Hindawi
Advances in Multimedia
Volume 2018, Article ID 9497083, 13 pages
https://doi.org/10.1155/2018/9497083

Review Article

Performance Comparison of Saliency Detection

Ning Li, Hongbo Bi, Zheng Zhang, Xiaoxue Kong, and Di Lu

Northeast Petroleum University, School of Electrical and Information Engineering, Development Road, Daqing 163000, China

Correspondence should be addressed to Hongbo Bi; bhbdq@126.com

Received 27 December 2017; Revised 29 March 2018; Accepted 29 April 2018; Published 3 June 2018

Academic Editor: Wei Zhang

Saliency detection has attracted significant attention in the field of computer vision technology over years. At present, more than 100 saliency detection models have been proposed. In this paper, a relatively more detailed classification is proposed. Furthermore, we selected 25 models and evaluated their performance using four public image datasets. We also discussed common problems, such as the influence to model performance by prior information and multiple objects. Finally, we offered future research directions.

1. Introduction

Human vision has the incredible capacity to detect visually distinct objects and regions of interest in these images. Research on human vision can effectively solve application problems in computer vision. And scientists have become interested in its capability to find objects or regions representing a scene, which is generally referred to as saliency detection. Figure 1 depicts the concept of saliency detection. The original images are shown in the first row. In saliency detection, significant objects or regions are determined and distinguished from the background, as shown in the second row in Figure 1.

Generally, the process includes two steps: (1) object location, accurate location is premise, and (2) segment object or region. So far, there have been kinds of segmentation algorithms proposed [1]. The proposed models have solved the first problem well on scenes with the single objects. However, how to accurately detect multiple objects on scenes with high-clutter background is still a problem.

1.1. Definitions. Saliency detection is generally described as the automatic estimation of significant (salient) objects and regions of images without any prior assumption or knowledge. Saliency usually is described as difference between a pixel and its surrounding neighborhood [2]. In addition, saliency models (i.e., models of saliency detection) include visual models, purely computational models, and their combination. Purely computational model does not consider the visual characteristics of the human eye whereas others do. Generally, more attention is on visual models, which can be roughly classified into two categories: attention models (predicting fixation points) and salient region detection models (highlighting the whole salient object areas). The former uses the selective visual attention mechanism to dynamically sample the important visual content in the scene. These models acquire a series of visual fixation points which locate the salient objects. The latter detects and segments the whole object or region. In this paper, we divide the latter into two more detailed categories: salient region detection and salient object detection.

1.2. Origin and Development. Visual attention was investigated in multiple disciplines, such as cognitive physiology, neurology, and computer vision, and had a long history. On the basis of cognitive theories, Treisman and Gelade [3] conceptualized feature integration theory in 1980 and thereafter proposed two stages of visual attention: feature registration and feature integration. However, visual attention only detect independent features (e.g., color, direction, and size) in the first stage for combination in the second stage. Following the theoretical approach, early versions of saliency models that aimed to predict attention points were proposed by Koch and Ullman [4] and Itti et al. [5]. The model explained human visual search strategies. Human visual system (HVS) decomposes a map into a set of topographic feature maps, then feature maps coming from different spatial location

Figure 1: Salient object or region.

are contrasted for saliency. As we know, locations which stand out their surrounding are persist. Finally, all of the feature maps integrate a master saliency map in a complete bottom-top manner. At present, more than 100 models have been proposed, including those for fixation points prediction. Zhang and Sclaroff [6] used a Boolean map topology to analyze a saliency map. However, the objects contours were not observed in their entirety, and thus, the consecutive points established in the visual attention prediction models were inappropriate for subsequent processing. Thereafter, considerable research focused on salient region detection. Many models were developed in this category, some of which were biologically represented. Li et al. [7] integrated a reconstruction error with sparse and dense representation to obtain saliency maps. Zhu et al. [8] constructed a scheme that added a simple boundary a priori different from those in previous models, thereby providing a robust background measure to characterize the spatial layout of image regions with image boundaries. Subsequently, a principal optimization method was proposed to integrate multiple low-level cues (e.g., background measure) and obtain uniform saliency maps. Scharfenberger et al. [9] proposed statistical pattern models, which expressly make full use of intrinsic heterogeneous textural feature of the image and also quantifying the saliency of regions in the image with an efficient mode.

Several models based on mathematical calculations were also developed. Hou and Zhang [11] defined the concept of residual spectrum and preserved an image phase spectrum. Then, saliency maps were developed based on the Fourier transform phase spectra. Salient region detection was conducted to obtain a binary image from the entire contour of the region using the uniform highlights of the dominant object. However, calculations were not continuously and accurately generated, thereby leading to false results and noticeable misses. Nonetheless, the models produced saliency maps with greater detail compared with those generated from previous approaches and thus could be used for salient object detection. However, the generated maps were considered as rough maps (i.e., a much more accurate saliency map could be obtained by subsequent processing). In addition, Achanta et al. [28] combined low-level features to obtain saliency maps with local contrast. These traditional models have achieved an excellent performance; however, models are hard to get better result due to the lack of high-level semantic feature. Recently, the rise of deep learning based on neural network provides a breakthrough for saliency algorithm.

At present, the typical classifications are sorted on the basis of different principles. By contrast, this study pays close attention on salient regions detection. We analyze classic models to obtain a relatively more detailed classification of saliency models. The 25 typical models evaluated in this study are listed in Table 1.

1.3. Application. Saliency detection technique is widely used in the fields of target detection and cognition [32–39], image retrieval [40], object discovery [41], image segmentation [42, 43], video summarization and skimming [44, 45], image and video compression [46], image automation pruning [47], and visual tracking [48–52], etc.

The rest of the paper is organized as follows: Section 2 analyzes and classifies several existing detection models. Section 3 evaluates 25 models. Finally, we provide the summary and recommendation.

2. Algorithm for Saliency Detection Classification and Analysis

From an angle of visual attention mechanism, models are generally divided into two categories. The first category is a rapid, bottom-up, and data-driven approach without semantic information, while the second category include slow, top-down, and task-driven methods with semantic information. At present, most models' research focuses on bottom-up detection. To give a more clearer overview of the existing bottom-up model, in this study, we divided bottom-up models into three categories: (1) attention point prediction, (2) salient region detection, and (3) salient object detection. Research shows that contrasts lead to visual attention, and low-level features such as color, texture, and direction are frequently used to calculate contrasts. Therefore, the contrast-based classification (global contrast and local contrast) is widely accepted. All bottom-up models also depend on several prior knowledge (e.g., a priori background, a priori boundary, and so on). Although prior knowledge varies, classifying models by their a priori foreground and a priori background are also generally acceptable.

2.1. Attention Point Prediction. Attention models have commonly been validated against eye movements of human observation. Eye movements convey important information regarding cognitive processes such as reading, visual search, and scene perception. Therefore, they often are treated as a proxy for shifts of attention [2]. Primates have a strong ability to process complex scenes in real time. Visual system will make choices in the collected information before further processing of visual information, which can greatly reduce the complexity of information, and this selection process is completed in a limited field of view (i.e., visual attention point). We analyze attention point prediction by referring to the physical structure of the human eye. A macular area is found in the retina, and the central depression is called central

Table 1: Classification of saliency detection.

Category			Referred	Model
Attention point prediction			IT [5]	A Model of Saliency-Based Visual Attention for Rapid Scene Analysis
			GB [10]	Graph-based visual saliency
			SR [11]	Saliency detection: A spectral residual approach
			SS [12]	Image Signature: Highlighting sparse salient regions
			COV [13]	Visual saliency estimation by nonlinearly integrating features using region covariances
Salient region detection	Local contrast	Foreground priori	FES [14]	Fast and Efficient Saliency Detection Using Sparse Sampling and Kernel Density Estimation
			CA [15]	Context-aware saliency detection
			DRFI [16]	Salient Object Detection: A Discriminative Regional Feature Integration Approach
			Stru [9]	Structure-guided statistical textural distinctiveness for salient region detection in natural images
		Background priori	RBD [17]	Saliency Optimization from Robust Background Detection
			MAP [18]	Saliency Region Detection Based on Markov Absorption Probabilities
Salient region detection	Global contrast	Foreground priori	HC [19]	Global Contrast based Salient Region Detection
			RC [20]	Global Contrast based Salient Region Detection
			MC [21]	Saliency Detection via Absorbing Markov Chain
			PCA [22]	What Makes a Patch Distinct
		Background priori	DSR [7]	Saliency Detection via Dense and Sparse Reconstruction
Salient region detection	Learning algorithm		AM [23]	Amulet: Aggregating Multi-level Convolutional Features for salient object detection
			BL [24]	Salient Object Detection via Bootstrap Learning
			LS [25]	A Unified Approach to Salient Object Detection via Low Rank Matrix Recovery
			UCF [26]	Learning Uncertain Convolutional Features for Accurate Saliency Detection
Salient object detection			FT [27]	Frequency-tuned salient region detection
			AC [28]	Salient region detection and segmentation
			GC [29]	Efficient Salient Region Detection with Soft Image Abstraction
			MS [30]	Saliency Detection with Multi-Scale Super-pixels
			SF [31]	Saliency Filters: Contrast Based Filtering for Salient Region Detection

fovea. In this area, visual nerve cells are dense; thus, the perceived visual information is accurate. Although the central fovea accounts for only 0.01% of the total visual surface, 10% of the information in the optic nerve is transmitted to the brain from the axons located at the central fovea. Human vision necessitates strong dynamic selectivity when sensing the external environment; in this case, dynamic sampling serves as the process of the visual attention point transfer. Furthermore, human vision can rapidly handle large amounts of image information. The above points explain the biological basis of attention point prediction.

The early versions of attention prediction models focused on human visual attention and eye movement prediction. Itti et al. [5] used computers to simulate visual attention point sequences, and their studies provided the foundation of subsequent saliency detection models. Attention point prediction models produce Blod maps to represent a series of visual attention points.

2.2. Salient Region Detection. In comparison with attention prediction research, several more studies were conducted to locate salient regions. Most models were developed to identify salient regions that could be integrated onto entire segments of salient objects.

Contrast is an important factor in salient region identification, and, thus, it is widely used in saliency detection. The brain is sensitive to stimulations due to high-contrast images. In fact, previous models determined significance on

the basis of differences (i.e., most differences are based on low-level characteristics). Accordingly, models can be divided according to global contrast and local contrast.

All bottom-up models rely on prior information, and thus, the models can be classified according to a priori foreground and a priori background. Models that can directly calculate target significance are called a priori foreground models; otherwise, they are known as a priori background models.

2.2.1. Salient Region Detection Based on Local Contrast. Contrast can be directly expressed as center-surround difference in order for the high-contrast parts of adjacent image regions to become relatively noticeable. In local detection, subsets (pixels, blocks, superpixels, and regions) are compared with adjacent subsets. The distance between features is called saliency value. The saliency detection algorithm of local contrast detection produces a high saliency value for the edge of a salient object, and this further results in unambiguous highlighting of the entire salience target.

2.2.2. Salient Region Detection Based on Global Contrast. Unlike in local contrast models, global contrast models compare all pixels or regions within the entire image. Then, the compared results of the pixels or regions are accumulated. Salient region detection based on the global contrast offers the following advantages.

(1) The defects from local contrast are compensated in the contour position to produce high saliency.

(2) Similar saliency values are distributed in similar regions to highlight the entire salient region.

In global contrast, the background is highlighted rather than the prominent area when the image appears with a complex background or when the salient region of the image is particularly large.

2.2.3. A Priori Foreground and A Priori Background . Early bottom-up models focus on detecting computational featurebased contrast followed by various mathematical principle [53]. Meanwhile, some other mechanisms [8, 54] have been proposed to adopt some prior knowledge, such as a priori background, to detect salient objects in still images. Most existing models directly calculate salient objects (foreground), and thus, these models can also be called a priori foreground models. These models mainly focus on the uniqueness, scarcity, and singularity of visual information. On the basis of a priori foreground, Ma et al. [44] obtained significant information via image contrast techniques. Li proposed a bootstrap learning model by using a boosting method to select the optimal single-core classifier during iteration instead of conducting a linear superposition of the classifier. The model was bootstrapped with samples using the bottom-up approach to reduce the time needed for offline training. The above a priori foreground models concentrated at obtaining high edge values for an image and neglected the low values at inner of an object.

The a priori background model was first proposed by Wei et al. [55]. This model extracts background according to a priori information and calculates the difference of all pixels and backgrounds. From the results, the greater the difference, the higher the saliency score, and vice versa. The main concept of this models is the extraction of the background. Jiang et al. [56] proposed a novel model based on homology continuity to identify the so-called real background. The identified background was used to compute saliency, the value of which improved the accuracy and precision of the saliency map.

2.2.4. Learning Algorithm. All the above traditional models just integrate the low-level features, and it is difficult to extract the object from clutter background that lacks the high-level semantic knowledge. The low-level features generate the detailed and sharp boundaries, and the high-level semantic features recognized and classify the region of raw image. It is helpful to combine all of the high- and low-level features. In recent years, deep learning, especially Convolutional Neural Network (CNN), has attracted much attention due to ability to extract the high semantic information, and it has shown commendable result in many datasets. For instance, Wang et al. [57] first proposed two neural networks DNN-L and DNN-G, which, respectively, learn local patch features and global features, and the final saliency map is generated by the combination of saliency regions. Li and Yu [58] introduced a neural network for extracting features at three scales, aggregating the multiscale saliency map in different level of segmentation. Though these models have shown preferable results, they extract features on special levels, and all levels of information are significant. Lu et al. present a multilevel feature aggregation network (AmuletNet) to integrate all level features (low-level features and high-level semantics) into multiple resolutions. The model also adds low-level features to refine the boundaries. However, how to exploit effectively multilevels features is a pivotal problem.

2.3. Salient Object Detection. Although a number of models have been proposed, the rapid and accurate identification of salient regions remains a challenge in salient region detection, particularly in cases of complex textures. Some models have ill-defined object boundaries. This ascribes to severe downsizing of the input image, which reduces the range of spatial frequencies in the original image considered in the creation of the saliency maps. Some models have incomplete internal information, which results from the limited range of spatial frequencies retained from the original image in computing the final saliency map. Several existing models have solved this problem to some extent. Generally speaking, in most papers, the terms salient objects and salient regions (or salient object regions) are used interchangeably. However, compared with former region detection models, salient object models outputs full resolution saliency maps with well-defined boundaries of salient object, and results remain a wider spatial frequency domain from the original image. Therefore, unlike the traditional classification, we distinguish salient objects and salient regions detection. In summary, rough saliency maps can be generated by the models to preserve most of the characteristics of objects and produce clear boundaries, as shown in Figure 4. However, there are also deficiencies in the overall highlighting of the entire

object. Our motivation is to try to find a way to improve the performance of the algorithm based on some subsequent processing

3. Performance and Analysis

3.1. Datasets. Considering that many models were proposed in literature, several datasets were also introduced. Subsequently, these datasets with multiple objects and complex backgrounds were collated.

Various datasets may produce different results. To verify the rationality of the test, we adopt the (1) ECSSD, (2) SED2, (3) JuddDB, and (4) DUTOMRON to evaluate the models. The ECSSD comprises 1,000 semantically meaningful but structurally complex images. The SED2 contains 100 images. Each image appears with two significant objects in order for the study to determine whether or not the model can ensure good performance during nonsingle target detection. In SED2, the objects of the image are positioned far from the center; in addition, target object is small. Generally, JuddDB contains multiple objects, and the most of saliency objects are smaller compared to the comprehensive background. Finally, I also has test models over DUTOMRON, and the dataset which includes 5000 images was to evaluate models on a large scale.

3.2. Evaluation Measure. Two methods are used to evaluate the performance of the saliency detection model: qualitative analysis and quantitative evaluation. Qualitative analysis can be used to visually observe the saliency maps and compare the Ground-Truth (GT) in the datasets. The criteria are expressed as follows: (1) if the target object in the image can be highlighted, and if height separation in the background can be maintained; (2) if the entire object can be unified; (3) and if a high degree of similarity with the GT can be observed. This qualitative method requires an extensive process, but its results may have low accuracy. By contrast, quantitative evaluation is simple and accurate. Nonetheless, the two universally accepted classical measures were conducted to effectively evaluate the models. Measurements are applied to the overlapping area between the prediction map and the GT.

Precision–Recall. Using saliency map A, the measure converts A to a binary mask B; then precision and recall are computed by comparing B with the GT. The main step in this process is the conversion of A to B. Three methods (fixed threshold, adaptive threshold, and saliency cut algorithm) are used to perform binarization. Adaptive threshold is selected by

$$\begin{aligned} \text{Precision} &= \frac{|M \cap G|}{|M|}, \\ \text{Recall} &= \frac{|M \cap G|}{|M|}. \end{aligned} \tag{1}$$

Precision and recall cannot comprehensively evaluate the advantages and disadvantages of saliency maps. For a more comprehensive comparison we therefore also evaluate the mean absolute error (MAE) between saliency map and GT, both normalized in the range $[0, 1]$. The MAE is as follows:

$$\text{MAE} = \frac{1}{w * H} \sum_{x=1}^{H} \sum_{y=1}^{H} \left| \overline{S}(x, y) - \overline{G}(x, y) \right|. \tag{2}$$

S-Measure. The PRF is measured in a pixel-by-pixel manner; thus, the structural information obtained from the PRF is insufficient. Moreover, the structural information of a saliency map is necessary for many applications. Fan et al. [59] proposed a new evaluation index called the *S*-measure, which is expressed as

$$\text{ssim} = \frac{2\overline{xy}}{(\overline{x})^2 + (\overline{y})^2} \cdot \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \cdot \frac{\sigma_{xy}}{\sigma_x \sigma_y}, \tag{3}$$

where x and y are vectors representing the pixel values of SM and GT, respectively. Variables σ_x, σ_y, $\overline{x}$, and $\overline{y}$ denote mean and covariance. In addition,

$$S_\gamma = \sum_{\kappa-1}^{K} \omega_\kappa * \text{ssim}(\kappa), \tag{4}$$

where ω_κ is the weighting coefficient. Then,

$$O_{\text{FG}} = \frac{2\overline{x}_{\text{FG}}}{(\overline{x}_{\text{FG}})^2 + 1 + 2\lambda * \sigma_{x_{\text{FG}}}}, \tag{5}$$

where x_{FG} and y_{FG} represent the probability values of the SM and GT foreground regions, respectively; $\overline{x}_{\text{FG}}$ and $\overline{y}_{\text{FG}}$ are their mean values; and σ is the standard deviation. The following expressions are also defined:

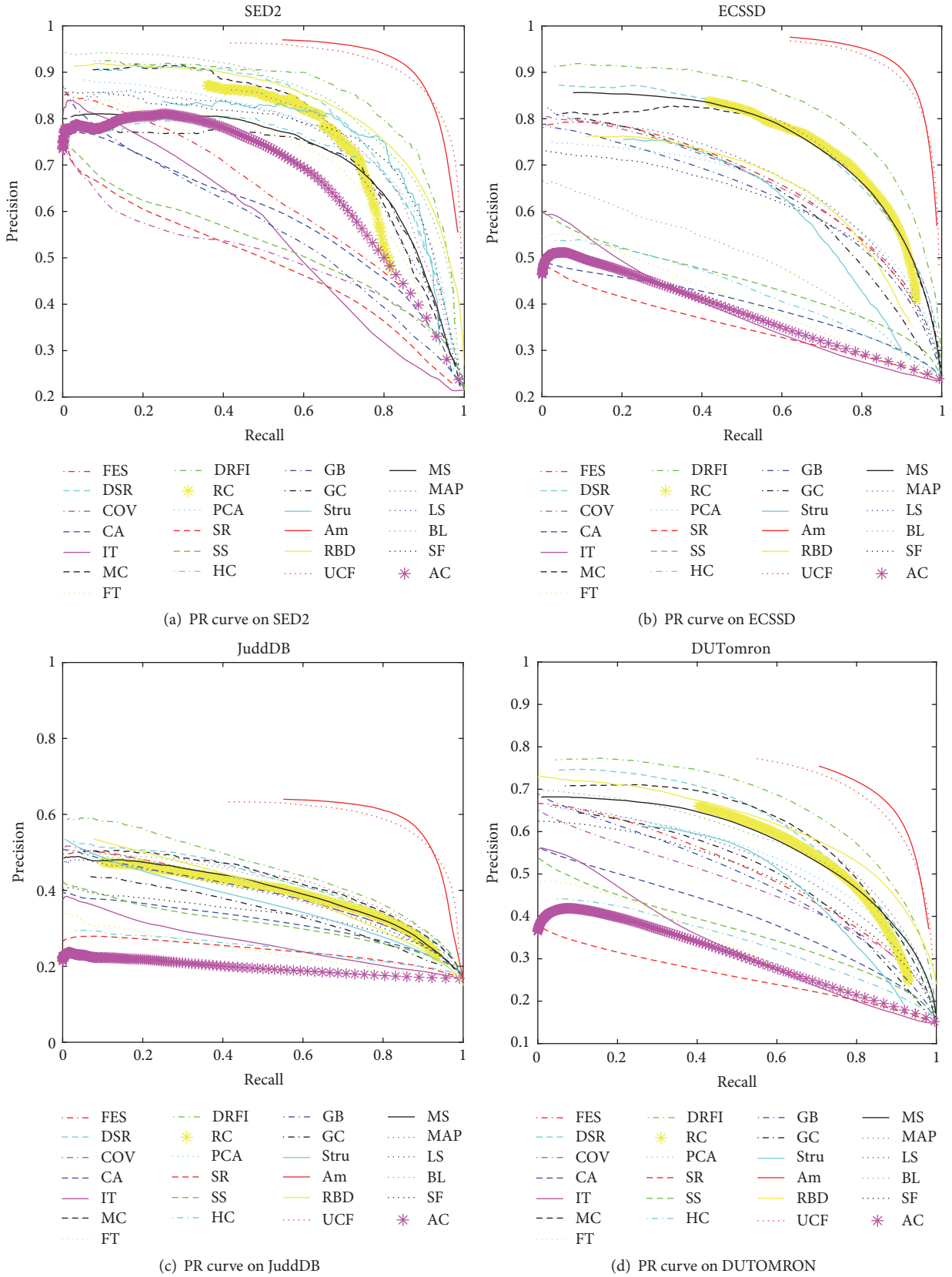
$$\begin{aligned} O_{\text{BG}} &= \frac{2\overline{x}_{\text{BG}}}{(\overline{x}_{\text{BG}})^2 + 1 + 2\lambda * \sigma_{x_{\text{BG}}}}, \\ S_o &= \mu * O_{\text{FG}} + (1 - \mu) * O_{\text{BG}}, \\ S &= \alpha * S_{\text{O}} + (1 - \alpha) * S_\gamma, \end{aligned} \tag{6}$$

where $\alpha \in [0, 1]$ with $\alpha = 0.5$, and μ is the ratio of foreground area in GT to image area (width $*$ height).

Note that the scores do not always work; for example, in SED2, there is server GT inverse, in other words, the saliency region is 0, and nonsaliency region is 1, and the result will be negative.

3.3. Results and Analysis. We evaluated the saliency maps by using four datasets and three different evaluation measures.

As shown by the large margins in Figure 2, with the exception of convolution network models, the traditional model DRFI achieved the best performance in each dataset. The DSR, MC, and RC models also attained good performances. The SR performed the weakest among all models in three datasets except JuddDB. Significantly, the convolution network modes are AM and UCF, the performance is perfect, and the experiments demonstrate that convolution network model performs favorably against traditional models. AM performs better on SED2, due to the edge-aware extraction,



(a) PR curve on SED2

(b) PR curve on ECSSD

(c) PR curve on JuddDB

(d) PR curve on DUTOMRON

Figure 2: PR curves of the saliency models on each datasets.

Table 2: MAE score (smaller is better; the top three models are highlighted in bold).

Model	SED2	ECSD	Judd	DUTOMRON
AC	0.206	0.263	0.239	0.190
BL	0.183	0.186	0.271	0.230
CA	0.229	0.308	0.282	0.254
COV	0.210	0.215	0.182	**0.156**
DRFI	**0.130**	**0.164**	0.213	0.155
DSR	0.140	0.171	0.196	0.139
FES	0.196	0.213	0.184	0.156
FT	0.206	0.289	0.267	0.250
GB	0.242	0.261	0.311	0.240
GC	0.185	0.212	0.258	0.197
HC	0.193	0.329	0.348	0.310
IT	0.245	0.271	0.200	0.198
LS	0.216	0.267	0.251	0.242
MAP	0.169	0.185	0.187	0.168
MC	0.182	0.202	0.231	0.186
MS	0.190	0.204	0.228	0.185
PCA	0.200	0.246	**0.181**	0.206
RC	0.147	0.185	0.270	0.189
SF	0.179	0.228	0.218	0.183
SR	0.220	0.264	0.200	0.181
SS	0.265	0.342	0.267	0.277
Stru	0.148	0.227	0.228	0.209
UCF	**0.067**	**0.058**	**0.115**	**0.113**
Am	**0.062**	**0.053**	**0.090**	**0.099**
RBD	**0.130**	0.173	0.212	0.144

while models extract the low-level features which are edge-aware features, and this helps to refine the boundaries and SED2 dataset has simple boundaries. How to integrate the high- and low-level information is key. The convolution network models exceed other models.

The results of MAE are shown in Table 2. There is no accident; AM and UCF models achieved relatively good results in each of the datasets. And then DRFI, DSR, RC, and MC also have a good performance. The DSR and DRFI model maintained its performance and ranked second among SED2, Judd, and DUT. However, except AM and UCF, COV and FES outperform other models in Judd. The results demonstrate that the convolution network model has good applicability whatever the raw images are. In addition, the region detection models performed relatively well on the average.

The latest evaluation indicator (S-measure) was used to highlight the limitations of traditional evaluation indicators, as shown in Table 3. Traditional approaches use pixels, which insufficiently supply overall structural information when saliency maps are evaluated. In the present study, AM and UCF still maintain the outstanding performance. However, RC models perform better in ECSSD. DSR model performed better than the MC model in terms of the PR. Nonetheless, the MC model produced a better saliency map compared with its counterpart models (Figure 3). The MC model also obtained a relatively complete shape of the target, which suggests good performance. Previous PR could not account for structural similarity; its results were ranked differently. However, if the evaluation measure could not capture the structural information of the object, then it could not also provide reliable information on model performance

Figure 3 shows the saliency maps of all 25 models for the two sample images. In the figure, the dark area indicates less significance while the white area corresponds to a high saliency value. The AM, UCF, DRFI, DSR, MC, and RC models performed relatively well. Their contours were nearly complete and their backgrounds were suppressed. The above models were not pixel-based.

Figure 4 shows the results of binarization (i.e., see section on salient object detection). The S-measure was used for the evaluation, and results proved good performance.

In terms of time spent for computation, we used a computer with Intel Core i3-2130 3.40 GHZ CPU with 4 GB RAM. The fastest model was LC (approximately 0.129 seconds per image), while the slowest was CA (approximately 66.426 seconds per image).

4. Prior Knowledge Problems

We found some issues after analyzing the characteristics of many models.

At present, many bottom-up saliency detection models integrate several a priori information, such as a priori boundary information and background information, among others,

Table 3: *S*-measure score (larger is better; the top three models are highlighted in bold).

Model	SED2	ECSD	Judd	DUTOMRON
AC	0.697	0.293	0.475	0.368
BL	0.934	0.565	0.522	0.583
CA	0.623	0.409	0.553	0.570
COV	0.598	0.410	0.753	0.495
DRFI	0.939	0.891	0.485	0.577
DSR	0.913	0.889	0.496	0.626
FES	0.611	0.319	0.481	0.487
FT	0.876	0.367	0.506	0.400
GB	0.662	0.509	0.509	0.606
GC	**0.951**	0.812	0455	0527
HC	0.932	0.287	0.476	0.486
IT	0.551	0.320	0.503	0.414
LS	0.815	0.590	0.607	0.631
MAP	0.917	**0.945**	0.702	0.617
MC	0.925	**0.954**	0.475	0.592
MS	0.936	0.551	0.468	0.520
PCA	0.848	0.733	0.542	**0.681**
RC	0.942	**0.908**	0.455	0.525
SF	0.919	0.770	**0.848**	0.502
SR	0.445	0.305	0.492	0.378
SS	0.577	0.423	0.437	0.676
Stru	0.606	0.577	0.543	0.563
UCF	**0.957**	0.768	**0.861**	**0.819**
Am	**0.961**	0.843	**0.883**	**0.891**
RBD	0.911	0.752	0.497	0.468

which can be used accurately for specific scenarios (e.g., when locating the salient region). However, this condition is not always the case for practical applications; in fact, the above approach may cause serious errors. For superior models that use center biases, the saliency targets are assumed to be not traceable in the boundary of an image. However, when part of the salient target is on the margin, the salient region is lost. The performance of the image with a salient object in the center is better than a salient object in the margins. Thus, four images were selected from the ECSSD (Figure 5). The area of the target decreased on the margins, and the *S*-measure gradually increased. Results indicate that appropriate prior knowledge can improve the model performance, but this result is not absolute.

In addition, practically all existing salient object detection models assume that one or several salient targets are found in a scene, but this assumption is suitable in experimental environments only. By contrast, problems occur in actual application. Thus, we selected four images with 300×400 resolution from the Internet. The test results for the MC model are shown in Figure 6. No obvious salient objects were shown by the four images. Theoretically, a good model can produce a black image if no salient objects are detected; that is, the grayscale of the pixels is zero. However, in the present study, the grayscales of the pixels greater than zero (i.e., 0.51%, 0.42%, 0.83%, and 1.35% of total pixels). The results differed significantly from theory, and this may be due to a number of reasons. First, the saliency detection model consistently assumed that the salient targets were found in the image and attempted to find them there, whereas a priori information also suggested that the salient region was at the center of the image. Second, contrast information was incorporated into the model. Although complex texture information was found in the image, several high-contrast areas were found within the surrounding area; instead, the model focused on the salient region. Finally, the model employed a superpixel unit, which reduced the complexity of the computation and inevitably resulted in less pronounced pixels that were considered highly salient because of clustering.

5. Conclusion

In this study, we thoroughly reviewed literature on saliency detection. 25 classical saliency detection models were selected and divided into three categories. Given that salient regional detection has recently attracted considerable attention, we also conducted a relatively more detailed classification for this topic. We discussed local comparison and global contrast classifications, and categories were divided into a priori foreground and a priori background. The performances of the 25 selected models over four datasets were tested with PR, MAE, and *S*-measure evaluation measures. On the basis of the evaluation results, apart from the convolution network, the performances of the MC, DSR, and DRFI models were

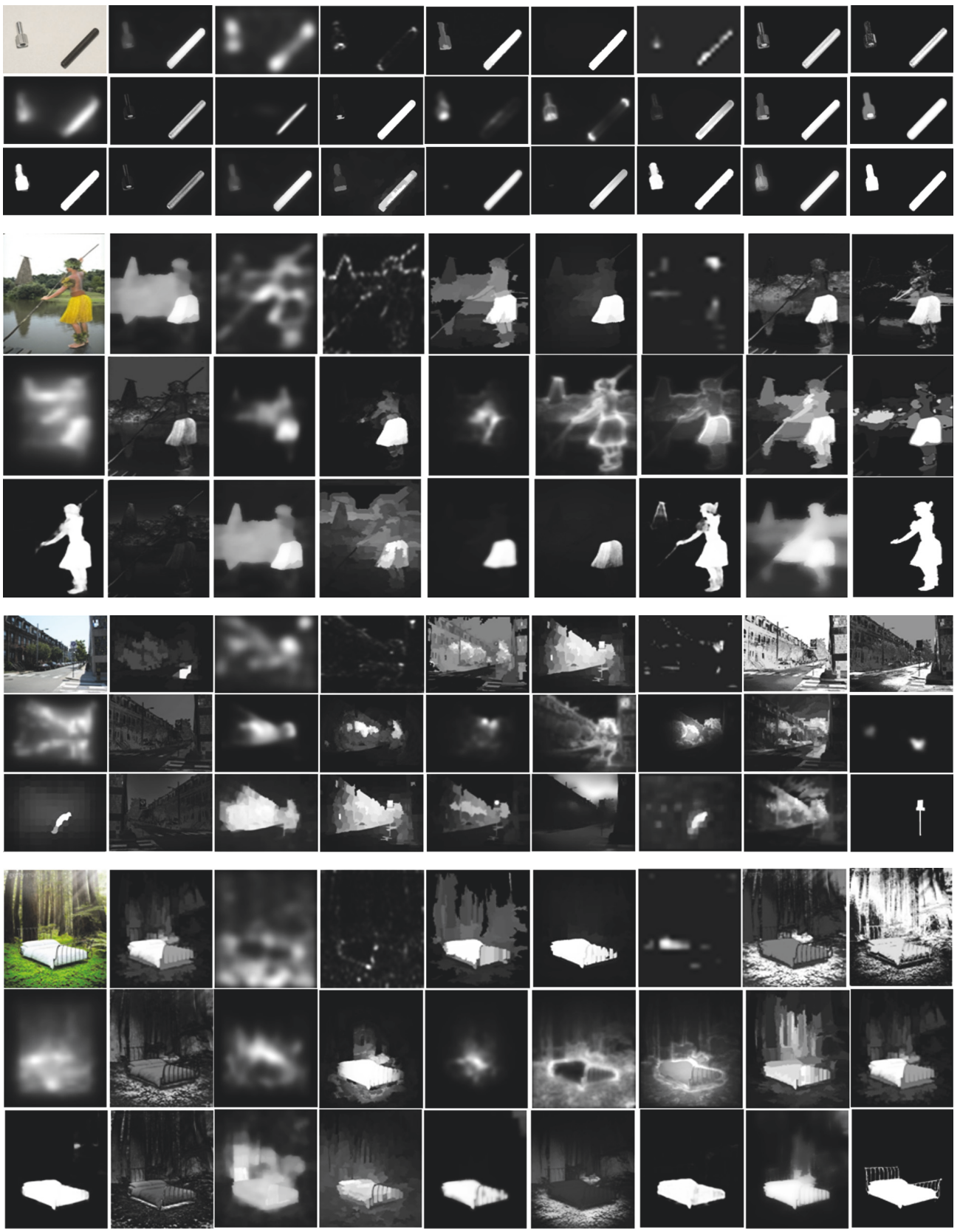


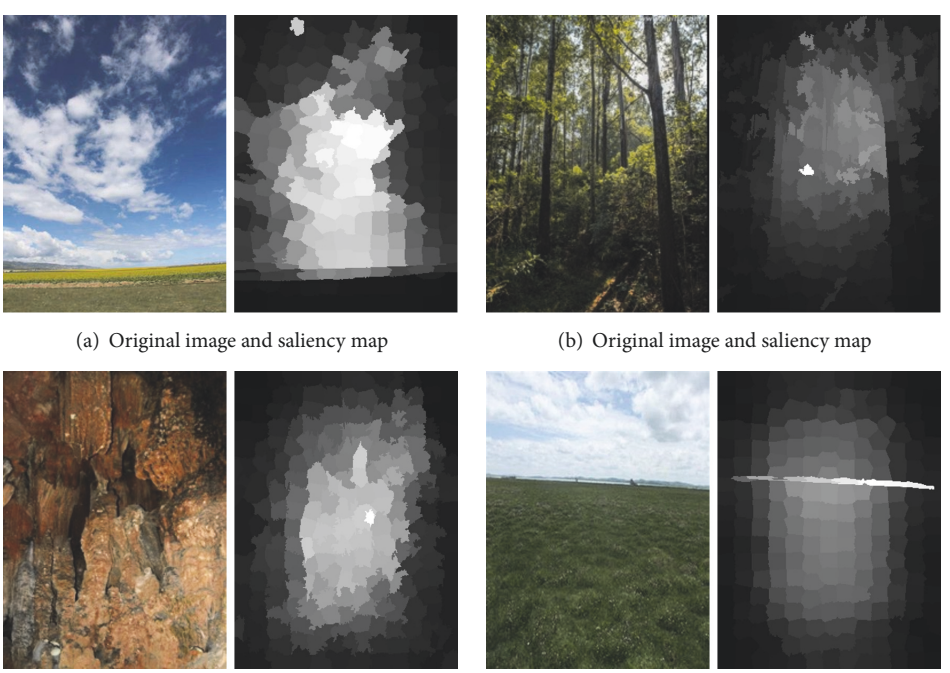
Figure 3: Saliency map (left to right: original image, RBD, SS, SR, RC, MC, IT, HC, GC, GB, FT, FES, DSR, COV, CA, PCA, DRFI, Stru, AM, AC, BL, LS, MAP, SF, UCF, MS, and GT. Top to down: SED2, ECSSD, JuddDB, and DUTOMRON).

(a) Rough map (b) Refined saliency map (c) GT

Figure 4: S-measure: (a) rough saliency map is 0.63; (b) refined saliency map is 0.73.

(a) Image (b) Image (c) Image (d) Image

Figure 5: S-measure (DSR and MC): (a) 0.464, 0.344; (b) 0.510, 0.481; (c) 0.789, 0.654; (d) 0.844, 0.730.



(a) Original image and saliency map (b) Original image and saliency map

(c) Original image and saliency map (d) Original image and saliency map

Figure 6: Original image and saliency map (MC).

relatively optimal. CNN model shows excellent performance than all of other models.

Three algorithms were considered for salient regional detection. On the basis of the test results, the PR and S-measures performed well. The common features of the models were as follows:

(1) Owing to the a priori background, the significant object was observed in the center of the image whereas the boundary area was found in the background.

(2) The models focused on superpixels or regions to emphasize the entire target, thereby significantly reducing calculation costs compared with those for pixel-based models.

(3) The models did not focus on a single feature only, given the practicality of feature combinations. Foregrounds differ from backgrounds when different cues are considered; that is, a single cue cannot be applied in all cases. For example, several images appear with obvious color differences and different textures.

Salient region detection has attracted considerable attention in recent years, but individual objects are mostly targeted in an image. The rapid development of CNN may provide a new direction for saliency detection. CNN can be used to distinguish salient and nonsalient regions of an image, thereby overcoming the limitations of traditional feature detection models, such as insufficient learning and poor robustness [60]. Compared to traditional models, the difference is the extracted features. The features extraction of CNN is classified into three kinds, from lower to high, from higher to low, and two-way. There is a supervised learning process. The existing problem is how to choose the network layers, although each network layer is significant, nevertheless, full convolution network layers increase workload. The other problem is what kind of low-level features to integrate, and the goal is to produce fine boundaries. Furthermore, multiple object and dynamic scene (e.g., videos) detection is still limited for use in actual applications, and thus, this gap may serve as a future research direction for salient region detection.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] A. Borji, "What is a salient object? A dataset and a baseline model for salient object detection," *IEEE Transactions on Image Processing*, vol. 24, no. 2, pp. 742–756, 2015.

[2] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 185–207, 2013.

[3] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, no. 1, pp. 97–136, 1980.

[4] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, no. 4, pp. 219–246, 1985.

[5] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[6] J. Zhang and S. Sclaroff, "Saliency detection: a boolean map approach," in *Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV '13)*, pp. 153–160, Sydney, Australia, December 2013.

[7] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV '13)*, pp. 2976–2983, Sydney, Australia, December 2013.

[8] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 2814–2821, June 2014.

[9] C. Scharfenberger, A. Wong, and D. A. Clausi, "Structure-guided statistical textural distinctiveness for salient region detection in natural images," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 457–470, 2015.

[10] J. Harel, C. Koch, and P. Perona, "Graph-Based Visual Saliency," in *Proceedings of the International Conference on Neural Information Processing Systems*, pp. 545–552, MIT Press, 2006.

[11] X. D. Hou and L. Q. Zhang, "Saliency detection: a spectral residual approach," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, June 2007.

[12] X. Hou, J. Harel, and C. Koch, "Image signature: highlighting sparse salient regions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 194–201, 2012.

[13] E. Erdem and A. Erdem, "Visual saliency estimation by nonlinearly integrating features using region covariances," *Journal of Vision*, vol. 13, no. 4, article 11, 2013.

[14] H. R. Tavakoli, E. Rahtu, and J. Heikkilä, "Fast and efficient saliency detection using sparse sampling and Kernel density estimation," in *Proceedings of the Scandinavian Conference on Image Analysis*, pp. 666–675, Springer, 2011.

[15] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1915–1926, 2012.

[16] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: a discriminative regional feature integration approach," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 2083–2090, Portland, Ore, USA, June 2013.

[17] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2814–2821, June 2014.

[18] J. Sun, H. Lu, and X. Liu, "Saliency region detection based on Markov absorption probabilities," *IEEE Transactions on Image Processing*, vol. 24, no. 5, pp. 1639–1649, 2015.

[19] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. Huang, and S. Hu, "Global contrast based salient region detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 409–416, Providence, RI, USA, June 2011.

[20] M. M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015.

[21] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang, "Saliency detection via absorbing Markov chain," in *Proceedings of the 14th*

IEEE International Conference on Computer Vision (ICCV '13), pp. 1665–1672, IEEE, Sydney, Australia, December 2013.

[22] R. Margolin, A. Tal, and L. Zelnik-Manor, "What makes a patch distinct?" in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 1139–1146, IEEE, Portland, Ore, USA, June 2013.

[23] P. Zhang, D. Wang, H. Lu, H. Wang, and X. Ruan, "Amulet: Aggregating Multi-level convolutional features for salient object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 202–211, Venice, Italy, October 2017.

[24] N. Tong, H. Lu, X. Ruan, and M.-H. Yang, "Salient object detection via bootstrap learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1884–1892, IEEE, June 2015.

[25] Y. Wu and X. Shen, "A unified approach to salient object detection via low rank matrix recovery," in *Proceedings of the IEEE Computer Vision and Pattern Recognition*, pp. 853–860, 2012.

[26] P. Zhang, D. Wang, H. Lu, H. Wang, and B. Yin, "Learning uncertain convolutional features for accurate saliency detection," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 212–221, IEEE Computer Society, Venice, Italy, October 2017.

[27] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 1597–1604, Miami, Fla, USA, June 2009.

[28] R. Achanta, F. Estrada, P. Wils, and S. Susstrunk, "Salient region detection and segmentation," in *Computer Vision Systems: 6th International Conference, ICVS 2008 Santorini, Greece, May 12–15, 2008 Proceedings*, vol. 5008 of *Lecture Notes in Computer Science*, pp. 66–75, Springer, Berlin, Germany, 2008.

[29] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and N. Crook, "Efficient salient region detection with soft image abstraction," in *Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV '13)*, pp. 1529–1536, Sydney, Australia, December 2013.

[30] N. Tong, H. Lu, L. Zhang, and X. Ruan, "Saliency detection with multi-scale superpixels," *IEEE Signal Processing Letters*, vol. 21, no. 9, pp. 1035–1039, 2014.

[31] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: contrast based filtering for salient region detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 733–740, IEEE Computer Society, June 2012.

[32] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition?" in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 2, pp. II-37–II-44, IEEE, July 2004.

[33] C. Kanan and G. Cottrell, "Robust classification of objects, faces, and flowers using natural image statistics," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 2472–2479, IEEE, San Francisco, Calif, USA, June 2010.

[34] F. Moosmann, D. Larlus, and F. Jurie, "Learning saliency maps for object categorization," in *Proceedings of the Workshop on the Representation & Use of Prior Knowledge in Vision (ECCV '06)*, 2006.

[35] A. Borji, M. N. Ahmadabadi, and B. N. Araabi, "Cost-sensitive learning of top-down modulation for attentional control," *Machine Vision and Applications*, vol. 22, no. 1, pp. 61–76, 2011.

[36] A. Borji and L. Itti, "Scene classification with a sparse set of salient regions," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '11)*, pp. 1902–1908, May 2011.

[37] H. Shen, S. Li, C. Zhu, H. Chang, and J. Zhang, "Moving object detection in aerial video based on spatiotemporal saliency," *Chinese Journal of Aeronautics*, vol. 26, no. 5, pp. 1211–1217, 2013.

[38] A. Borji, M. M. Cheng, H. Jiang et al., "Salient object detection: a survey," *Eprint ArXiv*, vol. 16, no. 7, p. 3118, 2014.

[39] Z. Ren, S. Gao, L. T. Chia et al., "Region-based saliency detection and its application in object recognition," *EEE Transactions on Circuits & Systems for Video Technology*, vol. 24, no. 5, pp. 769–779, 2014.

[40] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu, "Sketch2Photo: internet image montage," *ACM Transactions on Graphics*, vol. 28, no. 5, pp. 1–10, 2009.

[41] S. Frintrop, G. M. García, and A. B. Cremers, "A cognitive approach for object discovery," in *Proceedings of the IEEE 22nd International Conference on Pattern Recognition (ICPR '14)*, pp. 2329–2334, August 2014.

[42] B. C. Ko and J.-Y. Nam, "Object-of-interest image segmentation based on human attention and semantic region clustering," *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, vol. 23, no. 10, pp. 2462–2470, 2006.

[43] J.-Y. Zhu, J. Wu, Y. Wei, E. Chang, and Z. Tu, "Unsupervised object class discovery via saliency-guided multiple class learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 4, pp. 862–875, 2015.

[44] Y.-F. Ma, L. Lu, H.-J. Zhang, and M. Li, "A user attention model for video summarization," in *Proceedings of the 10th International Conference of Multimedia*, pp. 533–542, ACM, December 2002.

[45] Y. F. Ma and H. J. Zhang, "A model of motion attention for video skimming," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 1, pp. I-129–I-132, Rochester, NY, USA, 2002.

[46] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000 still image coding system: an overview," *IEEE Transactions on Consumer Electronics*, vol. 46, no. 4, pp. 1103–1127, 2002.

[47] S. Abidan, "Seam carving for content aware image resizing," *Siggraph*, vol. 26, no. 3, p. 10, 2007.

[48] S. Stalder, H. Grabner, and L. V. Gool, "Dynamic objectness for adaptive tracking," in *Proceedings of the Asian Conference on Computer Vision*, pp. 1–14, 2012.

[49] J. Li, M. D. Levine, X. An et al., "Visual saliency based on scale-space analysis in the frequency domain," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 35, no. 4, pp. 996–1010, 2013.

[50] G. M. García, D. A. Klein, J. Stückler, S. Frintrop, and A. B. Cremers, "Adaptive multi-cue 3D tracking of arbitrary objects," in *Pattern Recognition*, pp. 357–366, Springer, Berlin, Germany, 2012.

[51] A. Borji, S. Frintrop, D. N. Sihite, and L. Itti, "Adaptive object tracking by learning background context," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 23–30, June 2012.

[52] D. A. Klein, D. Schulz, S. Frintrop, and A. B. Cremers, "Adaptive real-time video-tracking for arbitrary objects," in *Proceedings of the 23rd IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '10)*, pp. 772–777, Taipei, Taiwan, October 2010.

[53] W. Wang, J. Shen, and L. Shao, "Video salient object detection via fully convolutional networks," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 38–49, 2018.

[54] W. Wang, J. Shen, L. Shao, and F. Porikli, "Correspondence driven saliency transfer," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5025–5034, 2016.

[55] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Proceedings of the 12th European conference on Computer Vision (ECCV '12)*, pp. 29–42, Florence, Italy, October 2012.

[56] Y. W. Jiang, L. Y. Tan, and S. J. Wang, "Saliency detected model based on selective edges prior," *Journal of Electronics & Information Technology*, vol. 37, no. 1, pp. 130–136, 2015.

[57] L. Wang, L. Wang, H. Lu, P. Zhang, and X. Ruan, "Saliency detection with recurrent fully convolutional networks," in *Proceedings of the European Conference on Computer Vision*, vol. 9908, pp. 825–841, Springer, Cham, Switzerland, 2016.

[58] G. Li and Y. Yu, "Deep contrast learning for salient object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 478–487, July 2016.

[59] D. Fan, M. Cheng, Y. Liu, T. Li, and A. Borji, "Structure-measure: a new way to evaluate foreground maps," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4558–4567, Venice, Italy, October 2017.

[60] Q. Hou, M. M. Cheng, X. Hu, A. Borji, Z. Tu, and P. H. Torr, "Deeply supervised salient object detection with short connections," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.