, sky, window	historic, house, buildings, sky, window		dancing, person	girls, hot, beautiful, asian, singapore, dancing, person
	animal, beach, dog, water	dog, beach, wet, water, mar, wave, animal		flowers, garden, plants	blue, holland, nature, spring, colours, flowers, garden, plants, netherlands		clouds, lake, sky, water	lake, art, water, boat, sunrise, clouds, sky
	clouds, military, plane, sky, sunset	sunset, plane, airplane, airport, force, aircraft, aviation, air, clouds, military, sky		clouds, grass, sky	blue, shadow, sky, tree, green, grass, wales, fence, landscape, clouds, countryside		animal, sand	animal, sand

Fig. 3 Examples of the images with the corresponding tags and manually labeled concepts in the subset of *NUS-WIDE*.

Table 1 Typical saliency detection methods of pixel-level (PL), region-level with simple features (RLSF) and region-level with complex features (RLCF).

Category	Method	Language
PL	FT [44]	Matlab
	MSSS [45]	Matlab
RLSF	MC [36]	Matlab
	GMR [35]	Matlab
RLCF	RBD [37]	Matlab
	RC [32]	C++

4 Saliency Detection on Sampled Images

We select six typical saliency detection methods of three categories, including pixel-level (PL), region-level with simple features (RLSF) and region-level with complex features (RLCF), in which two methods are selected for each category. Table 1 shows the selected methods and their categories and implementation languages.

In our experiments, we sample the input images to certain resolutions to reduce the influence of aspect ratio to image resolution. The sampled resolutions include 4×4 , 8×8 , 16×16 , 32×32 , 64×64 , 128×128 , and 256×256 . The generated saliency maps with the same resolution as the sampled input images are resized to the original resolution of the images using bilinear interpolation and further evaluated by comparing them with the manually labeled ground truths. Fig. 4 shows some examples of saliency detection results on *MSRA10K* and *HR100*, in which the top example is from *MSRA10K* and the bottom example is from *HR100*. In each example, the original image is shown in the bottom left with its resolution, and the saliency detection results of different methods on diverse resolutions are shown.

We use precision-recall (PR) curves to evaluate the performance of each method. A PR curve is generated by comparing the ground truth against the binary masks generated by thresholds sliding from 0 to 255. In addition, weighted F_β -measure (F_β^ω) [50] is also used in our performance evaluation:

$$F_\beta^\omega = (1 + \beta^2) \frac{P^\omega \cdot R^\omega}{\beta^2 \cdot P^\omega + R^\omega}, \quad (1)$$

where P^ω and R^ω are weighted precision and weighted recall, respectively [50]; β^2 is a parameter to equally treat precision and recall, which is set to 1 in our experiments.

All the experiments are executed on a computer with 3.5GHz CPU and 8GB memory. The algorithm implementations used in our experiments are provided by the original authors. In Table 2 to 3, the en-dash (–) represents the corresponding method fails to generate the saliency maps while sampling the images to such resolutions. And the bold values in Table 2 indicate the best performance in each column.

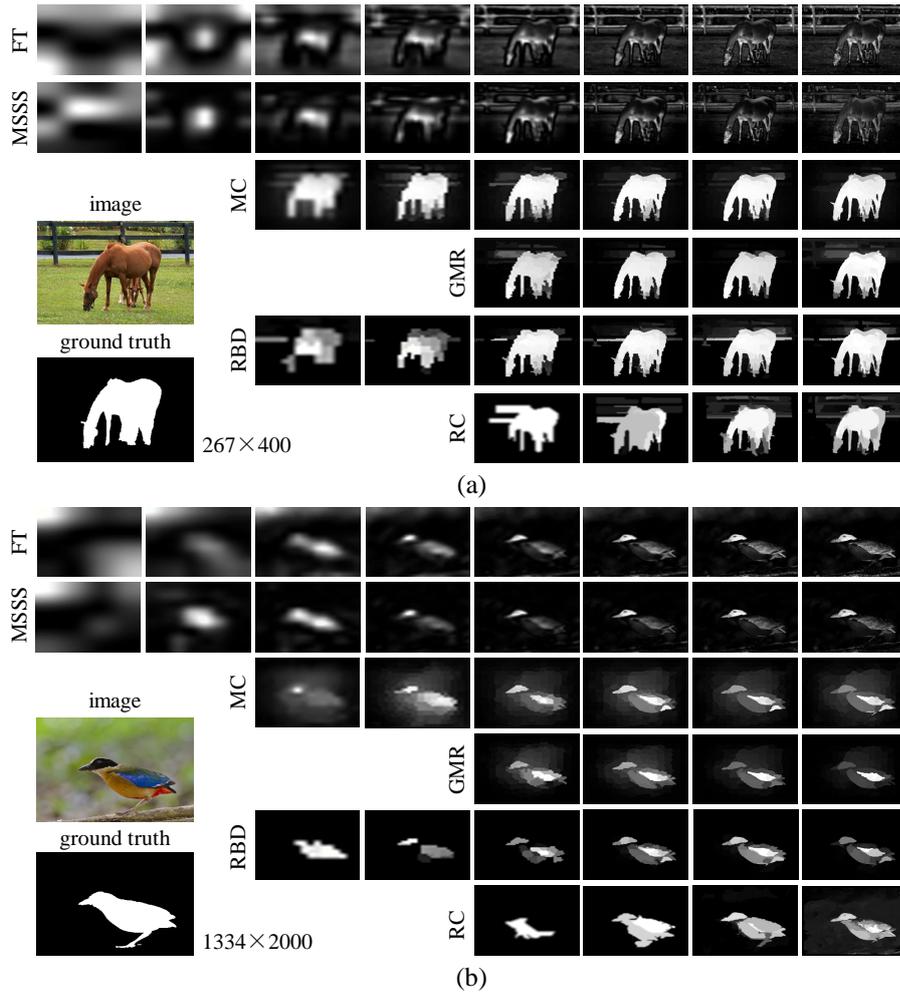


Fig. 4 Examples of saliency detection results on *MSRA10K* and *HR100*. (a) *MSRA10K*. (b) *HR100*.

4.1 Sampling for pixel-level methods

The first experiment is conducted to study the influence of sampling for pixel-level methods, including FT [44] and MSSS [45].

As shown in Fig. 5 and the first two columns of Table 2, the performance of these methods are keeping stable or slightly decreasing when the sampled image resolutions are higher than 16×16 . It shows that the pixel-level methods can obtain the acceptable performance on quite low sample resolutions.

The mean values in the first two columns in Table 3 show that FT and MSSS become much more efficient while decreasing the image resolutions. The standard variances of the running time on the sampled images are small, for

Table 2 Effectiveness evaluation with F_{β}^{ω} on diverse image resolutions.

	FT [44]		MSSS [45]		MC [36]		GMR [35]		RBD [37]		RC [32]		
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	
MSRA10K	4 × 4	0.1541	0.1016	0.2282	0.1493	–	–	–	–	–	–	–	
	8 × 8	0.2457	0.1331	0.3318	0.1274	–	–	–	–	–	–	–	
	16 × 16	0.2770	0.1371	0.3394	0.1223	0.5273	0.1747	–	–	0.5143	0.2288	–	–
	32 × 32	0.2866	0.1377	0.3264	0.1258	0.5673	0.1934	–	–	0.4468	0.2194	–	–
	64 × 64	0.2919	0.1384	0.3157	0.1274	0.5827	0.1986	0.6150	0.2228	0.6217	0.2062	0.5070	0.3376
	128 × 128	0.2988	0.1391	0.3109	0.1280	0.5869	0.2030	0.6221	0.2283	0.6702	0.1944	0.6490	0.2016
	256 × 256	0.3068	0.1393	0.3009	0.1279	0.5851	0.2047	0.6252	0.2295	0.6758	0.1934	0.6344	0.1794
	original	0.3101	0.1392	0.3094	0.1276	0.5835	0.2089	0.6238	0.2309	0.6726	0.1982	0.6076	0.1768
HR100	4 × 4	0.1225	0.0959	0.1797	0.1320	–	–	–	–	–	–	–	
	8 × 8	0.1773	0.1137	0.2486	0.1230	–	–	–	–	–	–	–	
	16 × 16	0.2061	0.1107	0.2599	0.1153	0.4201	0.1558	–	–	0.3287	0.2049	–	–
	32 × 32	0.2271	0.1133	0.2669	0.1106	0.4374	0.1614	–	–	0.3129	0.2084	–	–
	64 × 64	0.2369	0.1060	0.2633	0.1016	0.4667	0.1637	0.5128	0.1742	0.4810	0.1896	0.3777	0.3009
	128 × 128	0.2493	0.1077	0.2618	0.1023	0.4813	0.1638	0.5269	0.1761	0.5443	0.1838	0.5291	0.2093
	256 × 256	0.2580	0.1111	0.2591	0.1087	0.4724	0.1673	0.5254	0.1831	0.5495	0.1818	0.5361	0.1580
	original	0.2732	0.1160	0.2564	0.1116	0.4481	0.1711	0.4902	0.1886	0.5346	0.1800	0.4517	0.1672

each group of sampled images have the same resolution. It means that then efficiency improvement by sampling is stable.

4.2 Sampling for region-level methods

This experiment consists of two parts for we divide region-level methods into two categories, namely with simplex features and with complex features. Specifically, the methods with simplex features are MC [36] and GMR [35], while the methods with complex features are RBD [37] and RC [32]. Among these region-level methods, the ones based on SLIC [46] are set to have 150 super-pixels as a trade-off between efficiency and effectiveness.

Different from Section 4.1, the lowest resolution in this experiment is 16×16 instead of 4×4 . The reason is that generating 150 super-pixels requires at least 150 pixels, but 4×4 and 8×8 do not satisfy the requirement. Moreover, due to the limitation of the original implementation, GMR and RC fail to generate saliency maps when the input resolution is lower than 64×64 .

Figure 6 and 7 and the last four columns in Table 2 show that the performance of the region level methods on the sampled images. Similar to the pixel level methods, the region level methods can obtain the acceptable performance on low sample image resolutions. The required minimum sampled resolutions, such as 64×64 or 128×128 , are slightly higher than the ones for the pixel level methods, for too small regions cannot provide sufficient information for feature extraction.

An interesting phenomenon is that all of the region-level methods perform better on relatively low resolutions instead of the original ones. It is caused by the fact that the segmented region number is fixed to be 150 in each methods during the experiment, and too many pixels in a region may increase noise

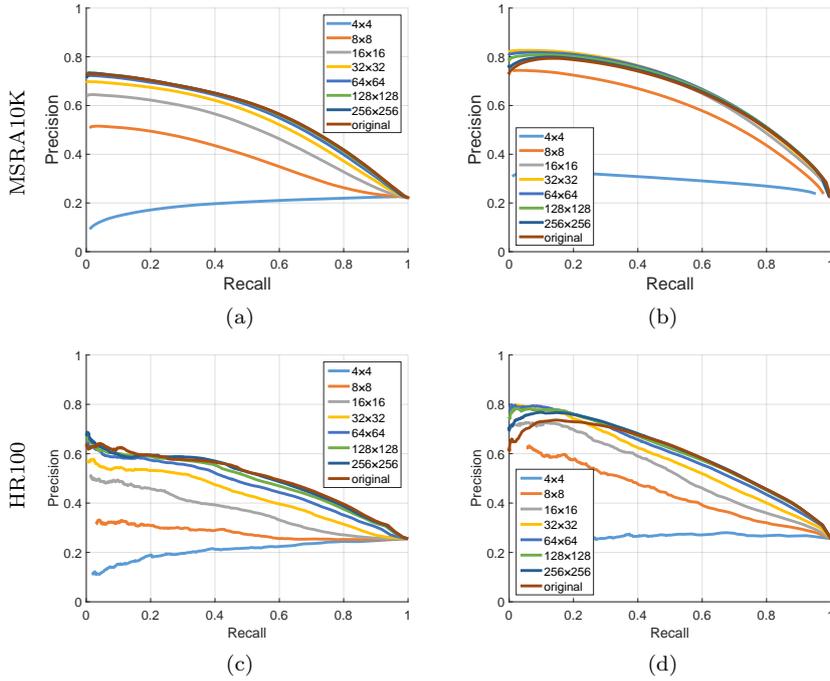


Fig. 5 Influence of sampling for pixel-level methods. (a) and (c) FT [44]. (b) and (d) MSSS [45].

and complexity instead of more information in feature extraction. Therefore, we could learn that the performance of region-level methods rely on the segmentation results. When applying the existing region-level method to high resolution images, it is necessary to increase the segmentation number at first.

From the mean value of the last four columns in Table 3, we can find that all of these region-level methods also obtain efficiency improvement by sampling.

4.3 Sampling strategy

By summarizing the experiments, we come to some conclusions about the relationship between image resolutions and the performance of salient object detection methods. Obviously, there is a trade-off between the efficiency and effectiveness in salient object detection depending on the sampled image resolution. The running time of all the salient object detection methods decreases on the images with lower resolutions, i.e., sampling could improve the efficiency of salient object detection. On the other hand, as for effectiveness, low sampled image resolution may cause the decline of salient object detection performance.

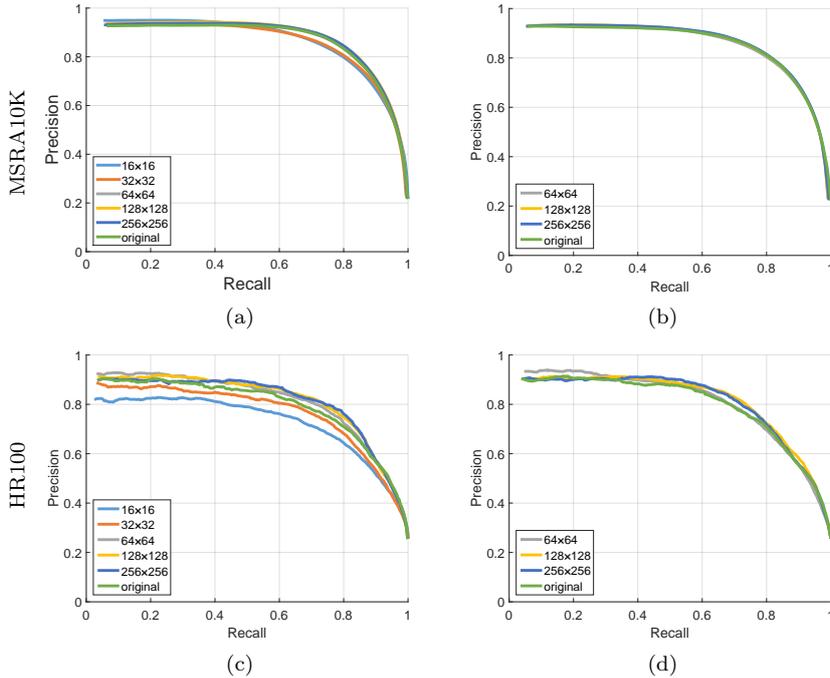


Fig. 6 Influence of sampling for region-level methods with simplex features. (a) and (c) MC [36]. (b) and (d) GMR [35].

The top two rows of Table 4 show the required minimum sampled resolutions for all the images while retaining the certain percentages of the F_{β}^{ω} values under the original image resolutions on MSRA10K and HR100, respectively. For example, the “16×16” at the intersection of line “pixel level” and column “75%” means at least one pixel level method cannot obtain 75% F_{β}^{ω} value of the one under the original resolution on one or more images if the sampled image resolution is less than “16×16”. It shows that the required minimum sampled resolutions for the pixel level methods increase slightly when the required percentage of the F_{β}^{ω} values increases. Compared to the pixel level methods, the region level methods require higher but more stable minimum resolutions. Nevertheless, we can find that all the salient object detection methods are tolerant of image sampling, i.e., they can obtain quite similar performance on low sampled image resolutions, such as 128×128 , and the original image resolutions. The reason is that sampling may retain the dominant color and structure of image content while removing the unnecessary details, which will benefit salient object detection. In the last row of Table 4, we suggest the sampled resolutions for all categories of methods under different percentages of the F_{β}^{ω} values to satisfy various original image resolutions, which can be used as an effective sampling strategy for the efficiency improvement of salient object detection in real applications.

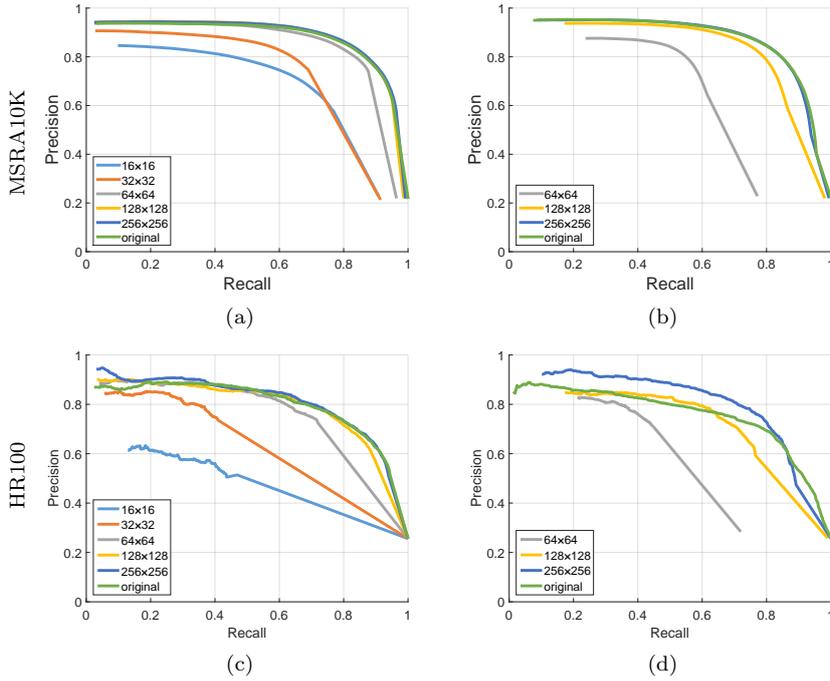


Fig. 7 Influence of sampling for region-level methods with complex features. (a) and (c) RBD [37]. (b) and (d) RC [32].

To illustrate the efficiency improvement by sampling, we choose 85% of the F_{β}^{ω} values on the original resolutions as an example, which leads to the minimum resolution 128×128 for region-level methods with complex features and it decreases to 64×64 for pixel-level methods and region-level methods with simple features. The efficiency improvements under other F_{β}^{ω} percentages can be analyzed in the same way. We use these two resolutions as the suggested resolution in our sampling strategy, i.e., sampling the images to 64×64 when using pixel-level methods and region-level methods with simple features and to 128×128 when using region-level methods with complex features. It means that all the salient object detection methods can obtain quite similar performance when decreasing image resolution to less than 1% of the original resolution. According to Table 3, the time cost of all the methods on the sampled images with the above resolutions can reduce 48% to 90% on *MSRA10K* and 98% to 99% on *HR100*. Hence, sampling could serve as a potential solution of efficiency problem when using the existing salient object detection methods in real word applications.

Table 3 Efficiency evaluation with running time on diverse image resolutions.

	FT [44]		MSSS [45]		MC [36]		GMR [35]		RBD [37]		RC [32]		
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	
MSRA10K	4 × 4	0.0123	0.0062	0.0054	0.0006	–	–	–	–	–	–	–	
	8 × 8	0.0108	0.0030	0.0056	0.0007	–	–	–	–	–	–	–	
	16 × 16	0.0110	0.0017	0.0059	0.0006	0.0128	0.0018	–	–	0.0177	0.0056	–	–
	32 × 32	0.0118	0.0021	0.0070	0.0003	0.0051	0.0010	–	–	0.0358	0.0063	–	–
	64 × 64	0.0138	0.0023	0.0106	0.0008	0.0086	0.0016	0.0738	0.0072	0.0469	0.0076	0.0506	0.1918
	128 × 128	0.0172	0.0028	0.0223	0.0016	0.0160	0.0028	0.1091	0.0067	0.0529	0.0068	0.0588	0.1461
	256 × 256	0.0397	0.0027	0.0711	0.0029	0.0469	0.0070	0.2226	0.0114	0.1071	0.0092	0.1032	0.2281
	original	0.0658	0.0082	0.1101	0.0184	0.0675	0.0112	0.2520	0.0240	0.1278	0.0181	0.1147	0.1175
HR100	4 × 4	0.0142	0.0009	0.0096	0.0537	–	–	–	–	–	–	–	
	8 × 8	0.0102	0.0014	0.0043	0.0006	–	–	–	–	–	–	–	
	16 × 16	0.0105	0.0013	0.0046	0.0007	0.0121	0.0076	–	–	0.0169	0.0254	–	–
	32 × 32	0.0114	0.0015	0.0052	0.0009	0.0095	0.0019	–	–	0.0306	0.0035	–	–
	64 × 64	0.0139	0.0009	0.0079	0.0007	0.0120	0.0022	0.0892	0.0051	0.0415	0.0046	0.0370	0.0144
	128 × 128	0.0189	0.0010	0.0174	0.0028	0.0192	0.0024	0.1150	0.0060	0.0591	0.0057	0.0445	0.0143
	256 × 256	0.0442	0.0019	0.0540	0.0076	0.0492	0.0053	0.2123	0.0093	0.1067	0.0087	0.0783	0.0156
	original	1.3848	0.1768	2.1502	0.2782	1.6477	0.2053	5.0167	0.5748	2.6256	0.3275	2.6274	0.8030

Table 4 The required minimum sampled resolutions for retaining the certain percentages of the F_{β}^{ω} values under the original image resolutions.

	Category	Minimum sampled resolution for diverse F_{β}^{ω} percentages				
		70%	75%	80%	85%	90%
MSRA10K	PL	4 × 4	8 × 8	16 × 16	16 × 16	32 × 32
	RLSF	64 × 64	64 × 64	64 × 64	64 × 64	64 × 64
	RLCF	64 × 64	64 × 64	64 × 64	128 × 128	128 × 128
HR100	PL	16 × 16	16 × 16	32 × 32	64 × 64	128 × 128
	RLSF	64 × 64	64 × 64	64 × 64	64 × 64	64 × 64
	RLCF	64 × 64	64 × 64	64 × 64	128 × 128	128 × 128
suggested	PL	16 × 16	16 × 16	32 × 32	64 × 64	128 × 128
	RLSF	64 × 64	64 × 64	64 × 64	64 × 64	64 × 64
	RLCF	64 × 64	64 × 64	64 × 64	128 × 128	128 × 128

5 Validation in Tag Ranking

We validate the sampling strategy for improving the efficiency of saliency detection methods in tag ranking. Inspired by the existing tag saliency ranking methods [22, 34], we first need to build a semantic mapping between tags of images and segmented regions of images. Since tags are associated with images instead of segmented regions in *NUS-WIDE* data set, multi-instance learning provides a good approach to model such weakly-supervised learning problem. Specifically, we select RBF-MIP [51] as the multi-instance learning algorithm in our experiments for its effectiveness and efficiency. Note here, the selection of multi-instance learning algorithm has little influence on our validation, for we validate the performance of saliency detection results on different resolutions in tag ranking based on the same tag assignment result from images to regions.

In RBF-MIP algorithm, the labels are associated with the bags, which are the groups of the instances. We treat each segmented region as an instance, so that an image, which is a group of segmented regions, is represented as a bag, and a tag annotated to an image is treated as a bag-level label. In this way, our data setting conforms to the requirement of RBF-MIP.

For each segmented region of an image, we extract a 21-dimension feature vector consisting the following features: region location, region size, average RGB color, standard variance of RGB color, Gabor magnitude with twelve orientations. The dimensions of the above features are 2, 1, 3, 3, and 12, respectively.

Given some images with tags, RBF-MIP trains a two-layer neural network which responds to segmented regions. When we input the feature vector of a segmented region, the trained network outputs the probability of each tag assigned to this region.

We randomly select 16,000 images from the subset of *NUS-WIDE* with 20,000 images as the training data, and use the rest 4,000 images as the test data. We train an RBF-MIP network for all the tags appearing in the data set. For each region in a given image, we assign it with a tag which has the highest probability in the output of RBF-MIP network. The sum of saliency of all assigned regions to a tag in an image is treated as the weight of the tag in the image. We determine the order of tags in each image according to their weights, and evaluate the performance of tag ranking on each image by comparing with the concepts of the image. We calculate the average precision of tag ranking result on the i th image as follows:

$$AP_i = \frac{1}{N_i} \sum_{k=1}^{N_i} \frac{p_k^i}{k}, \quad (2)$$

where AP_i is the average precision of tag ranking result on the i th image; N_i is the number of concepts of the i th image; p_k^i is the number of tags in the top- k tags belong to the concepts of the i th images.

We calculate the mean value of average precision on all the images, i.e., mean average precision (MAP), on test data, and compare the mean average precision generated by different saliency detection results. Table 5 shows the comparison results, in which the performance of each saliency detection method in the original image resolution and the suggested sampled resolution by the summarized sampling strategy is compared in tag ranking. We can find that the saliency detection method can achieve similar performance in the suggested sampled resolution to those in the original image resolution in tag ranking, even slightly better in some conditions. Fig. 8 shows some examples of tag ranking results with saliency detection on the sampled images, in which the first row shows the images, the second row and the third row show the corresponding concepts and tags respectively, and the rest rows show the tag ranking results by using the saliency detection results with different methods on the sampled images. It shows that the saliency detection methods can obtain good performance in the suggested sampled resolution in tag ranking.

Table 5 Comparison of the performance of the typical saliency detection methods in the original image resolution and the suggested sampled resolution in tag ranking.

Method	Suggested sampled resolution	MAP (sampled resolution)	MAP (original resolution)
FT [44]	64×64	63.17%	63.14%
MSSS [45]	64×64	65.82%	65.20%
MC [36]	64×64	66.08%	65.48%
GMR [35]	64×64	65.60%	65.77%
RBD [37]	128×128	65.21%	65.25%
RC [32]	128×128	64.55%	64.76%

6 Conclusion

In this paper, we investigate the relationship between image resolution and the performance of saliency detection methods. The extensive experiments show that sampling images to suitable resolutions can obviously improve the efficiency of the existing saliency detection methods, namely pixel-level methods, region-level methods with simplex features and region-level methods with complex features, while retaining their effectiveness. It benefits to apply the existing saliency detection methods in efficiently handling massive images in tag ranking.

Acknowledgements The authors would like to thank the anonymous reviews for their helpful suggestion. This work is supported by National Science Foundation of China (61321491, 61202320), Research Fund of the State Key Laboratory for Novel Software Technology at Nanjing University (ZZKT2016B09), and Collaborative Innovation Center of Novel Software Technology and Industrialization.

References

1. Sang, J., Xu, C.: Right buddy makes the difference: An early exploration of social relation analysis in multimedia applications. In: *ACM International Conference on Multimedia*, ACM (2012) 19–28
2. Zhang, H., Yang, Y., Luan, H., Yang, S., Chua, T.S.: Start from scratch: Towards automatically identifying, modeling, and naming visual attributes. In: *ACM International Conference on Multimedia*, ACM (2014) 187–196
3. Zheng, L., Wang, S., Guo, P., Liang, H., Tian, Q.: Tensor index for large scale image retrieval. *Multimedia Systems* **21**(6) (2015) 569–579
4. Cheng, Z., Shen, J.: On very large scale test collection for landmark image search benchmarking. *Signal Processing* **124** (2016) 13–26
5. Corridoni, J.M., Del Bimbo, A., Pala, P.: Image retrieval by color semantics. *Multimedia Systems* **7**(3) (1999) 175–183
6. Xu, X., Geng, W., Ju, R., Yang, Y., Ren, T., Wu, G.: Obsir: Object-based stereo image retrieval. In: *IEEE International Conference on Multimedia and Expo*, IEEE (2014) 1–6
7. Cao, W., Liu, N., Kong, Q., Feng, H.: Content-based image retrieval using high-dimensional information geometry. *Science China Information Sciences* **57**(7) (2014) 1–11

image			
concept	animal, coral, person, water	buildings, clouds, nighttime, water	animal, birds, lake
tag	canada, vancouver, water, aquarium, interestingness, animal, coral, person,	beautiful, norway, village, quality, buildings, clouds, nighttime, water	lake, nature, birds, flying, wings, flight, landing, animal
FT (64×64)	animal, water, interestingness, person	clouds, village, water, nighttime	animal, birds, lake
MSSS (64×64)	animal, water, aquarium, interestingness	clouds, water, village, beautiful	animal, birds, lake
MC (64×64)	person, aquarium, water, animal	water, nighttime, clouds, buildings	birds, animal, lake
GMR (64×64)	person, animal, water, coral	water, clouds, beautiful, nighttime	animal, flying, lake
RBD (128×128)	animal, person, water, interestingness	clouds, village, buildings, water	animal, lake, birds
RC (128×128)	animal, person, aquarium, water	water, clouds, village, buildings	animal, birds, wings

Fig. 8 Examples of tag ranking results with saliency detection on the sampled images on the subset of *NUS-WIDE* with 20,000 images.

8. Zhu, L., Shen, J., Xie, L.: Unsupervised visual hashing with semantic assistant for content-based image retrieval. *IEEE Transactions on Knowledge and Data Engineering* **29**(2) 472 – 486
9. Zhao, N., Hong, R., Wang, M., Hu, X., Chua, T.S.: Searching for recent celebrity images in microblog platform. In: *ACM International Conference on Multimedia*, ACM (2014) 841–844
10. Zhang, H., Shang, X., Luan, H., Wang, M., Chua, T.S.: Learning from collective intelligence: Feature learning using social images and tags. *ACM Transactions on Multimedia Computing, Communications, and Applications* **13** (2016)
11. Sang, J., Xu, C., Liu, J.: User-aware image tag refinement via ternary semantic analysis. *IEEE Transactions on Multimedia* **14**(3) (2012) 883–895
12. Zhong, S.H., Liu, Y., Liu, Y.: Bilinear deep learning for image classification. In: *ACM International Conference on Multimedia*, ACM (2011) 343–352
13. Tang, J., Zha, Z.J., Tao, D., Chua, T.S.: Semantic-gap-oriented active learning for multilabel image annotation. *IEEE Transactions on Image Processing* **21**(4) (2012) 2354–2360
14. Gao, Z., Zhang, L.f., Chen, M.y., Hauptmann, A., Zhang, H., Cai, A.N.: Enhanced and hierarchical structure algorithm for data imbalance problem in semantic extraction under massive video dataset. *Multimedia Tools and Applications* **68**(3) (2014) 641–657

15. Zhang, H., Shang, X., Yang, W., Xu, H., Luan, H., Chua, T.S.: Online collaborative learning for open-vocabulary visual classifiers. In: IEEE International Conference on Computer Vision and Pattern Recognition, ACM (2016)
16. Liu, A.A., Su, Y.T., Nie, W.Z., Kankanhalli, M.: Hierarchical clustering multi-task learning for joint human action grouping and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**(1) (2017) 102–114
17. Sang, J., Xu, C.: Browse by chunks: Topic mining and organizing on web-scale social media. *ACM Transactions on Multimedia Computing, Communications, and Applications* **7**(1) (2011) 30
18. Wang, M., Ni, B., Hua, X.S., Chua, T.S.: Assistive tagging: A survey of multimedia tagging with human-computer joint exploration. *ACM Computing Surveys* **44**(4) (2012) 25
19. Gao, Y., Wang, M., Zha, Z.J., Shen, J., Li, X., Wu, X.: Visual-textual joint relevance learning for tag-based social image search. *IEEE Transactions on Image Processing* **22**(1) (2013) 363–376
20. Liu, D., Hua, X.S., Yang, L., Wang, M., Zhang, H.J.: Tag ranking. In: International World Wide Web Conference, ACM (2009) 351–360
21. Li, X., Snoek, C.G., Worring, M.: Learning tag relevance by neighbor voting for social image retrieval. In: ACM International Conference on Multimedia Information Retrieval, ACM (2008) 180–187
22. Feng, S., Lang, C., Xu, D.: Beyond tag relevance: integrating visual attention model and multi-instance learning for tag saliency ranking. In: ACM International Conference on Image and Video Retrieval, ACM (2010) 288–295
23. Nguyen, T.V., Xu, M., Gao, G., Kankanhalli, M., Tian, Q., Yan, S.: Static saliency vs. dynamic saliency: a comparative study. In: ACM International Conference on Multimedia, ACM (2013) 987–996
24. Jian, M., Lam, K.M., Dong, J., Shen, L.: Visual-patch-attention-aware saliency detection. *IEEE Transactions on Cybernetics* **45**(8) (2015) 1575–1586
25. Zhong, S.H., Liu, Y., Ng, T.Y., Liu, Y.: Perception-oriented video saliency detection via spatio-temporal attention analysis. *Neurocomputing* (2016)
26. Ren, T., Liu, Y., Wu, G.: Image retargeting based on global energy optimization. In: IEEE International Conference on Multimedia and Expo, IEEE (2009) 406–409
27. Du, H., Liu, Z., Jiang, J., Shen, L.: Stretchability-aware block scaling for image retargeting. *Journal of Visual Communication and Image Representation* **24**(4) (2013) 499–508
28. Wei, Y., Xia, W., Lin, M., Huang, J., Ni, B., Dong, J., Zhao, Y., Yan, S.: Hcp: A flexible cnn framework for multi-label image classification. *IEEE transactions on pattern analysis and machine intelligence* **38**(9) (2015) 1901–1907
29. Lu, Y., Lai, Z., Fan, Z., Cui, J., Zhu, Q.: Manifold discriminant regression learning for image classification. *Neurocomputing* **166** (2015) 475–486
30. Bao, B.K., Liu, G., Xu, C., Yan, S.: Inductive robust principal component analysis. *IEEE Transactions on Image Processing* **21**(8) (2012) 3794–3800
31. Kuang, H., Chong, Y., Li, Q., Zheng, C.: Mutualcascade method for pedestrian detection. *Neurocomputing* **137** (2014) 127–135
32. Cheng, M.M., Mitra, N.J., Huang, X., Torr, P.H.S., Hu, S.M.: Global contrast based salient region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **37**(3) (2015) 569–582
33. Wang, W., Lang, C., Feng, S.: Contextualizing tag ranking and saliency detection for social images. In: International Conference on Multimedia Modeling, Springer (2013) 428–435
34. Cao, Y., Kang, K., Zhang, S., Zhang, J., Wang, Z.: Automatic tag saliency ranking for stereo images. *Neurocomputing* **172** (2016) 9–18
35. 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, IEEE (2013) 3166–3173
36. Jiang, B., Zhang, L., Lu, H., Yang, C., Yang, M.H.: Saliency detection via absorbing markov chain. In: IEEE International Conference on Computer Vision, IEEE (2013) 1665–1672

37. Zhu, W., Liang, S., Wei, Y., Sun, J.: Saliency optimization from robust background detection. In: IEEE Conference on Computer Vision and Pattern Recognition, IEEE (2014) 2814–2821
38. Zhang, J., Sclaroff, S., Lin, Z., Shen, X., Price, B., Měch, R.: Minimum barrier salient object detection at 80 fps. In: IEEE International Conference on Computer Vision, IEEE (2015)
39. Hou, X., Zhang, L.: Saliency detection: A spectral residual approach. In: IEEE Conference on Computer Vision and Pattern Recognition, IEEE (2007)
40. Guo, J., Ren, T., Bei, J.: A comprehensive study of salient object detection on diverse image resolutions. In: National Conference on Multimedia Technology, CCF (2016)
41. Zhuang, J., Hoi, S.C.: A two-view learning approach for image tag ranking. In: ACM International Conference on Web Search and Data Mining, ACM (2011) 625–634
42. Tang, J., Li, M., Li, Z., Zhao, C.: Tag ranking based on salient region graph propagation. *Multimedia Systems* **21**(3) (2015) 267–275
43. Feng, S., Lang, C., Liu, H., Huang, X.: Adaptive all-season image tag ranking by saliency-driven image pre-classification. *Journal of Visual Communication and Image Representation* **24**(7) (2013) 1031–1039
44. Achanta, R., Hemami, S., Estrada, F., Süsstrunk, S.: Frequency-tuned salient region detection. In: IEEE Conference on Computer Vision and Pattern Recognition, IEEE (2009) 1597–1604
45. Achanta, R., Süsstrunk, S.: Saliency detection using maximum symmetric surround. In: IEEE Transactions on Image Processing, IEEE (2010) 2653–2656
46. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(11) (2012) 2274–2282
47. Ju, R., Liu, Y., Ren, T., Ge, L., Wu, G.: Depth-aware salient object detection using anisotropic center-surround difference. *Signal Processing: Image Communication* **38** (2015) 115–126
48. Guo, J., Ren, T., Bei, J.: Salient object detection for rgb-d image via saliency evolution. In: IEEE International Conference on Multimedia and Expo, IEEE (2016)
49. Chua, T.S., Tang, J., Hong, R., Li, H., Luo, Z., Zheng, Y.: NUS-WIDE: a real-world web image database from national university of singapore. In: ACM International Conference on Image and Video Retrieval, ACM (2009) 48
50. Margolin, R., Zelnik-Manor, L., Tal, A.: How to evaluate foreground maps. In: IEEE Conference on Computer Vision and Pattern Recognition, IEEE (2014) 248–255
51. Zhang, M.L., Zhou, Z.H.: Adapting rbf neural networks to multi-instance learning. *Neural Processing Letters* **23**(1) (2006) 1–26