A Brief Survey of Visual Saliency Detection

Inam Ullah[1], Muwei Jian[2,3,4], Sumaira Hussain[1,4], Jie Guo[1], Hui Yu[5], Xing Wang[6], Yilong Yin[1,*]*

[1]School of Software Engineering, Shandong University, Jinan, China.

[2]School of Information Science and Engineering, Linyi University, Linyi, China.

[3]School of Computer Science and Technology,Shandong University of Finance and Economics, Jinan, China.

[4]Department of Computer Science, Sindh Madressatul Islam University, Karachi-74000, Pakistan.

[5]School of Creative Technologies, University of Portsmouth, Portsmouth, UK.

[6]School of Electronic and Information Engineering, Liaoning Technical University, China.

[*](Dr. Jian and Dr. Yin are co-corresponding authors) E-mail: jianmuweihk@163.com; ylyin@sdu.edu.cn

Abstract:

Salient object detection models mimic the behavior of human beings and capture the most salient region/object from the images or scenes. This field has many important applications in both computer vision and pattern recognition tasks. Despite hundreds of models proposed in this field, it still has a large room for research. This paper demonstrates a detailed overview of the recent progress of saliency detection models in terms of heuristic-based techniques and deep learning-based techniques. We have discussed and reviewed its co-related fields, such as Eye-fixation-prediction, RGBD salient-object-detection, co-saliency object detection, and video-saliency-detection models. We have reviewed the key issues of the current saliency models and discussed future trends and recommendations. The broadly utilized datasets and assessment strategies are additionally investigated in this paper.

1. Introduction:

The human vision system (HVS) has the incredible capability to recognize and focus the impressive objects or regions quickly, which are more visually distinct and prominent in the images/sceneries This process has been explored in computer vision [1-4] to detect those salient objects which have more importance and valuable information inside the images or videos, such as object recognition tasks, scene perception, and underwater vision, etc. This is an emerging topic and has recently engrossed the wide consideration of researchers from various disciplines. The mechanism of detecting a salient-object from an image is called saliency detection or salient-object-detection. The basic concept of salient-object-detection is shown in Figure 1. The first row represents the original images, and the corresponding ground-truth of each image is shown in the second row.

Saliency detection process first locates and identifies the correct location/region of the object, and then segments it from its background. For this purpose, a lot of models have been proposed, which have achieved a good performance in simple images/scenes having a single object. however, it is still difficult to find a salient-object in complex scenes, which have a more complex and cluttered background [5].

Figure 1. An example of salient-objects and their corresponding ground-truth.

Thereinto, bottom-up saliency detection is the mechanism that automatically captures the more focused and stimuli objects' regions of human visual attention without any prior knowledge [6]. Usually, saliency is termed as variance and contrast between a pixel and its surrounding locality [7]. Moreover, saliency-map is used to describe the degree of image saliency. In the saliency-map, each saliency value represents the pixel values of its corresponding regions in the image. It has a long history and it is still considered as an active research area in computer vision research.

In general, good saliency detection approaches must ensure precise object detection, high resolution and

computational efficiency[8]. Currently, different researchers have been classified as state-of-the-art methods based on different principles. In this work, we discuss comprehensively salient-object-detection models. We also discuss the common datasets and evaluation measures used for saliency- detection approaches. We summarized the related work and suggest some recommendations for future research work.

The remaining paper is organized as follows: In section 2, we briefly review various salient-object-detection models such as RGBD salient-object-detection models, Co-saliency-detection models, and video-saliency-detection models. In section 3, we discuss briefly the co-related databases for saliency detection. In section 4, we enlist the databases and applications of salient-object detection and finally, we provide the conclusion and future recommendations.

2. Review of visual saliency detection models:

Visual attention has been explored in multiple disciplines of computer vision [9-12]. Based on the early cognitive theories, in 1980, Treisman and Gelade [13] presented a theory of feature integration and proposed feature integration model and feature registration model for visual attention. Wolfe et al. [14] proposed a biological structure (Guided-Search-Model) and Koch and Ullman [15] proposed a Computational Attention framework. These theories are founded on bottom-up center-surround mechanisms. In 1998, Itti et al. presented a visual attention model [12,1] to describe human visual attention, which generates a map for saliency detection by combining three different feature maps (i.e., color, orientation, and intensity) at various scales based on center-surround mechanisms. recently hundreds of visual attention models have been proposed, including fixation point prediction models.

afterward, Liu et al. [16] defined saliency detection as a binary segmentation work. Zhang and Sclaroff [17] analyzed the saliency-map by using Boolean map topology. To get a saliency-map, Li et al. [9] incorporated a reconstruction error scheme via dense and sparse representation. Zhu et al. [18] added a simple boundary to compute the background measure and find the spatial format in the image regions along with their corresponding boundaries. Consequently, an optimization method was adopted to incorporate different low-level cues such as background measures and obtained uniform saliency-maps. Scharfenberger et al. [19] presented a statistical pattern scheme, which robustly uses the essential heterogeneous textural features of the image and computes the relevant saliency of every region in the image effectively. In addition, there are several other techniques rely on mathematical calculation. Hou and Zhang [2] proposed the residual spectrum framework by using the Fourier transform phase spectra to generate a saliency map. Achanta et al. [20] obtained a saliency map based on local contrast by integrating low-level features. These classic models have yet achieved an admirable performance, but, due to the absence of high-level semantic information, these low-level models are still getting tough to achieve the desired results.

Nowadays, the resurgence of the deep-learning-based Convolutional Neural Network [21] and especially fully Convolutional Neural Network [22] provides a feasible technology for saliency detection. Different than traditional methods, which use the low-level visual information mostly based on contrast-priors [23], CNN based methods use high-level semantic information and abolish the need for handcrafted features. A CNN normally has hundreds or even ten thousands of parameters and neurons with various receptive field sizes. Neuron with the large receptive-field size is used to identify global information for the most salient-regions of the image, and the small receptive fields are used to identify the local information between the small regions of the image. The interest of researchers is rising in the CNNs model due to its tremendous performance and more desirable properties compared to classical hand-crafted feature-based models.

From the viewpoint of information processing mechanisms, saliency detection approaches can be generally categorized as bottom-up approaches and top-down models. The bottom-up methods are based on low-level visual features without high-level semantic information. On the other hand, the top-down approaches assume that the extrinsic cues for saliency detection with more semantic information. The top-down methods [2,24] are generally task-driven and require abundant training data with human-labeled ground truths. Thus these models can extract high-level semantic features from images to describe the specific objects (e.g. car, pedestrian). However, due to the complication and variation of daily tasks and behaviors, the high-level methods are not much explored.

In the last two decades, research work in this zone has developed in two directions: visual-attention-prediction (i.e., eye fixation-prediction) and saliency detection in computer vision. The earlier class emphasizes on locating the fixation-points of a human observer at the first glimpse [25,26], whereas the latter class tries to identify or/and segment the most prominent and salient objects from the original image [27]. In the following sub-section, we briefly review the fixation prediction models while providing comprehensive detail on salient-object-detection models.

2.1 Fixation Prediction Models:

To simulate visual attention, eye-fixation-prediction models have generally been corroborated against eye actions of human attention. Eyeball movements express important information concerning cognitive procedures such as analysis, scene perception, and visual search. Thus, they are frequently preserved as a proxy for changes of attention [7]. Primates have a strong talent to analyze complicated scenes in real-time. Visual systems will first make selections in the collected information before the extra processing of visual information. it can lessen dramatically the complication of obtained information. This selection method is accomplished in a limited field of view, named visual-attention-prediction. HVS imposes a solid dynamic selectivity process when sensing the exterior surroundings; in that scenario, dynamic selection functions as the procedure of the visual-attention-point transfer. Moreover, HVS can quickly grasp huge volumes of image information. The overhead sentiments elucidate the biological foundation of attention-point-prediction. Figure 2 shows some samples of human eye fixation prediction, where the red light blobs show the more salient-regions.

The initial classes of attention-prediction models are engrossed in human-visual-attention and eye gaze prediction. Itti et al.'s basic model used three simple feature channels (i.e., color, orientation, and intensity). This model becomes the basis of future models in this field and the standard benchmark for assessment. It has been presented to associate with human eye flux in free-viewing tasks [28,29]. Le Meur et al. [30] presented a method for bottom-up saliency detection constructed on contrast-sensitivity functions, perceptual-decomposition, center-surround interactions, and visual-masking. Later, Le Meur et al. [31] prolonged this model to the spatiotemporal field by combining chromatic, achromatic and spatial-temporal based information. In this modified model, they extracted the early visual-features from the visual input into some single parallel channels. A feature map is achieved for each channel, and then a distinctive saliency-map is constructed from the union of those channels. Kootstra et al. [32] proposed three symmetry-saliency basic operators and made their comparison with human eye-tracking data. This technique is constructed on the radial symmetry operators and isotropic-symmetry of Reisfeld et al. [33] and the color-symmetry of Heidemann [34].

Figure 2. Examples of Human eye-fixation-prediction.

2.2 Saliency detection models

In this paper, the literature of saliency detection has been classified into heuristic-based and learning-based approaches. In saliency detection, contrast is a very important factor for salient region identification [35] [36]. The brain is very sensitive to high-contrast objects/ regions in an image. Traditional heuristic approaches of the saliency detection are mostly based on low-level visual features and most of the computational frameworks are unsupervised [37]. These conventional bottom-up methods follow the heuristic features approach (i.e., such as contrast, location, and texture) during saliency detection. Heuristic features are usually called visual priors or cues for saliency detection [38] [39]. The contrast-prior is a very crucial feature and one of the most used priors. Concretely, the contrast priors comprise of local-contrast prior and global-contrast prior, and the contrast-prior assumes that the salient-regions are always dissimilar from their neighborhoods or scenes [40]. Beside contrast-priors, location priors consist of center-priors and background-priors. Center-priors describe the salient-object appears in the middle of the image, while the background-priors state that a border of an image has more chances to be part of the background. In this sub-section, we will discuss some important cues or priors in heuristic-based saliency detection models.

2.2.1 Heuristic-based saliency detection models

A. Saliency detection based on local contrast

Contrast represents the obvious difference between two or more pixels/regions in an image. The distance

between the two features is called a contrast-based saliency value. The edges of a salient-object produce a high saliency score in local-contrast saliency methods [39], thus highlighting the entire salient target. Local-contrast based saliency detection [41-44,9,45,46,24,47] has been proposed, which calculates the saliency value map by considering local features (i.e., color, illumination, orientation, and other motion information) between different regions.

Itti et al. [41] presented a center-surround method and by using a linear and non-linear combination of multi-scale saliency-map to extract low-level elements (i.e., color, intensity, texture, and orientation). Ma and Zhang [42] used color contrast as a saliency measure in a local neighborhood. In [43], Jiang et al. introduced a regional level saliency descriptor primarily based on local-contrast, backgroundness, and other well-known features. Jiang et al. [44] proposed a strategy based on multi-scale local contrast regions, which computes saliency values throughout different regional segmentation to create robustness and combines each value of these regions to obtain a pixel-wise saliency map. In [9], the authors adopted a similar framework by estimating regional saliency using multiple hierarchical segmentation. Li et al. [45] lengthened the pairwise local-contrast with the aid of creating a hypergraph, which is made by a non-parametric multi-scale non-parametric gathering of superpixels, in order to obtain both interior consistency and exterior separation of regions. Salient object detection is then achieved via looking for salient vertices and hyperedges in the hypergraph. Liu et al. [24] proposed a multi-scale contrast based saliency-detection algorithm by linearly merging local features in a Gaussian image pyramid. Goferman et al. [47] consecutively devised a model based on local low-level contrast, global-contrast, visual organization policies and other high-level elements to capture conspicuous salient items along with their contexts. Jian et al. [48] designed a saliency-detection model based on principal local color contrast.

B. Saliency detection based on global contrast

Unlike local-contrast based methods, a global-contrast based method [23,49-55] usually separates an object from its surroundings. Global-contrast based methods have advantages over local-contrast based techniques as they generate excessive saliency values at their object boundaries. In global feature consideration, similar saliency values are disseminated in similar regions leading to generate high saliency cost.

Cheng et al. proposed a color histogram as the global-contrast and calculated the weighted sum of color difference for every region with all other regions of the same image [23]. Harel et al. [49] presented a global-based saliency-detection method based on graph theory. Zhai and Shah [50] computed the saliency score by calculating the sum of the color difference of each pixel with all other pixels. Achanta et al. [51] presented a frequency-tuned model that estimates pixels level saliency score by directly computing the color difference from its average image color. Perazzi et al. [48] measured the global-contrast by applying the uniqueness of the element and the spatial distribution of the image. Goferman et al. [52] proposed a patch uniqueness method for saliency estimation by considering global contrast with respect to other patches. Yan et al. [53] introduced a hierarchical saliency-detection approach to address the small-scale changes in the high contrast structure. Shen et al. [54] introduced a low-rank recovery technique to add low-level visual structures with high-level priors for saliency detection. Imamoglu et al. [55] used the wavelet transform to produce multiscale structures that curb local contrast with global saliency. Perazzi et al. [48] applied Gaussian filters to compute the global uniqueness and spatial distribution for salient object detection. Though adequate research has been carried on global priors, however, it still has weaknesses in capturing the semantic information.

C. Saliency detection based on center-prior

The primitive center-prior is actually based on the idea that a salient object frequently lies close to the middle of the image [53,56,57,44,43]. The center-prior tries to highlight the center region or combines with other cues to highlight the salient region/object as a spatial feature during saliency detection. However, we know that the salient object does not appear every time in the image center. To conquer this drawback, Xie et al. [57] utilized a convex hull of interest points to predict the coarse center of the salient object. Jian et al. [35] used perceptual directional patches based on a discrete wavelet frame transformed to a fixed position of the salient object.

D. Saliency detection based on backgroundness-prior

Backgroundness prior [58,9,59-61] deems the narrow border as a background region of the image. The saliency score can be calculated as the contrast against the background by considering the background seeds as a reference. Jiang et al. [58] offered a saliency-detection method by using absorbing Markov Chain, in which superpixels are the transit and absorbing nodes around the center and border of the image. Li et al. [9] proposed a saliency-detection scheme based on dense and sparse reconstruction errors by using image boundaries as background templates. Wei et al. [60] constructed an undirected weighted graph and estimated the saliency value as the shortest distance to

the background. Yang et al. [61] proposed a two-scheme saliency computation model by performing a manifold ranking approach on the basis of an undirected weighted graph by considering the relevance score of each side in the background queries. Saliency detection may fail based on pseudo-background, specifically when the item attaches the boundary. Boundary connectivity prior [23,18] is used to resolve this problem. Naturally, the background is more connected to the border than any salient object. Zhu et al. [18] used this idea to find the boundary connectivity score by estimating the length of the image border with respect to the spanning area of the salient region. Recently, a saliency-detection model based on background seeds by object proposals and extended random walk is proposed [38].

E. Saliency detection based on objectness prior

Beyond these techniques, objectness prior can also be used to assist salient object detection by using object proposals, which was introduced by Alexe et al. [62] to measure the probability value that there exists a whole object by assessing score of an objectness for every random window of the image. Chang et al. [63] presented a computational scheme by combining the regional saliency and objectness into a graphical saliency. Jiang et al. [64] computed regional objectness based on average objectness value of its all regional pixels. According to the objectness prior, Jia and Han [65] calculated the saliency score for each region and then compared these to the soft foreground and background. To connect objectness with the saliency score, local saliency is calculated by randomly taking a great number of sampling windows [66]. For images of complex scenes, Li et al. [67] proposed a three-center-biased objectness measure. They proposed a co-transduction approach to fuse boundary superpixels and objectness labels with each other. Moreover, Jiang et al. [64] computed the saliency score by non-linearly fusing the scores of uniqueness, objectness, and focusness.

F. Saliency detection based on Bayesian framework

Regarding saliency computation, the Bayesian model [57] is presented for finding salient objects by approximating the pixel x posterior probability as the foreground in the image. For saliency prior calculation, the interest pixels are estimated via a convex-hull function, which splits the image into inner and outsides regions and then obtains a rough estimation score for foreground and background. Liu et al. [68] used an optimization model based on a Bayesian framework for saliency detection by roughly estimating a convex-hull to classify the input image into potential foreground and pure background regions. To generate a saliency map, a common Linear Elliptic mechanism with Dirichlet boundary is presented using these cues to model the diffusion of the seeds to other regions.

Table 1 shows some representative methods of visually heuristic-based models using different cues/priors.

Referred	Pub	Year	Key priors	Code
FCS[41]	HVEI	2001	LC+CLP	C++
FG[42]	MM	2003	LC, GC	NA
FT[51]	CVPR	2009	GC+CS	C++
CA[47]	CVPR	2010	LC+CS+GC	NA
CB[44]	BMVC	2011	LC+CP	C+M
ULR[54]	CVPR	2011	GC+CP+CLP	C+M
SVO[63]	ICCV	2011	CS+O	C+M
RC[23]	CVPR	2011	BC+GC	C++
FES [69]	SCIA	2011	CS+LC	M
SWD[70]	CVPR	2011	LC+CS	M
SF[48]	CVPR	2012	GC+SD	C
GS[60]	ECCV	2012	BC	NA
DSR[9]	ICCV	2013	LC+BA	C+M
CHM[45]	ICCV	2013	LC+CS	C+M
HSD[53]	CVPR	2013	GC	EXE
WT[55]	TM	2013	LC+GC+B	M
LMLC[57]	TIP	2013	CS+BA	C+M
MC[58]	ICCV	2013	BC	C+M
GMR[61]	CVPR	2013	B	M
UFO[64]	ICCV	2013	GC+F+O	C+M
CIO[65]	ICCV	2013	GC+O	NA
PISA[71]	CVPR	2013	SD+CP	NA
GR[72]	SPL	2013	GC+CS	M
PCA[73]	CVPR	2013	GC	C++
COV[74]	JOV	2013	LC+CS	M
RBD[18]	CVPR	2014	BC	M
SLF[75]	CVPR	2014	F+ B	M
PDE[68]	CVPR	2014	CP+B+CLP	NA
ILP[67]	ITOIP	2015	GC+SD	NA

Table 1. A list of Traditional-based models using different cues/priors, where LC=local-contrast, GC= global-contrast, CP=center-prior, BA=Bayesian, CS=center-surround, CLP=color prior, B=background-prior, BC=background connectivity, O=objectness prior, F=focusness-prior, IN=informative feature, OR= orientation cue SD= spatial distribution, NA=not available and M= Matlab code.

G. Discussion

The above-discussed priors are the most common priors used in the heuristic-based saliency detection models. There are some other traditional techniques also introduced for saliency detection such as frequency domain analysis [51], cellular automata [76], sparse representation [9], random walks[59], low-rank recovery[77], compactness prior [78] and orientation prior[12].

These traditional-based approaches for salient-object-detection consist of intrinsic cues, which aim to withdraw different cues from the given input image by

itself to highlights the target regions and to suppress backgrounds. Moreover, much more complementary saliency priors can be utilized for saliency detection in order to enhance the performance and robustness, such as backgroundness, background connectivity, foregroundness, focusness, objectness, orientation, contrast, etc.

In this overview, based on common priors/cues, our classification only specifies the supremacy of the priors, because a model can consist of the single or the combination of different priors. The local and global are the most frequently used uniqueness saliency priors for saliency detection [235].

The traditionally heuristic-based approaches for saliency detection have got a great achievement in the field of computer vision, but still, it fails in some spatial cases, especially when the image contains a very complex scene, low contrast (e.g. underwater images) and interlaced objects. To overcome these problems, the learning-based approaches (supervised learning, semi-supervised learning or unsupervised learning-based approaches) are applied which we will introduce in the next section.

2.2.2 Learning-based saliency detection:

All of the above-mentioned methods which we studied among traditional-based approaches are using intrinsic low-level cues and based on unsupervised techniques, and these techniques are sometimes insufficient to detect accurately salient targets especially when the image is complex and shares common visual features. To tackle these issues, learning-based methods with training data are utilized to find a salient object in the complex background image.

A. **Classic Learning-based saliency detection methods**

These are supervised or semi-supervised learning based saliency detection methods, also called data-driven approaches, in which high-level features and supervised information are integrated to enhance the degree of accuracy for saliency maps. Judd et al. [79] proposed a model for Eye-fixation-prediction via a Support Vector Machine (SVM) classifier based on a training dataset including fixation locations of fifteen viewers. In [24], Liu et al. presented a binary saliency estimation scheme based on a conditional random field (CRF). Yang et al. [80] proposed a method that trains a Conditional random field (CRF) and a discriminative dictionary for saliency detection. The designed method includes a layered structure starting from the top-down manner, which is trained under structured supervision and then followed a max-margin mechanism for efficient learning. In [81], Borji et al. integrated low-level features (e.g., orientation, color, and intensity) with high-level visual-features (e.g., humans, faces and cars, etc.) to train a direct mapping approach by means of AdaBoost classifier for eye fixations. Wang et al. [82] proposed a method from multiple instances learning, where low-level, mid-level and high-level features are integrated for salient object detection. Jiang et al. [43] proposed a model of saliency detection as a regression structure and then trained a regression forest classifier to generate saliency values. In [91], Lu et al. presented a model and trained it to learn optimal seeds, and then these seeds are propagated through a diffusion process. Tong et al. [91] [35] put forward a salient-object-detection model via bootstrap learning technique, instead of training only a classifier in a large dataset. They also train a group of weak SVMs in order to obtain a strong classifier by incorporating the weak classifier through the multi-kernel boosting method.

As the classic learning-based methods utilize prior knowledge and occasionally outperform the traditional hand-crafted feature-based saliency-detection techniques, these methods boost the performance of saliency detection. Owing to the classic learning-based approaches are still hand-crafted features, which may degrade the performance of the models if they are not carefully collected. But, recently the development of CNNs-based approaches turned the trend of researchers to deep-learning approaches instead of classical machine learning algorithms, due to their tremendous performance.

B. **Deep-Learning based saliency detection models**

In this section, we introduce Deep-Learning based saliency detection models, especially CNNs and FCN-based based approaches.

Convolutional-Neural-Networks (CNNs) [21] has attracted great attention from researchers for its functionality in representing high-level semantics and has been successfully applied in many computer vision problems [22,83]. Recently, CNNs [84,85] has also shown its effectiveness in the field of saliency detection and has the capability to capture the most salient regions without prior knowledge. Generally, saliency-detection approaches established on CNNs can be classified into two basic classes: (1) region-based models, and (2) FCN-based (i.e., pixels-based) models, according to their processing with input images. The region-based approaches divide the input images into multi-scale or smaller regions. Then, CNN is utilized to extract the high-level features of these small regions and then input to multi-layer perceptrons (MLPs) to get the saliency value of each small region. The region-based models achieved a good performance against traditional state-of-the-art models, however, these models can't persevere the spatial information due to the segmentation of small regions. To overcome

this demerit, a Fully Convolutional-Neural-Network (i.e. FCN-based approach) is designed, also called end-to-end models by predicting saliency map directly with the end-to-end network.

Wang et al. [86] developed deep networks for saliency computation by combining shape, texture and contrast information from the local regions of the input image. In the global search stage, a list of candidate object regions is created via an object proposal method [87]. In [88], Lee et al. proposed a unified deep learning framework for saliency detection by utilizing high-level and low-level features of the image. The VGGNet [89] is trained to extract the high-level features and then the low-level features are integrated to identify the salient regions. He et al. [84] proposed a region-based model to learn feature representations from superpixels. It can reduce the computational cost as compared to pixel-wise CNN. Zou et al. [90] proposed a hierarchical-related feature (HARF) framework for saliency detection, which integrates the basic features from regions using a multi-level deep learning network. Kim et al. [91] proposed a two-bran CNN based saliency-detection model by considering the coarse representation and fine representation. A number of region candidates are generated through selective search [92] method and then taken as inputs to the CNN. Wang et al. [93] proposed a fast R-CNN based multi-scale mask framework for saliency detection, which segments the input image into multi-scale regions and an edge-based propagation approach is used to refine the saliency map. In [94], Kim et al. proposed a CNN model to estimate the saliency values of each image patches/region. Li et al. [95] utilized both low-level features captured through hand-crafted methods and high-level features by using CNNs methods to enhance the saliency accuracy. In this model, candidate bounding boxes with interior region masks are produced by using a selective search method [92]. Li et al. [85] captured deep features from multi-scale regions for saliency detection. And a superpixel refinement scheme is utilized to obtain an enhanced spatial coherence result. Zhao et al. [96] introduced a multi-context deep learning model, which captures the local and global scale features from the given superpixels to predict the corresponding saliency value of each region. In [97], Hariharan et al. presented a hypercolumn approach for salient object segmentation, and the features of different type layers are fused for further classification purposes. Liu et al. [97] proposed a hierarchically refine scheme which gradually produces a saliency map by exploiting the VGG net to produce a global coarse prediction. In [98], a refinement subnetwork recurrent convolutional-layers (RCL) are designed to fine-tune the coarse-level prediction map into fine-level saliency map.

The recent advanced CNNs saliency-detection frameworks have gotten considerably better results than earlier hand-crafted features methods. Furthermore, the CNNs extracted features comprises more high-level features because these CNNs are typically pre-trained for visual recognition activities on very large datasets. However, the Region-based CNNs are functioned at the segment-based or patch level rather than utilizing pixel-level, where each pixel is basically allocated the saliency score of its enclosing segment. As a result, it gives a blurred saliency map that lacks the fine details of the salient objects and their boundaries. Moreover, all the segmented patches or regions of the images are processed as an independent sample for classification purposes; even they may overlap each other. This redundancy causes a significant increase in computation as well as requires more space during training and testing. Furthermore, the region-based CNNs models cannot preserve the contextual information well. Thus, to overcome the shortcomings of region-based CNNs, the well-known end-to-end based Fully Convolution Network is adopted, which predicts pixel-wise saliency maps.

As we know that the region-based CNNs techniques can't well preserve the contextual information of the salient object because CNN is operated independently for each image patches or regions. To dispose of the above issue, Fully-Convolutional-Networks (FCNs) [22] operates on pixel-levels instead of regions or patches level. FCNs based saliency-detection techniques can eliminate problems such as vague predictions over the blurriness boundaries of the salient objects. FCNs-based models for salient object detection also have drawn the attention of the researchers due to its tremendous performance. Long et al. [22] introduced an FCNs based saliency-detection model, which is trained pixels-to-pixels by presenting the meaningful information obtained by deep and coarse layers. Li et al. [99] presented a model with a spatial pooling stream (SPS) and a pixel-wise fully convolutional stream (FCS) to generate a saliency map. Tang et al. [100] used the deeply supervised net [101] and designed a holistically-nested edge detector (HED) [83] for saliency detection.

In [102], Tang et al. proposed a saliency-detection scheme via fusing both pixel-level CNN and region-level CNN saliency prediction. Kruthiventi et al. [103] proposed an incorporated deep architecture for fixation prediction and salient object detection by fully connected CRF [104]. In [105], the authors designed a recurrent attentional convolutional-deconvolution (RACDNN) approach for saliency detection. In RACDNN, a segment of the input image is chosen in each time-step by a spatial transformer [106]. Zhang et al. [107] proposed a saliency-detection method based on CNNs and a multi-level amalgam framework. The

Deeplab[108] scheme is employed to get the high-level features, and a multi-scale binary-pixel-labeling framework is also employed to recover spatial coherency. Li et al. [109] presented a multi-task CNN

Model	Pub	Year	#Training Images	Training Set	Pre-trained Model
LCIR[89]	IVPR	2014	4600	VOC-2012	-
LEGS [86]	CVPR	2015	3,340	MSRA-B, PASCALS	-
SuperCNN [84]	IJCV	2015	800	ECSSD	-
HARF [90]	ICCV	2015	2500	MSRA-B	-
MDF[85]	CVPR	2015	4447	HKU-IS	-
MC [96]	CVPR	2015	8000	MSRA-10K	GoogLeNet
ELD [88]	CVPR	2016	approximately 9000	MSRA10K	VGGNet
SSD-DL [91]	ECCV	2016	2500	MSRA-B	AlexNet
SFRLC [93]	ICIP	2016	4000	DUT-OMRON	VGGNet
SPSD [94]	ICPR	2016	2500	MSRA-B	AlexNet
LCNN [95]	Neuro	2017	2900	MSRA-B + PASCALS	AlexNet
FCNSS[22]	CVPR	2015	4600	VOC2012	VGGNet
DCL [99]	CVPR	2016	2,500	MSRA-B	VGGNet
DHSNET [97]	CVPR	2016	6,000	MSRA10K	VGGNet
DSRCNN [100]	MM	2016	10,000	MSRA10K	VGGNet
CRPSD [102]	ECCV	2016	10,000	MSRA10K	VGGNet
SU [103]	CVPR	2016	10,000	MSRA10K	VGGNet
RACDNN [105]	CVPR	2016	10,565	DUTS+NJU2000+RGBD	VGG
DS [109]	TIP	2016	nearly 10,000	MSRA10K	VGGNet
DISC [110]	TNNLS	2016	5233	MSRA10K	VGG-16
IMC [107]	WACV	2017	nearly 6,000	MSRA10K	ResNet
MSRNet [111]	CVPR	2017	2,500	MSRA-B + HKU-IS	VGGNet
DSS [112]	CVPR	2017	2,500	MSRA-B	VGGNet
SRM[113]	ICCV	2017	10,553	DUT-OMRON	ResNet
NLDF[114]	CVPR	2017	2500	MSRA-B	VGG-16
DSLM[115]	ITOC	2018	15,000	DUT-OMRON+ MSRA10K	VGG-16
PAGR[116]	CVPR	2018	10,553	DUTS-TR	VGG-19
CKT [117]	ECCV	2018	10K	MSRA10K	VGG-16
RAS[118]	ECCV	2018	10000 plus	MSRA-B, DUT-OMRON	VGG-16
THR[119]	ICCV	2019	5000 plus	DUTS,HRSOD	VGG-16
AFN[120]	CVPR	2019	10,553	DUTS-TR	VGG-16
BASNet [121]	CVPR	2019	10,553	DUTS-TR	ResNet-34
Refinet [122]	ITOM	2019	3000	MSRA-B.	VGG-16
SPBR [123]	arXiv	2019	5000	MSRA-B, HKU-IS	VGG-16

Table 2. A brief summary of deep-learning-based saliency detection models.

a framework, which works for both salient-object-detection and semantic segmentation. They replaced the originally connected layers in VGGNet [89] with convolutional-layers. Li et al. [111] designed s a multi-scale CNN to simultaneously locate contours and regions for salient object detection. In [124], a deep architecture is exploited to pick up a small amount of candidate bounding boxes/regions that are well-segmented to provide support in the generation of the salient maps. A CRF model [125] is applied to refine the spatial coherency. In [112], Hou et al. proposed a structure model that semantic information from upper layers is propagated to lower layers for locating salient objects. Chen et al. [110] presented a coarse-to-fine based approach, in which progressive representation learning are used for saliency map prediction. Wang et al. [113] presented a stage-wise scheme established on spatial pyramid pooling method [126] which combines multi-scale global contextual priors and fuses high-level syntactic information (ciphered in the master network layers) along with the contextual rich information of low-level features (ciphered in the refinement network module). In [115], Yuan et al. propose a dense and sparse-labeling network for saliency detection. Inspired by the MumfordShah (MS) functional loss [127], Luo et al. [114] proposed a non-

local deep feature (NLDF) framework, which captures local and global features via a multi-resolution grid structure during saliency detection. Fu et al. [122] proposed a refinement- network (refinet) model to firstly locate boundaries of salient objects and then generated a saliency map through the refinet framework. Zhang et al. [1], proposed a multi-level attention-guided network by introducing multi-path recurrent feedback to utilize the local and global information. In Li et al. [2] proposed a Contour-to-Saliency network approach, which can generate saliency masks from a well-trained contour network and feedback the result for further training. the model updates the parameters gradually during training. Jiang et al. [123] proposed a pooling-based approach and merged two independent CNNs to collect global and local information. Chen et al. [118] employed residual learning and reverse attention at side-output and obtained a concise model appropriate for embedding devices. In [119], Zeng et al. merged three independent CNNs for global-features, local-features, and spatial consistency. Feng et al. [120] designed Attentive Feedback and Boundary-Enhanced Loss for extracting structure-wise and boundary-wise features. Similarly, in [121], Qin et al. proposed a predict-refine architecture with an encoder-decoder module to get a saliency map with more refine boundaries.

Discussion:

As compared to region-based CNNs models, FCNs based models are the end-to-end based CNNs models utilizing pixel-level values for predicting saliency-maps, and hence, also called pixel-to-pixel CNNs models. The FCN-based approaches are very efficient and overcome the limitation of region-based CNNs models. It can also preserve the contextual information in a very good manner and hence, provide a more robust result. As region-based CNN models use a separate network for utilizing local and global features, the FCN-based models learn local and global features in one network. While the shallower layers provide global information and more details about edges of the object, while the deeper layers provide the high-semantic, local and more meaningful information. These FCN-based networks are mostly pre-trained/learned on ImageNet dataset [128] for image classification purpose, and these learned models can be then fine-tuned for multiple purpose (e.g., object detection [129], object-localization [130], and saliency detection [96,122]. The pre-trained models minimize the training cost and provide more sophisticated results than training from scratch. Furthermore, the FCN models contain a stack of different types of layers, which can perform a different type of function, and hence, provide structure-wise flexibility and diversity than previous region-based CNNs models. A brief summary of deep learning-based models is shown in Table 2, and a visual comparison of some conventional heuristic-based and new learning-based methods is shown in Figure 7.

Although Deep learning techniques, especially FCNs based methods, have achieved a very great performance, yet it fails in many circumstances that need to improve in the future. For example, it needs improvements in low-contrast images, which have more common foreground and background similarity, transparent objects, and images that contain complex backgrounds. Similarly, the repetition of poolings and strides operations in FCNs minimize image resolution and degrade the performance of the models. more time and large memory is also a challenging issue for these deep models. Also, these methods require a large amount of training data.

To resolve these issues, there are several different types of CNN-based architectures proposed in recent years. Some approaches have shown tremendous response and need to be further explored in the future. For example, multi-scale and multi-level deep networks can utilize the features at different layers by using fusion, skip-connections, and short-connections among different levels. Similarly, the encoder-decoder architecture is the most promising approach and has shown a great performance in different classification and segmentation tasks. In these types of methods, the high-level features are back-propagated to lower-layer and making a stronger union of multi-level features. Another good approach for the promising result is to use ResNet [131] which is a deep network and can perform the complicated task very well. ResNet is more powerful than VGGNet [132]. The fusion of different cross-models also can boost performance. A standard training-loss function can also boost performance and require more attention in the future. Similarly, the embedded applications such as mobiles, robotics, autonomous driving, etc., need a lot of research in the salient object-detection area to reduce time, memory space and energy consumption

2.3 RGBD saliency detection

RGBD saliency detection is an emerging topic and still has a large research gap for improvements. Dissimilar from 2D-image saliency detection methods, the depth cue has to be incorporated in saliency detection for 3D-images. RGBD saliency detection methods utilize color information and depth cue at the same time to identify the salient-object. There are commonly two ways to incorporate the depth cues with 2-D images[133]: (1) Depth feature-based methods [134-139], which aim to incorporate the depth facts as an additional material

along with color measurements. (2) Depth-measure based methods [140-143], which capture the comprehensive information from the depth cue, such as shape and structure via utilizing designed Depth-measures.

These are the hand-crafted-features based methods with depth cues to detect a salient-object in an image. Various studies have worked on saliency detection for 3D multimedia content. Lang et al. [134] perceived salient-objects by incorporating global-context depth priors into 2D models. Ju et al. [135] presented the RGBD saliency process created on anisotropic center-surround variance, in which saliency is estimated as how much an object is different from its surroundings. In [139], Fang et al. extracted color, texture, luminance, and depth feature from the RGBD based images to estimate the contrast feature maps. Then, the combination and improvement methods are exploited to get the resultant 3D saliency-map. Song et al. [136] utilized the depth information as a regional feature for computing low-level contrast-based saliency, and also used as a weighting feature for measuring mid-level saliency. Then high-level location priors are applied to build the high-level saliency-map. In the last stage, a multiscale discriminative saliency fusion technique is applied to combine the multiple saliency-maps and get the concluding saliency output.

Furthermore, motivated by the assessment that the salient-regions are definitely dissimilar from their local and global surroundings in the depth feature map, a "depth contrast" is a general depth property to be calculated. For this purpose, Niu et al. [138] computed global-contrast with domain knowledge to estimate the stereo saliency. In [137], Peng et al. proposed a multi-contextual contrast framework for calculating depth saliency by considering the contrast-prior, global uniqueness, and background-prior to the depth-map. Then a multi-level RGBD saliency approach is exploited to fuse the low contrast features, medium-level local alliance, and high-level prior techniques.

Ju et al. [140] proposed a depth-aware framework for saliency detection by applying an anisotropic center-surround difference (ACSD) measure, Furthermore, they built a huge dataset for stereo saliency detection, which contains 1985 stereo images and estimated depth-maps. Coalescing the ACSD measure method with color saliency-map, In [141], Guo et al. proposed a salient-object-detection model for RGB-D images established on evolution strategy. It is a re-iterative generation process to enhance the early saliency-map and produce the final output. As the backgrounds include the regions that are extremely mutable in depth-map, some high contrast background regions might raise false-positive. To get a ride over this disturbing, Feng et al. [142] used a Local-Background-Enclosure measure (LBE) framework to straightly extract a salient region from depth-map, which calculates the ratio of object margins located in frontal of background. Wang et al. [143] introduced a multi-stage salient-object-detection scheme for RGBD images by joining the Minimum-Barrier Distance transform saliency-map and multi-layer cellular automata-based saliency-map.

Recently, deep learning [144-146,37] is also applied in RGBD saliency detection to learn more discriminatory RGBD features. In [144], Qu et al. proposed a CNN model for RGBD saliency detection. They combined the low-level saliency features such as local-contrast, global-contrast, spatial-prior and background-prior and generated coarse saliency vectors. These vectors are then combined with depth modalities and fed into CNN to train it from scratch to produce the RGBD hyper-features. Han et al. [145] proposed a two-stream late-time fusion structure to combine RGBD deep features. A stage-wise approach is followed to train the network and obtained optimistic performance. Similarly in [37], Wang, et al. proposed RexNet which produces end-to-end saliency-map with a sharp-edged object. In this method, first, the image is divided into two independent segments: edge regions and superpixel regions. The network then produced end-to-end saliency score for these regions, and the context in multiple layers are combined with regional saliency scores. The proposed model is then extended to RGBD saliency detection by applying depth refinement. Chen et al. [146] proposed an end-to-end RGBD salient-object-detection network, which is correspondent-aware for combining cross-modal and cross-level features. The presented cross-modal connections and level-wise supervisions clearly motivate the capturing of complementary facts from the counterpart, and thus, growing fusion capability by decreasing fusion uncertainty. In [147], Wang et al. proposed a two-stream CNN by utilizing a fusion strategy. Similarly, in [148], Liu et al. proposed a fusion-based two-stream network for RGBD saliency detection. The depth structure information help in the foreground and background identification. Then a propagation-based module is used for the identification of object boundaries.

Discussion:

Currently, there are three ways to capture the depth-map for 3D-images: (1) structured light technique [149] are used to extract the depth information by the variation of a light signal produced by the camera. This is a good technique but mostly sensitive to illumination. (2) Time-of-Flight (TOF) [150], utilize the round-trip time of the light signals for estimating the depth cue. This is also a robust technique but commonly has a low resolution. (3) The stereo imaging system (i.e.,

a b c

Figure 3. A 3D saliency conditions in RGBD images. (a) Color-depth saliency: both RGB images and depth images are salient. (b)Color saliency: only RGB images are salient. (c) Depth saliency: only depth images are salient.

binocular imaging) [151], captures two photos by using two cameras at different positions and finds the distance of the object through triangular rules. This method has a low cost but requires post-processing steps. So, it is true that RGBD images need further research on how to get good quality depth information and then how to utilize it in a proper way because the improper use of the depth information leads to performance degradation. Figure 3 shows some different conditions of depth-maps, and Table 3 represents a brief summary of RGBD based saliency detection.

Model	Pub	Year
DM[134]	ECCV	2012
DSA [135]	ICIP	2014
SDS [139]	TIP	2014
DSM [136]	TIP	2017
LSA [138]	CVPR	2012
ROD [137]	ECCV	2014
DSDA [140]	SPIC	2015
ISE [141]	ICME	2016
LBE [142]	CVPR	2016
MBDT [143]	SPL	2017
RDF [144]	TIP	2017
CTF [145]	ITC	2017
EPMC[37]	TIP	2018
PCF [146]	CVPR	2018
AFD[147]	IEEE Access	2019
TSR [148]	ICIP	2019

Table 3. A brief summary of RGBD saliency detection.

2.4 Co-Saliency-Detection

Co-saliency-detection is the process that tries to discover the most common and salient-objects from a given group of images. For this purpose, the inter-image correspondence feature is used as a simple attribute check to distinguish the shared objects (attributes-wise) from all other salient-objects. The low-level or high-level features are first calculated for every image in the sequence to obtain a co-saliency-map. The low-level features are the heuristic characteristics of an image, represent color, texture, and luminance, etc. while the high-level features represent the semantic information obtained via deep learning techniques, two types of models are utilized to extract the intra-image and inter-image features for co-saliency detection. The intra-image saliency models are used to extract a feature from an individual image, and the inter-image saliency models are used to extract the features from a group of images. For intra-image co-saliency, the common saliency detection methods can be utilized, however, the inter-image models use different types of techniques, such as similarity based-matching, low-rank based analysis, clustering, and method of propagation. After calculating these two types of models, a fusion scheme is utilized to incorporate these models and obtain a final co-saliency-map.

Co-saliency-detection is often nearly correlated to the notion of a co-segmentation scheme that plans to segment most identical objects or regions from multiple images [152]. As indicated in [153], there are three main variations between the co-saliency process and the co-segmentation process. First, Co-saliency-detection approaches focus only on encountering the salient-objects that are common, while on the other hand similar non-salient parts of the background can also be considered in co-segmentation methods [154,155]. Second, a few co-segmentation approaches, e.g.,[156], want user response to lead the process of segmentation in a vague situation. Third, salient-object-detection frequently performs as a pre-processing step, and hence more real and efficient approaches are favored than co-segmentation approaches, particularly over a huge number of images.

The traditional-based methods are basically the earliest and the simple methods for Co-saliency-detection by using hand-captured co-saliency features for scoring each pixel/region in the image group. Generally, these are low-level methods that are comprised of four basic components containing pre-processing, feature extraction, applying low-level cues, and weighted combination.

Chang et al. [157] proposed a fully unsupervised method to resolve the co-segmentation problem. They produced an optimized CRF model by establishing a co-saliency prior to the clue about conceivable foreground locations to substitute user input data and a unique global-energy term to get the co-segmentation procedure efficiently. Tan et al. [158] presented an autonomous Co-saliency-detection scheme that originated on the similarity matrix, which measures the

co-saliency process by using the bipartite superpixel-level mechanism of graph matching across the set of image pairs. Fu et al. [153] presented a cluster-based Co-saliency-detection approach by utilizing the global contrast and spatial distribution cues on a single image, and use the corresponding cues over a group of images to find the saliency co-occurrence. Li et al. [159] presented a co-saliency model by utilizing a low-rank matrix recovery scheme for computing intra saliency detection and a region-level fusion scheme. The region-level fusion scheme utilizes the similarities that exist among different regions and the global uniformity measures over the image set. The pixel-level refinement scheme is utilized to measure the similarities between pixel and region as well as their object priors. Ye et al. [160] proposed a saliency detection framework based on object discovery and recovery using gross similarity matching. They first generated an exemplar saliency map by discovering the consistent exemplars for co-salient objects. Then a local and global recovery of co-salient object regions, foci of attention area and border connectivity of the regions are exploited to create final co-saliency maps for all corresponding image set. Li et al. [161] introduced a saliency-guided co- saliency detection scheme, where the first step recuperates the co-salient chunks, lost in the single saliency map by using the efficient manifold ranking scheme, and the second step extracts the correlated relationship via a ranking scheme with different types of queries. Ge et al. [162] proposed a two-stage propagation method for co-saliency detection, where the inter-saliency propagation stage is exploited to recognize shared features and build the pairwise shared foreground cue maps, and the intra-saliency propagation stage is utilized to suppress the background locations and refine the processing of the first stage. Song et al. [163] proposed an RGBD Co-saliency-detection model by using bagging-based clustering. The candidate object regions are created by utilizing region pre-segmentation and RGBD single saliency maps. Then a clustering via feature bagging technique is executed recurrently to compute various weak co-saliency measures based on the cluster level. Finally, an adaptive fusing multiple (WCS) map is utilized to evaluate the clustering quality. In [164], Huang et al. designed a scheme for Co-saliency-detection by considering color feature reinforcement method, and co-saliency map are obtained by utilizing feature coding coefficients and salient foreground dictionary.

In [165], Cong et al proposed an energy function refinement and hierarchical sparsity reconstruction framework for RGBD co-saliency detection. A hierarchical sparsity reconstruction scheme is utilized to formulate the inter-image correspondence with the help of an intra saliency map. The global sparsity reconstruction framework is utilized with the ranking scheme and captures the global characteristics among the entire image via a common dictionary, and the pairwise sparsity reconstruction model is utilized to find the co-relationship among the images via a set of a pairwise dictionary. Finally, an energy function is adapted to improve inter-image consistency and intra-image smoothness. In [166], Li et al planned a low-rank weighted Co-saliency-detection framework through a two-stage EMR. A two-stage ranking method is utilized to create multiple co-saliency maps for each input image, and then for each image, a group of variable sizes of salient regions is extracted and fused the co-saliency maps with their corresponding superpixels. Then an adaptive weight for each co-saliency map is designed via sparse error matrix. Finally, the co-saliency maps and their corresponding weights are multiplied to obtain the fusion results and optimized further by using Graph Cuts.

Recently, learning-based Co-saliency-detection methods have attracted much research attention and attained a reasonable performance, comprising deep learning, self-paced learning, and metric learning. These methods directly learn the features of the co-salient-objects from a given image group, instead, relying on hand-crafted cues. In [167] Zhang et al proposed a co-saliency object detection framework by introducing looking deep and looking at wide perceptions under the Bayesian framework. The term looking deep aims that the high-level features are extracted by using CNN with multiple layers to discover better representation, and the term looking wide tries to detect some visually identical neighbors to effectually suppress the mutual background regions. Zhang et al. [168] proposed a self-paced multiple-instance-learning (SP-MIL) framework by integrating the MIL and SPL models, where the Multi-Instance-Learning (MIL) model specifies to train a predictor for every instance via rising inter-class differences and reducing the intra-class difference. The self-paced learning (SPL) aims to progressively learn from the easy/faithful examples to more composite/confusable ones. In [169], Wei et al. proposed a pixel-to-pixel based group-wise deep Co-saliency-detection framework. A block of thirteen convolutional-layers are introduced to capture the basic features, and then, the group-wise properties and individual properties are extracted to specify the group-wise properties and single image properties. Finally, a combined learning scheme with the convolution-deconvolution process is devised to get the co-saliency map. To cope with the wide variation in the image scene, Han et al. [170]proposed a metric learning co-saliency model through a new objective function, in which metric learning aims to learn a distance metric to bring the

same-class sample closer and make the different-class samples far away from each other.

.

Figure 4. An example of Co-saliency-detection by using the iCoseg dataset. The 1st row displays the input images and the 2nd row represents the corresponding ground-truth images

Discussion:

Co-saliency-detection is an emerging topic for the research community and achieved considerable progress in the last few years, there is still a very large space for future improvement in this field. Here we enlist some major issues that need development in this field: (1) image complexity, co-saliency models need considerable improvements for complex and clutter images. (2) if the foreground consists of different types of objects with multiple colors, then it is difficult to find only salient objects. (3) Co-saliency cannot perform well on large-scale data, because it contains more outliers, noise and variation (4) co-saliency models are not efficient and consume more time. (5) Inter-correspondence constraint needs a lot of improvements, to effectively monopolize the common attributes among multiple images. A summary of co-saliency techniques is presented in Table 4 and Figure 4, shows the common salient objects among several images.

Model	Pub	Year	Main techniques
FCC [157]	CVPR	2011	based on clustering and similarity matching
SA [158]	ICASSP	2013	Based on superpixel-level graph matching
CCS [153]	TIP	2013	Based on clustering-with-multiple cues
CRPR [159]	ICME	2014	Based on low-rank matrix recovery and similarity matching
CODR[160]	SPL	2015	Based on gross similarity-ranking
SCS [161]	SPL	2015	Based on ranking-scheme
CSP [162]	SPIC	2016	Based on two-stage propagation
CDBC[163]	SPL	2016	Based on bagging clustering
CFR [164]	SPL	2017	Based on color-feature reinforcement
CLDW [167]	CVPR	2015	Deep-learning based on Bayesian framework
SMIL [168]	ICCV	2015	self-paced based on multi-instance-learning
GWDC [169]	APA	2017	Pixel-to-pixel deep co-saliency network
UMLCD [170]	TCSVT	2018	Distance based metric-learning
HSCSR[165]	ITOM	2018	Based on hierarchical sparsity reconstruction and global sparsity reconstruction ranking scheme
EMR [166]	MTA	2019	Based on a low efficient manifold ranking
MGFCN [171]	CVPR	2019	Based on the mask-guided fully convolutional network

Table 4. A brief summary of Co-saliency detection.

2.5 Video Saliency

Video sequences utilize the sequential feature, motion and color appearance information for the perceiving and identification of scenes. In video-saliency, an object is salient if it has some repetition, motion-relevancy and some other distinctive targets in the video sequences. These are the unsupervised methods exploiting the low-level cues, such as color-appearance, motion-cue, and some other prior constraints. The traditional-based video-saliency methods further split into the Fusion-based Model and Direct-pipeline-based Models [133]. Fusion-based models first compute the spatial saliency (i.e., spatial-cue, describe the intra-frame information in each frame) and their corresponding temporal-saliency (temporal cue, represents the inter-frame association among different frames). Then, the results of these two saliency-maps are combined to obtain video-saliency-detection. Spatial saliency detection utilizes the center-surround, contrast-prior, background-prior, sparse re-construction and low-rank analysis to get the saliency representation in each separate frame, while the temporal saliency

detection exploits the motion cue to describe the moving objects in the video.

Fang et al. [172] obtained static saliency using luminance, color and texture features in a compressed domain, and get motion saliency using motion cue and then, a fusion method is utilized to achieve the final saliency-map for each video frame. Ren et al. [173] obtained a spatial saliency by using a sparse reconstruction method to detect the regions with high center-surround contrast. For temporal saliency, a reconstruction process for the target patch and their neighboring overlapping patches are used to reconstruct the target patch. Finally, a fusion mechanism is applied for video-saliency. In [174], Liu et al. extracted superpixel-wise low-level features and frame-wise global features for spatial saliency. For temporal saliency integrated the motion uniqueness of superpixels and finally fused the spatial and temporal saliency-maps by using the adaptive fusion method. Xi et al. [175] used background-prior for spatial saliency and SIFT flow and bidirectional consistent propagation for temporal saliency and fused these both saliencies by using simple addition to get the final saliency. In [176], Chen et al used color contrast and gradient guided contrast for spatial and temporal saliency-maps respectively and applied a fusion method to get the final saliency.

The models in this class use spatiotemporal features to directly discover the salient-object. Xue et al. [177] used a low-rank and sparse decomposition scheme on video slices as a temporal feature and separated the foreground from backgrounds. The spatial information is utilized to keep the completeness of the discovered motion objects. Wang et al. [178] proposed a spatiotemporal saliency approach built on the gradient flow and energy improving scheme, which is good for complicated scenes, different motion arrangements, and dissimilar looks. The gradient flow field describes the salient parts by integrating the intra-frame and inter-frame. Liu et al. [179] proposed a dynamic pipeline scheme for video-saliency-detection by utilizing the graph-based motion saliency based on superpixel-level, spatial propagation, and temporal propagation. Guo et al. [180] presented the video-saliency method by computing spatial saliency and motion saliency and then applied object proposal scheme for ranking and voting, to filter non-salient-regions and estimated the initial saliency. Finally, initial saliency is refined by considering temporal consistency and appearance diversity. In [181], Kim et al. random walk with restart is used to identify the salient-object, in which the temporal consistency and motion distinctiveness are exploited to extract temporal consistency and a quick variation is utilized as the restarting distribution of the random walker. Similarly, [182] and [183] proposed a geodesic distance-based method to compute superpixel-wise saliency by using undirected inter-frame and intra-frame graphs constructed from spatiotemporal edges, appearance, and motion. In summary, fusion techniques are comparatively more natural than direct-pipeline techniques. Furthermore, the spatial saliency methods are image saliency methods which can provide a basis for spatiotemporal saliency and can be used directly in video-saliency.

Indeed, deep-learning-based video-saliency methods have demonstrated a great performance over the existing traditional-based (hand-crafted features based) methods. These learning-based methods independently extract the features from each individual frame and then utilize frame-by-frame processing to calculate saliency. Le et al. [184] presented a deep learning model to extract the Spatio-temporal deep-features (STF). The region-based CNN is applied to extract the local features and the global features are extracted from temporal-segments by using a block-based CNN. Using the STF features, a Random Forest (RF) and Spatio-temporal CRF (CRF) are presented to achieve the ultimate saliency. In [185], Wang et al. proposed a deep learning model to detect salient-objects in the video. The static network generates a fixed saliency-map for every frame using FCNs and then the frame-pairs map and static saliency are fed into a dynamic network to generate the dynamic saliency-map. Le et al. [186] presented an end-to-end 3D Recurrent Fully-Convolutional-Network (DSRFCN3D) for salient-region-detection in video streams, which contains an encoder, decoder and refinement networks respectively. The encoder network captures 3D features (both spatial and temporal information) from a feeding video block. The decoder network estimates the precise saliency voxel from the 3D deep feature by gradually refining the intermediate saliency voxel through supervised learning at every hidden 3D deconvolution layer [101]. On the other hand, the refinement method along with skip-connection layers and 3D recurrent-convolution-layer (RCL) is designed to learn the relevant contextual evidence. In [187] Li et al introduced an unsupervised video-saliency by using the saliency-guided stacked scheme of autoencoders. First, the saliency cues captured from the spatiotemporal acquaintances at three different stages (i.e., pixel-, superpixel- and object-levels) are collected as a feature-vector of high-dimension properties. In the second step, the initial saliency-map is obtained by learning the stacked auto-encoders by the unsupervised way. At last, some post-processing actions are applied to further enhance the salient-object and demolish the false clue., Similarly, Cong et al. [188] proposed a sparse reconstruction and propagation method to detect salient objects in video.

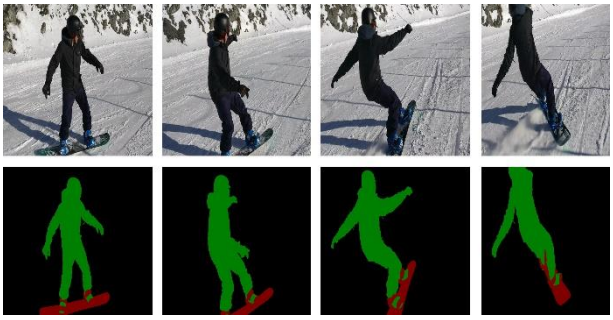


Figure 5. A video-saliency-detection example on the DAVIS dataset. The first row represents the original video frames of input data and the second row represents the corresponding ground-truths.

Discussion:

To sum up, video-saliency-detection is also an emerging field for future research, as it is largely unexplored and there are still many challenges that need to be addressed. The key issue in video-saliency-detection is how to abolish the background and fixed objects in order to find more relevant salient items in the video. For this purpose, mostly optical flow is used, but it is not an efficient technique and also does not provide much more accuracy. Recently deep learning techniques outperformed the traditional techniques, but the major issue in deep learning is the non-availability of large annotated datasets for video-saliency-detection. The next key issue in the video-saliency-detection is to find robust techniques to capture the inter-frames attributes that provide a consistent appearance saliency-map for all frames, for this purpose some energy function is adapted to improve the consistency, but still, it needs further improvements. Video-saliency also needs improvements, where most of the frames consist of complex backgrounds and multiple objects. A video-saliency-detection summary is shown in Table 5, and some example video-frames are shown in Figure 5**Error! Reference source not found.**, which shows the same salient object among different frames of the same video.

Model	Pub	Year	Main techniques
LRSD [177]	ICASSP	2012	Low-rank , sparse decomposition and spatial information about object completeness
SSDSR [173]	ICME	2012	Sparse reconstruction process for both temporal and spatial-saliency
SBSSD[174]	TCSVT	2014	Superpixel-level motion features as a spatiotemporal and global contrast and spatial sparsity as a spatial saliency
VSDMC [172]	TCSVT	2014	luminance, color, and texture for static saliency map and motion saliency map
GFGR [178]	TIP	2015	Gradient flow field for salient regions, then local and global contrast and energy optimization function
SGVS [182]	CVPR	2015	Geodesic distance is used for Spatiotemporal saliency map, global appearance, and location features
STBP [175]	TIP	2017	Spatiotemporal background prior, SIFT flow and superpixel
SGSP [179]	TOC	2017	Superpixel-level graph, temporal propagation and spatial propagation
VOP [189]	TOC	2017	Object proposal ranking and saliency refinement optimization process
USGD [183]	PAMI	2018	Geodesic distance and energy optimization techniques
SUDF[184]	ICME	2017	Deep learning STF feature, STRCP, and Random Forest
VFCN[185]	TIP	2018	Directly capturing spatial and temporal saliency information with the help of deep learning
DSFCN [186]	BMVC	2017	An end-to-end 3D FCN method learns spatial and temporal information directly
DSVS[190]	TIP	2019	3D stereoscopic video saliency with two main STSM and SSAM modules
SBRP [188]	ITOIP	2019	A sparse reconstruction and propagation-based approach
TASED-Net [191]	ICCV	2019	An encoder-Decoder-based approach

Table 5. A brief summary of video-saliency detection Models.

3. Datasets and Applications

3.1 Datasets for saliency detection:

In this section, we presented the most common datasets used for saliency detection techniques like RGB-D saliency-detection, co-saliency-detection, and video-saliency-detection. As the advancement in saliency detection techniques, more challenging datasets have been introduced to further challenge the state-of-the-art models. The early datasets contain a very simple background and a single image in the foreground, having the ground-truth being annotated with the bounding-box methods, such as MSRA-A and MSRA-B [192]. The recent datasets are very complex and cluttered background having more than one object, being annotated with pixel-level ground-truth annotation, Pixel-based annotation datasets carry more

accurate results than bounding-box annotation.

For simple RGB image saliency detection, we collected a total of 10 datasets, as shown in table 6, such as Judd-A[193,5], UCSB [194], OSIE [195], ECSSD [25], DUT-OMRON [61], MSRA10K [196], ACSD [51], PASCAL-S [197] and XPIE [198]. There are some datasets which also hold the fixation data, collected for each image during the free-viewing process, such as Judd-A, UCSB, and OSIE. The list of RGBD datasets consists of RGBD1000 [137], NJUD [140], DES [199] as shown in Table 7. some example images from PASCAL challenging dataset is shown in Figure 6.

For Co-saliency-detection we listed a total of 8 datasets that are used commonly, as shown in Table 8. The first 5 are for simple RGB image saliency detection, such as MSRC [200], iCoseg [156], Image Pair [201], Cosal2015 [202], INCT2016 [203], which comprises more than two images in each group except Image Pair dataset which contains only one image pairs. The last 2 datasets, of Table 8 , such as RGBD Coseg183 [204] and RGBD Cosal150 [205] are for RGBD co-saliency-detection. The RGBD Coseg183 is a dataset, containing 183 images with depth-cue, distributed in 16 groups. The RGBD Cosal150 dataset, have 150 RGBD images, distributed in 21 image groups.

For video-saliency-detection, there are several datasets available, such as USVD [179], ViSal [178], SegTrackV1 [206], SegTrackV2 [207]], MCL [181], VOS [187] and DAVIS [208], as shown in Table 9. The DAVIS dataset is one of the frequently used and more challenging datasets, containing 50 video series along with pixel-wise ground-truth for every video frame. The UVSD dataset is a new dataset and particularly designed for video-saliency-detection which contains 18 unrestricted videos with complex motion patterns and more scattered scenes, with pixel-wise annotated ground-truth for each video frame. An extended video-saliency-detection dataset called VOS is created, which comprises 116103 total frames that distributed in two-hundred (i.e. 200) video sequences. This dataset contains 7467 binary ground-truth annotated frames, which is good enough to train and learn a deep learning model to capture the salient-objects in the video.

A dataset is the collection of data for a specific application domain. Unfortunately, each dataset may suffer from different types of biases, which can affect the performance of the models. For example, Torralba and Efros acknowledged three biases in the field of computer vision, called selection bias, capture bias (i.e., center-bias) and negative-set-bias [209]. Selection bias occurs, when someone prefers a specific type of image during data assembling and it may produce an error because the individual prefers his own choice while violating standard rules for selection. The selection bias collects more similar images in the dataset and hence, lacks variability in the dataset. To avoid selection bias, it is necessary to have an independent selection. The Capture bias transmits the effect of image structure into the dataset (i.e., People tend to capture the images of similar objects in a similar way), which also lack variability in the dataset. For example, center-bias means that most of the captured objects lie in the center of images. This type of bias makes the dataset challenging for quantitative comparison and sometimes even produces an ambiguous comparison. For example, a petty saliency method that contains a Gaussian blob at the center of an image, always produce the best score than many fixation prediction methods [79]. The Negative-set bias represents that an individual personally not like to include a particular object into the dataset, while a dataset must represent every possible thing. The Negative-set-bias can disturb the ground-truth by employing the annotator's particular favorite to some particular object. Hence, it is encouraged to have more varieties of images in a good dataset.

3.2 Applications of saliency detection:

Saliency-detection technique is usually used in the field of image retrieval [210,211], image segmentation [212-214], object discovery [214], target detection and cognition [215-219], video summarization and skimming [220,221], image and video compression [222], image resizing, image automation pruning [223], content-based image retrieval [224-226], photo collage [43,227] image editing and manipulating [228,229], human-robot interaction [230,231] and visual tracking [36,232,233].

3.3 Evaluation Measures

The qualitative and quantitative evaluation techniques are the two common techniques to assess the performance of salient-object-detection models. The qualitative technique visually compares the predicted saliency maps with their corresponding ground-truth masks. It is the more simple technique but it has no fixed value and hence, varies from person to person. On the other hand, a quantitative evaluation gives a fixed value, acceptable for each observer. There are different types of evaluation techniques available in the literature for comparing predicted saliency maps with their corresponding ground-truth. Here we only discuss the standard top-five techniques that consider as a standard in salient object detection. All of these techniques consider overlapping regions between predicted maps and their corresponding ground-truth masks. For mathematical notation, we use G for ground-truth mask and S for predicted saliency map.

We use $|\cdot|$ for both binary masks to indicate the number of entries in the mask.

Figure 6. Some example images from PASCAL challenging dataset.

Dataset	Pub Year	Image No	Max Res.	Object property	Background Property
ACSD [51]	2009	1K	400 × 400	Single moderate	Clean, simple
ECSSD [25]	2012	1K	400 × 400	Single, large	Clean, simple
DUT-OMRON [61]	2013	5168	400 × 400	Single, small	complex
Judd-A[193,5]	2014	900	1024×768	Single, moderate	Clean and simple
UCSB [194]	2014	700	405 × 405	Single, large	Clean and simple
OSIE [195]	2014	700	800 × 600	Multiple, moderate	simple
MSRA10K [196]	2014	10K	400 × 400	Single, large	Clean, simple
PASCAL-S [197]	2014	1K	500 × 500	Multiple, moderate	simple
HKU-IS [85]	2015	850	400 × 400	Multiple, moderate	complex
XPIE [198]	2017	4447	300 × 300	Single, moderate	complex

Table 6. A list of salient-object detection datasets for RGB.

Dataset	Pub Year	Image No	Max Res	Object property	Depth attribute
RGBD1000 [137]	2014	10K	640 × 640	Single, moderate	Kinect capturing
LFSD dataset [75]	2014	100	1080 × 1080	Complex	Lytro light field capturing
DES [199]	2014	135	640×480	Single, moderate	Kinect capturing
NJUD [140]	2015	2K	600 × 600	single, moderate	depth estimation

Table 7. A list of salient-object detection datasets for RGBD Images.

Dataset	Pub Year	Image No	Group No	Group size	Max Res.	Object property
MSRC [200]	2005	230	7	30-53	320 × 210	Complex
Caltech[234]	2006	101	257	30607	500 × 800	Complex
iCoseg [156]	2010	643	38	4-42	500 × 300	Multiple
Image Pair[201]	2011	210	115	2	128 × 100	Single
Cosal2015[202]	2016	2015	50	26-52	500 × 333	Multiple
INCT2016 [203]	2016	291	12	15-31	500 × 375	Multiple
RGBD Coseg183[204]	2015	183	16	12-36	640 × 480	Multiple
RGBD Cosall50[205]	2018	150	21	2-20	600 × 600	single

Table 8. A list of Co-saliency-detection datasets.

Dataset	Pub Year	Frame No	Video No	Video size	Max Res.	Object property	Background Property
SegTrackV1 [206]	2010	244	6	21-71	414 × 352	Single	Diverse
SegTrackV2[207]	2013	1065	14	21-279	640 × 360	Single	Diverse
FBMS [182]	2014	13860	59	720	960 × 540	Single	Diverse
ViSal [178]	2015	963	17	30-100	512 × 228	Single	Diverse
MCL [181]	2015	3689	9	131-789	480 × 270	Single, small	Complex
DAVIS [208]	2016	3455	50	25-104	1920×1080	Multiple	Complex
VOS-E	2016	49206	97	83-962	800 × 640	Single	Simple
UVSD [179]	2017	6524	18	71-307	352 × 288	Single, small	Clustered, complex
VOS [187]	2018	116103	200	~500	800 × 800	single	Complex

Table 9. A list of video-saliency-detection datasets

1. Precision-recall (PR)

Precision and Recall can be calculated by translating the saliency-map S into a binary mask B and then comparing B with its corresponding ground-truth G.

$$Precision = \frac{|B \cap G|}{|B|}, and\ Recall\ = \frac{|B \cap G|}{|G|}. \quad (1)$$

The key phase in this process is the binarization of S to B. Three most frequent methods such as fixed threshold, adaptive threshold [51] and GrabCut method [235]masks are used for the binarization process.

2. F-Measure.

Precision and recall cannot comprehensively estimate the excellence of the saliency map. For this purpose, the F-measure method The qualitative as a harmonic weighted-mean of the Precision and Recall methods with a non-negative weight β^2

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

whereas P is the Precision and R represents Recall. the β^2 value is often set to 0.3 to raise the weight of precision more than recall [51].

3. Receiver-Operating-Characteristics (ROC) curve.

Similarly, true positive rates (TPR) and false positive (FPR) can be calculated by applying a fixed threshold during saliency-map binarization.

$$TPR = \frac{|B \cap G|}{|G|}, and\ \text{FPR}\ = \frac{|B \cap G|}{|B \cap G| + |\bar{B} \cap \bar{G}|}, \quad (3)$$

where $\overline{B}$ and $\overline{G}$ indicate the complement sets of binary mask B and ground-truth G correspondingly. The ROC curve is the plotting of TPR values against FPR values by trying all probable thresholds.

4. Arear under the ROC curve (AUC).

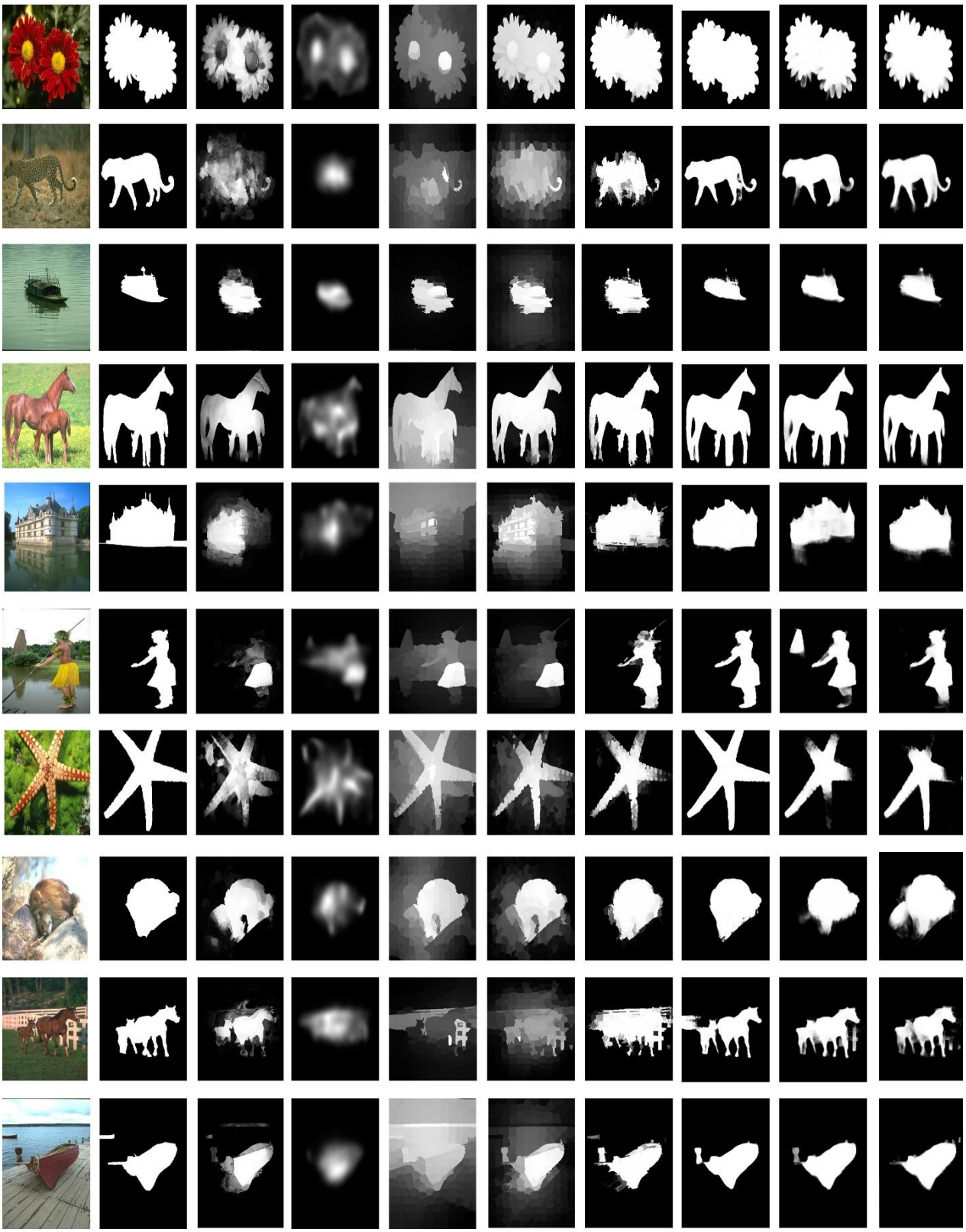
As the name indicates, it is computed as the area under the ROC curve. The performance of AUC over a perfect saliency method will get exactly 1 score, while the performance of AUC at random guessing will get around about 0.5 scores.

5. Mean-Absolute-Error (MAE).

The above overlap-based assessment measures actually do not focus on the assignment of the true negative saliency value (i.e., the pixels marked correctly as non-salient). They prefer those approaches that can effectively allocate high saliency values to salient pixels but mostly they neglect the detection of non-salient-regions. Furthermore, for some applications [223], the saliency-map sometimes requires more consideration than its binary mask. Hence, Mean-absolute-error (MAE) is an easy and reliable assessment metric for saliency-map. It is calculated as the average of pixel-wise absolute error between the saliency-map S and the corresponding ground-truth G, normalized to [0, 1], which is defined as follows:

$$MAE = \frac{1}{W \times H} \sum_{I=1}^{H} \sum_{J=1}^{H} \left| S_{ij} - G_{ij} \right|, \quad (4)$$

where H and W denote the height and width of the image respectively.



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Figure 7. A visual comparison of saliency-maps. (a) Original image, (b) Ground-truth, (c) DSR[9], (d) FES[69], (e) GR [72], (f) MC [58], (g) ELD[88], (h) PiCANet[236], (i) NLDF[70], (j) RAS [118].

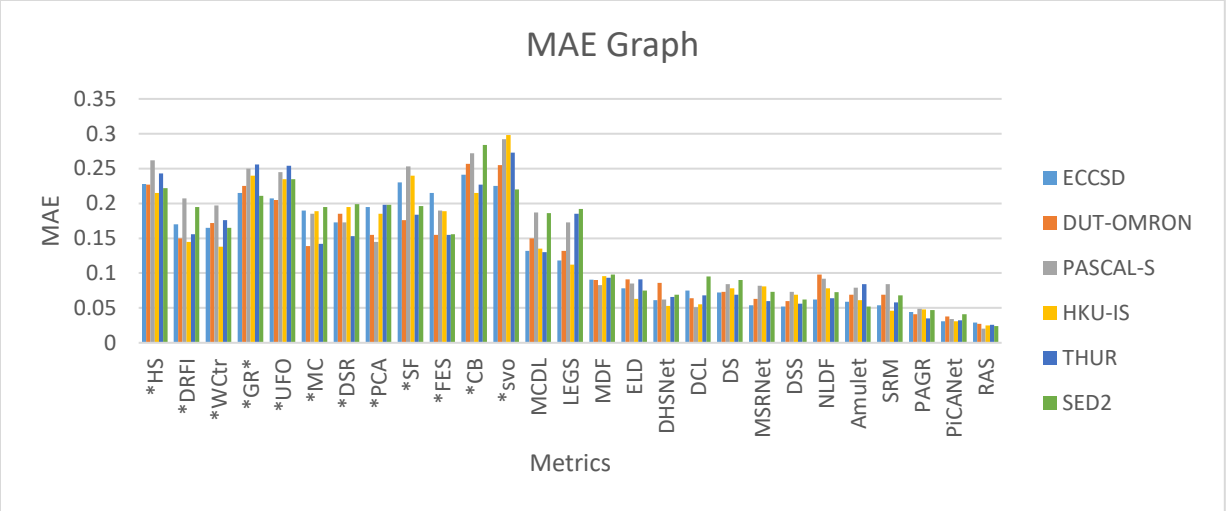


Figure 8. The MAE score graph for non-learning and learning-based models. The * means the heuristic-based salient object detection models.

3.1. Discussion and Future recommendations:

In this review, we comprehensively presented a survey on salient object detection and discussed the conventional-heuristic-based approaches and new learning-based approaches. We also discussed the co-related areas such as fixation prediction, RGBD-saliency detection, Co-saliency-detection, and video-saliency-detection. A visual comparison of some example heuristic and learning-based models are shown in Figure 7, which shows clearly that deep learning-based models outperform in the state-of-the-art models. This review provides depth insights and guidelines for upcoming progress in saliency detection. The heuristic-based approaches follow the intrinsic cues, due to which these methods are working well in a specific environment and cannot generalize well in other scenarios. Recently, deep learning-based models have shown great performance over the conventional heuristic-based methods. Deep learning-based methods follow extrinsic cues and can collect high-level semantic knowledge from large datasets, and hence have the power to generalize well in different scenarios. These methods also called task-driven methods, because they can learn features from a specific dataset and can effectively apply the learned knowledge for other environments. Although the deep-learning methods outperformed all conventional heuristic-based methods, yet they have many issues that need to be tackled in the future. the following are some considerable issues that need to be tackled in the future:

Needs large-data for training: The learning-based techniques require a large number of data for extracting features during training, it is very difficult to have a large number of data in different environments. To tackle this issue, different augmentation techniques have been proposed for creating false data. however still, the performance is not as the original dataset and needs further efforts. The other option is to design such a model that can be trained on little data. Encoder-decoder models require fewer data comparatively and require further exploration.

Dataset bias: Dataset bias also can degrade the performance of the data, if the collector violates the standard rules. For this purpose, proper knowledge will be needed to collect the dataset.

Feature-loss due to pooling and strides: In learning-based methods, the resolution of the image becomes smaller and smaller due to different pooling and stride operation and causes to lose important features during training. For this purpose, different multi-scale, multi-level, skip-connection, short-connection networks are encouraged to recover the loss features.

Manual-annotations: The learning-based methods require manual-annotations for each corresponding instance in the dataset. It is very difficult to generate large data with the corresponding pixel-level annotation. For this purpose, unsupervised-learning is encouraged in the future. Unsupervised-learning methods are most time-efficient than supervised-learning approaches.

Complex background: CNNs techniques achieved great success in simple background images. However, the complex and clutter background images still require much improvement in salient-object detection.

As we know, saliency detection has vast applications and attracted much attention from researchers. For this

purpose, the following research trends may play an important role in the future.

1. Instance level salient object detection: the recent approaches of salient-object detection are object-agnostic (i.e., the salient regions do not split into objects), however, the humans have the talent to split the salient or stimuli objects at instance-level. Instance-level saliency approach can be used in several applications, such as video compression and photo editing.

2. Flexible Network Architecture: it is verified that the deeper CNNs model can capture more accurate salient objects based on their high-level semantic knowledge. For this purpose, deeper networks like ResNet can be the more preferable choice in the future. Similarly, to avoid features losing, encoder-decoder and multi-level network can perform well in model selection.

3. Collaboration among different modules: In computer vision, the collaboration and sharing of information among common tasks such as object segmentation, object-detection, object-tracking, and object-categorization strongly boost each other. Similarly, the contextual and prior information from other modules can also boost the salient object detection. Especially, exploring the association between salient object detection, fixation prediction, and semantic perception models can benefit each other.

4. Extending the salient object detection behavior into other fields: apart from image and video, the visual-saliency concept can be extended into speech recognition, auditory perceptions, touch behavior, and scene-captioning.

5. 3D Object Detection: RGB-D images can improve the performance of salient object detection, however, there is very narrow work in this field.

6. Co-saliency and video saliency need more advanced techniques: In the case of Co-saliency detection, the inter-image correspondence technique is used to find the common salient objects among a group of images. For this purpose, different techniques have been adopted. However, it needs much consideration in the future. similarly, in video saliency, the inter-frame correspondence techniques also need further exploration to find a robust association among multiple frames for salient-object detection.

7. Interpretable deep learning Models: Inerpretablity techniques can help in understanding the predictions of a specific model in a specific scenario. By using these approaches, we can learn which type of dataset, model, and hyper-parameters can perform excellently in salient object detection.

8. Emotion-based saliency detection: The combination of visual-based saliency models with emotion-based models can be used to extend the performance of saliency detection. These models find the relationship of saliency with emotion, that how images can invoke human emotion.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (NSFC) (61601427, 61602229, 61771230); Royal Society - K. C. Wong International Fellowship (NIF\R1\180909); Fostering Project of Dominant Discipline and Talent Team of Shandong Province Higher Education Institutions.

References:

1. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on pattern analysis machine intelligence 20 (11):1254-1259
2. Hou X, Zhang L Saliency detection: A spectral residual approach. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, 2007. IEEE, pp 1-8
3. Liu L, Zhang H, Jing G, Guo Y, Chen Z, Wang W, Graphics C (2018) Correlation-Preserving Photo Collage. IEEE Transactions on Visualization (6):1956-1968
4. Borji A, Sihite DN, Itti L (2013) What stands out in a scene? A study of human explicit saliency judgment. Vision research 91:62-77
5. Borji A (2015) What is a salient object? A dataset and a baseline model for salient object detection. IEEE Transactions on Image Processing 24 (2):742-756
6. Sang N, Li Z-l, Zhang T-x (2004) Applications of human visual attention mechanisms in object detection. Infrared Laser Engineering 33 (1):38-42
7. Borji A, Itti L (2013) State-of-the-art in visual attention modeling. IEEE transactions on pattern analysis machine intelligence 35 (1):185-207
8. Borji A, Sihite DN, Itti L (2012) Salient object detection: A benchmark. In: Computer Vision–ECCV 2012. Springer, pp 414-429
9. Li X, Lu H, Zhang L, Ruan X, Yang M-H Saliency detection via dense and sparse reconstruction. In: Proceedings of the IEEE international conference on computer vision, 2013. pp 2976-2983
10. Liu Y, Han J, Zhang Q, Wang L (2018) Salient Object Detection Via Two-Stage Graphs. IEEE Transactions on Circuits Systems for Video Technology

11. Wang Q, Zheng W, Piramuthu R Grab: Visual saliency via novel graph model and background priors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. pp 535-543
12. Jian M, Lam K-M, Dong J, Shen L (2015) Visual-patch-attention-aware saliency detection. IEEE transactions on cybernetics 45 (8):1575-1586
13. Treisman AM, Gelade G (1980) A feature-integration theory of attention. Cognitive psychology 12 (1):97-136
14. Wolfe JM, Cave KR, Franzel SL (1989) Guided search: an alternative to the feature integration model for visual search. Journal of Experimental Psychology: Human perception performance 15 (3):419
15. Koch C, Ullman S (1987) Shifts in selective visual attention: towards the underlying neural circuitry. In: Matters of intelligence. Springer, pp 115-141
16. Liu T, Sun J, Zheng N-N, Tang X, Shum H-Y Learning to detect a salient object. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, 2007. IEEE, pp 1-8
17. Zhang J, Sclaroff S Saliency detection: A boolean map approach. In: Proceedings of the IEEE international conference on computer vision, 2013. pp 153-160
18. Zhu W, Liang S, Wei Y, Sun J Saliency optimization from robust background detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014. pp 2814-2821
19. Scharfenberger C, Wong A, Clausi DA (2015) Structure-guided statistical textural distinctiveness for salient region detection in natural images. IEEE Transactions on Image Processing 24 (1):457-470
20. Achanta R, Estrada F, Wils P, Süsstrunk S Salient region detection and segmentation. In: International conference on computer vision systems, 2008. Springer, pp 66-75
21. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proceedings of the IEEE 86 (11):2278-2324
22. Long J, Shelhamer E, Darrell T Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. pp 3431-3440
23. Cheng M-M, Mitra NJ, Huang X, Torr PH, Hu S-M (2015) Global contrast based salient region detection. IEEE transactions on pattern analysis machine intelligence 37 (3):569-582
24. Liu T, Yuan Z, Sun J, Wang J, Zheng N, Tang X, Shum H-YJIToPa, intelligence m (2011) Learning to detect a salient object. 33 (2):353-367
25. Shi J, Yan Q, Xu L, Jia J (2016) Hierarchical image saliency detection on extended CSSD. IEEE transactions on pattern analysis machine intelligence 38 (4):717-729
26. Fan Q, Qi C (2016) Saliency detection based on global and local short-term sparse representation. Neurocomputing 175:81-89
27. Peng H, Li B, Ling H, Hu W, Xiong W, Maybank S (2017) Salient object detection via structured matrix decomposition. IEEE transactions on pattern analysis machine intelligence 39 (4):818-832
28. Parkhurst D, Law K, Niebur E (2002) Modeling the role of salience in the allocation of overt visual attention. Vision research 42 (1):107-123
29. Itti L (2005) Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. Visual Cognition 12 (6):1093-1123
30. Le Meur O, Le Callet P, Barba D, Thoreau D (2006) A coherent computational approach to model the bottom-up visual attention. IEEE transactions on pattern analysis machine intelligence 28:802-817
31. Le Meur O, Le Callet P, Barba D (2007) Predicting visual fixations on video based on low-level visual features. Vision research 47 (19):2483-2498
32. Kootstra G, Nederveen A, De Boer B Paying attention to symmetry. In: British Machine Vision Conference (BMVC2008), 2008. The British Machine Vision Association and Society for Pattern Recognition, pp 1115-1125
33. Reisfeld D, Wolfson H, Yeshurun Y (1995) Context-free attentional operators: the generalized symmetry transform. International Journal of Computer Vision 14 (2):119-130
34. Heidemann G (2004) Focus-of-attention from local color symmetries. IEEE Transactions on Pattern Analysis Machine Intelligence 26 (7):817-830
35. Jian M, Zhang W, Yu H, Cui C, Nie X, Zhang H, Yin Y (2018) Saliency detection based on directional patches extraction and principal local color contrast. Journal of Visual Communication Image Representation 57:1-11
36. Jian M, Zhou Q, Cui C, Nie X, Luo H, Zhao J, Yin Y (2019) Assessment of feature fusion strategies in visual attention mechanism for saliency detection. Pattern Recognition Letters 127:37-47
37. Wang X, Ma H, Chen X, You S (2018) Edge preserving and multi-scale contextual neural network for salient object detection. IEEE Transactions on Image Processing 27 (1):121-134
38. Jian M, Zhao R, Sun X, Luo H, Zhang W, Zhang H, Dong J, Yin Y, Lam K-M (2018) Saliency detection based on background seeds by object proposals and extended random walk. Journal of Visual Communication Image Representation 57:202-211
39. Jian M, Qi Q, Dong J, Sun X, Sun Y, Lam K-M (2018) Saliency detection using quaternionic distance based weber local descriptor and level priors. Multimedia tools applications 77 (11):14343-14360
40. Wang L, Dong S-L, Li H-S, Zhu X-B A brief survey of low-level saliency detection. In: Information

System and Artificial Intelligence (ISAI), 2016 International Conference on, 2016. IEEE, pp 590-593
41. Itti L, Koch C Comparison of feature combination strategies for saliency-based visual attention systems. In: Human vision and electronic imaging IV, 1999. International Society for Optics and Photonics, pp 473-483
42. Ma Y-F, Zhang H-J Contrast-based image attention analysis by using fuzzy growing. In: Proceedings of the eleventh ACM international conference on Multimedia, 2003. ACM, pp 374-381
43. Jiang H, Wang J, Yuan Z, Wu Y, Zheng N, Li S Salient object detection: A discriminative regional feature integration approach. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2013. pp 2083-2090
44. Jiang H, Wang J, Yuan Z, Liu T, Zheng N, Li S Automatic salient object segmentation based on context and shape prior. In: BMVC, 2011. vol 7. p 9
45. Li X, Li Y, Shen C, Dick A, Van Den Hengel A Contextual hypergraph modeling for salient object detection. In: Proceedings of the IEEE international conference on computer vision, 2013. pp 3328-3335
46. Zhou Z, Wang Y, Wu QJ, Yang C-N, Sun X (2017) Effective and efficient global context verification for image copy detection. IEEE Transactions on Information Forensics security 12 (1):48-63
47. Goferman S, Zelnik L L. manor, and A. Tal. Context-aware saliency detection. In: CVPR, 2010. vol 2. p 3
48. Perazzi F, Krähenbühl P, Pritch Y, Hornung A Saliency filters: Contrast based filtering for salient region detection. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012. IEEE, pp 733-740
49. Harel J, Koch C, Perona P Graph-based visual saliency. In: Advances in neural information processing systems, 2007. pp 545-552
50. Zhai Y, Shah M Visual attention detection in video sequences using spatiotemporal cues. In: Proceedings of the 14th ACM international conference on Multimedia, 2006. ACM, pp 815-824
51. Achanta R, Hemami S, Estrada F, Susstrunk S Frequency-tuned salient region detection. In: Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on, 2009. IEEE, pp 1597-1604
52. Goferman S, Zelnik-Manor L, Tal A (2012) Context-aware saliency detection. IEEE transactions on pattern analysis machine intelligence 34 (10):1915-1926
53. Yan Q, Xu L, Shi J, Jia J Hierarchical saliency detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013. pp 1155-1162
54. Shen X, Wu Y A unified approach to salient object detection via low rank matrix recovery. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012. IEEE, pp 853-860
55. Imamoglu N, Lin W, Fang Y (2013) A saliency detection model using low-level features based on wavelet transform. IEEE transactions on multimedia 15 (1):96-105
56. Qi W, Cheng M-M, Borji A, Lu H, Bai L-F (2015) SaliencyRank: Two-stage manifold ranking for salient object detection. Computational Visual Media 1 (4):309-320
57. Xie Y, Lu H, Yang M-H (2013) Bayesian saliency via low and mid level cues. IEEE Transactions on Image Processing 22 (5):1689-1698
58. Jiang B, Zhang L, Lu H, Yang C, Yang M-H Saliency detection via absorbing markov chain. In: Proceedings of the IEEE international conference on computer vision, 2013. pp 1665-1672
59. Li C, Yuan Y, Cai W, Xia Y, Dagan Feng D Robust saliency detection via regularized random walks ranking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015. pp 2710-2717
60. Wei Y, Wen F, Zhu W, Sun J Geodesic saliency using background priors. In: European conference on computer vision, 2012. Springer, pp 29-42
61. Yang C, Zhang L, Lu H, Ruan X, Yang M-H Saliency detection via graph-based manifold ranking. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2013. pp 3166-3173
62. Alexe B, Deselaers T, Ferrari V What is an object? In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, 2010. IEEE, pp 73-80
63. Chang K-Y, Liu T-L, Chen H-T, Lai S-H (2011) Fusing generic objectness and visual saliency for salient object detection.
64. Jiang P, Ling H, Yu J, Peng J Salient region detection by ufo: Uniqueness, focusness and objectness. In: Proceedings of the IEEE international conference on computer vision, 2013. pp 1976-1983
65. Jia Y, Han M Category-independent object-level saliency detection. In: Proceedings of the IEEE international conference on computer vision, 2013. pp 1761-1768
66. Vikram TN, Tscherepanow M, Wrede B (2012) A saliency map based on sampling an image into random rectangular regions of interest. Pattern Recognition 45 (9):3114-3124
67. Li H, Lu H, Lin Z, Shen X, Price B (2015) Inner and inter label propagation: salient object detection in the wild. IEEE Transactions on Image Processing 24 (10):3176-3186
68. Liu R, Cao J, Lin Z, Shan S Adaptive partial differential equation learning for visual saliency detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014. pp 3866-3873

69. Tavakoli HR, Rahtu E, Heikkilä J Fast and efficient saliency detection using sparse sampling and kernel density estimation. In: Scandinavian conference on image analysis, 2011. Springer, pp 666-675
70. Duan L, Wu C, Miao J, Qing L, Fu Y Visual saliency detection by spatially weighted dissimilarity. In: CVPR 2011, 2011. IEEE, pp 473-480
71. Shi K, Wang K, Lu J, Lin L Pisa: Pixelwise image saliency by aggregating complementary appearance contrast measures with spatial priors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013. pp 2115-2122
72. Yang C, Zhang L, Lu H (2013) Graph-regularized saliency detection with convex-hull-based center prior. IEEE Signal Processing Letters 20 (7):637-640
73. Margolin R, Tal A, Zelnik-Manor L What makes a patch distinct? In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013. pp 1139-1146
74. Erdem E, Erdem A (2013) Visual saliency estimation by nonlinearly integrating features using region covariances. Journal of vision 13 (4):11-11
75. Li N, Ye J, Ji Y, Ling H, Yu J Saliency detection on light field. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014. pp 2806-2813
76. Qin Y, Lu H, Xu Y, Wang H Saliency detection via cellular automata. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015. pp 110-119
77. Peng H, Li B, Ji R, Hu W, Xiong W, Lang C Salient Object Detection via Low-Rank and Structured Sparse Matrix Decomposition. In: AAAI, 2013. pp 796-802
78. Zhou L, Yang Z, Yuan Q, Zhou Z, Hu D (2015) Salient region detection via integrating diffusion-based compactness and local contrast. IEEE Transactions on Image Processing 24 (11):3308-3320
79. Judd T, Ehinger K, Durand F, Torralba A Learning to predict where humans look. In: Computer Vision, 2009 IEEE 12th international conference on, 2009. IEEE, pp 2106-2113
80. Yang J, Yang M-H, intelligence m (2017) Top-down visual saliency via joint CRF and dictionary learning. IEEE transactions on pattern analysis 39 (3):576-588
81. Borji A Boosting bottom-up and top-down visual features for saliency estimation. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012. IEEE, pp 438-445
82. Wang Q, Yuan Y, Yan P, Li X (2013) Saliency detection by multiple-instance learning. IEEE transactions on cybernetics 43 (2):660-672
83. Xie S, Tu Z Holistically-nested edge detection. In: Proceedings of the IEEE international conference on computer vision, 2015. pp 1395-1403
84. He S, Lau RW, Liu W, Huang Z, Yang Q (2015) Supercnn: A superpixelwise convolutional neural network for salient object detection. International journal of computer vision 115 (3):330-344
85. Li G, Yu Y Visual saliency based on multiscale deep features. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. pp 5455-5463
86. Wang L, Lu H, Ruan X, Yang M-H Deep networks for saliency detection via local estimation and global search. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015. pp 3183-3192
87. Krähenbühl P, Koltun V Geodesic object proposals. In: European conference on computer vision, 2014. Springer, pp 725-739
88. Lee G, Tai Y-W, Kim J Deep saliency with encoded low level distance map and high level features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. pp 660-668
89. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv
90. Zou W, Komodakis N Harf: Hierarchy-associated rich features for salient object detection. In: Proceedings of the IEEE international conference on computer vision, 2015. pp 406-414
91. Kim J, Pavlovic V A shape-based approach for salient object detection using deep learning. In: European Conference on Computer Vision, 2016. Springer, pp 455-470
92. Uijlings JR, Van De Sande KE, Gevers T, Smeulders AW (2013) Selective search for object recognition. International journal of computer vision 104 (2):154-171
93. Wang X, Ma H, Chen X Salient object detection via fast R-CNN and low-level cues. In: Image Processing (ICIP), 2016 IEEE International Conference on, 2016. IEEE, pp 1042-1046
94. Kim J, Pavlovic V A shape preserving approach for salient object detection using convolutional neural networks. In: Pattern Recognition (ICPR), 2016 23rd International Conference on, 2016. IEEE, pp 609-614
95. Li H, Chen J, Lu H, Chi Z (2017) CNN for saliency detection with low-level feature integration. Neurocomputing 226:212-220
96. Zhao R, Ouyang W, Li H, Wang X Saliency detection by multi-context deep learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015. pp 1265-1274
97. Liu N, Han J Dhsnet: Deep hierarchical saliency network for salient object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. pp 678-686
98. Liang M, Hu X Recurrent convolutional neural network for object recognition. In: Proceedings of the

IEEE Conference on Computer Vision and Pattern Recognition, 2015. pp 3367-3375
99. Li G, Yu Y Deep contrast learning for salient object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. pp 478-487
100. Tang Y, Wu X, Bu W Deeply-Supervised Recurrent Convolutional Neural Network for Saliency Detection. In: Proceedings of the 2016 ACM on Multimedia Conference, 2016. ACM, pp 397-401
101. Lee C-Y, Xie S, Gallagher P, Zhang Z, Tu Z Deeply-supervised nets. In: Artificial Intelligence and Statistics, 2015. pp 562-570
102. Tang Y, Wu X Saliency detection via combining region-level and pixel-level predictions with cnns. In: European Conference on Computer Vision, 2016. Springer, pp 809-825
103. Kruthiventi SS, Gudisa V, Dholakiya JH, Venkatesh Babu R Saliency unified: A deep architecture for simultaneous eye fixation prediction and salient object segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. pp 5781-5790
104. Koltun V Efficient inference in fully connected crfs with gaussian edge potentials. In: NIPS, 2011. Citeseer,
105. Kuen J, Wang Z, Wang G Recurrent attentional networks for saliency detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. pp 3668-3677
106. Jaderberg M, Simonyan K, Zisserman A Spatial transformer networks. In: Advances in neural information processing systems, 2015. pp 2017-2025
107. Zhang J, Dai Y, Porikli F Deep salient object detection by integrating multi-level cues. In: Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on, 2017. IEEE, pp 1-10
108. Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL, intelligence m (2017) Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis 40 (4):834-848
109. Li X, Zhao L, Wei L, Yang M-H, Wu F, Zhuang Y, Ling H, Wang J (2016) Deepsaliency: Multi-task deep neural network model for salient object detection. IEEE Transactions on Image Processing 25 (8):3919-3930
110. Chen T, Lin L, Liu L, Luo X, Li X (2016) DISC: Deep Image Saliency Computing via Progressive Representation Learning. IEEE Trans Neural Netw Learning Syst 27 (6):1135-1149
111. Li G, Xie Y, Lin L, Yu Y Instance-level salient object segmentation. In: Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, 2017. IEEE, pp 247-256
112. Hou Q, Cheng M-M, Hu X, Borji A, Tu Z, Torr P Deeply supervised salient object detection with short connections. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. IEEE, pp 5300-5309
113. Wang T, Borji A, Zhang L, Zhang P, Lu H A stagewise refinement model for detecting salient objects in images. In: Proceedings of the IEEE International Conference on Computer Vision, 2017. pp 4019-4028
114. Luo Z, Mishra AK, Achkar A, Eichel JA, Li S, Jodoin P-M Non-local Deep Features for Salient Object Detection. In: CVPR, 2017. vol 6. p 7
115. Yuan Y, Li C, Kim J, Cai W, Feng DD (2018) Dense and sparse labeling with multidimensional features for saliency detection. IEEE Transactions on Circuits Systems for Video Technology 28 (5):1130-1143
116. Zhang X, Wang T, Qi J, Lu H, Wang G Progressive Attention Guided Recurrent Network for Salient Object Detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018. pp 714-722
117. Li X, Yang F, Cheng H, Liu W, Shen D Contour knowledge transfer for salient object detection. In: Proceedings of the European Conference on Computer Vision (ECCV), 2018. pp 355-370
118. Chen S, Tan X, Wang B, Hu X Reverse attention for salient object detection. In: Proceedings of the European Conference on Computer Vision (ECCV), 2018. pp 234-250
119. Zeng Y, Zhang P, Zhang J, Lin Z, Lu H Towards High-Resolution Salient Object Detection. In: Proceedings of the IEEE International Conference on Computer Vision, 2019. pp 7234-7243
120. Feng M, Lu H, Ding E Attentive Feedback Network for Boundary-Aware Salient Object Detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019. pp 1623-1632
121. Qin X, Zhang Z, Huang C, Gao C, Dehghan M, Jagersand M BASNet: Boundary-Aware Salient Object Detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019. pp 7479-7489
122. Fu K, Zhao Q, Gu I (2018) Refinet: A Deep Segmentation Assisted Refinement Network for Salient Object Detection. IEEE Transactions on Multimedia
123. Liu J-J, Hou Q, Cheng M-M, Feng J, Jiang J (2019) A Simple Pooling-Based Design for Real-Time Salient Object Detection. arXiv preprint arXiv:09569
124. Pont-tuset J Multiscale combinatorial grouping. In: In CVPR, 2014. Citeseer,
125. Krähenbühl P, Koltun V Efficient inference in fully connected crfs with gaussian edge potentials. In: Advances in neural information processing systems, 2011. pp 109-117

126. Zhao H, Shi J, Qi X, Wang X, Jia J Pyramid scene parsing network. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2017. pp 2881-2890
127. Mumford D, Shah J (1989) Optimal approximations by piecewise smooth functions and associated variational problems. Communications on pure applied mathematics 42 (5):577-685
128. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M (2015) Imagenet large scale visual recognition challenge. International Journal of Computer Vision 115 (3):211-252
129. Girshick R, Donahue J, Darrell T, Malik J Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014. pp 580-587
130. Oquab M, Bottou L, Laptev I, Sivic J Learning and transferring mid-level image representations using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014. pp 1717-1724
131. He K, Zhang X, Ren S, Sun J Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. pp 770-778
132. Simonyan K, Zisserman AJapa (2014) Very deep convolutional networks for large-scale image recognition.
133. Cong R, Lei J, Fu H, Cheng M-M, Lin W, Huang Q (2018) Review of Visual Saliency Detection with Comprehensive Information. arXiv preprint arXiv:03391
134. Lang C, Nguyen TV, Katti H, Yadati K, Kankanhalli M, Yan S (2012) Depth matters: Influence of depth cues on visual saliency. In: Computer vision–ECCV 2012. Springer, pp 101-115
135. Ju R, Ge L, Geng W, Ren T, Wu G Depth saliency based on anisotropic center-surround difference. In: Image Processing (ICIP), 2014 IEEE International Conference on, 2014. IEEE, pp 1115-1119
136. Song H, Liu Z, Du H, Sun G, Le Meur O, Ren T (2017) Depth-aware salient object detection and segmentation via multiscale discriminative saliency fusion and bootstrap learning. IEEE Trans Image Processing 26 (9):4204-4216
137. Peng H, Li B, Xiong W, Hu W, Ji R Rgbd salient object detection: a benchmark and algorithms. In: European conference on computer vision, 2014. Springer, pp 92-109
138. Niu Y, Geng Y, Li X, Liu F Leveraging stereopsis for saliency analysis. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012. IEEE, pp 454-461
139. Fang Y, Wang J, Narwaria M, Le Callet P, Lin W (2014) Saliency detection for stereoscopic images. IEEE Trans Image Processing 23 (6):2625-2636
140. Ju R, Liu Y, Ren T, Ge L, Wu G (2015) Depth-aware salient object detection using anisotropic center-surround difference. Signal Processing: Image Communication 38:115-126
141. Guo J, Ren T, Bei J Salient object detection for rgb-d image via saliency evolution. In: Multimedia and Expo (ICME), 2016 IEEE International Conference on, 2016. IEEE, pp 1-6
142. Feng D, Barnes N, You S, McCarthy C Local background enclosure for RGB-D salient object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. pp 2343-2350
143. Wang A, Wang M (2017) RGB-D salient object detection via minimum barrier distance transform and saliency fusion. IEEE Signal Processing Letters 24 (5):663-667
144. Qu L, He S, Zhang J, Tian J, Tang Y, Yang Q (2017) RGBD salient object detection via deep fusion. IEEE Transactions on Image Processing 26 (5):2274-2285
145. Han J, Chen H, Liu N, Yan C, Li X (2017) CNNs-based RGB-D saliency detection via cross-view transfer and multiview fusion. IEEE Transactions on Cybernetics
146. Chen H, Li Y Progressively Complementarity-Aware Fusion Network for RGB-D Salient Object Detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018. pp 3051-3060
147. Wang N, Gong X (2019) Adaptive Fusion for RGB-D Salient Object Detection. IEEE Access 7:55277-55284
148. Liu D, Hu Y, Zhang K, Chen Z Two-Stream Refinement Network for RGB-D Saliency Detection. In: 2019 IEEE International Conference on Image Processing (ICIP), 2019. IEEE, pp 3925-3929
149. Han J, Shao L, Xu D, Shotton J (2013) Enhanced computer vision with microsoft kinect sensor: A review. IEEE transactions on cybernetics 43 (5):1318-1334
150. Gokturk SB, Yalcin H, Bamji C A time-of-flight depth sensor-system description, issues and solutions. In: Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on, 2004. IEEE, pp 35-35
151. "Stereo camera. https://en.wikipedia.org/wiki/Stereo_camera.
152. Rother C, Minka T, Blake A, Kolmogorov V Cosegmentation of image pairs by histogram matching-incorporating a global constraint into mrfs. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, 2006. IEEE, pp 993-1000

153. Fu H, Cao X, Tu Z (2013) Cluster-based co-saliency detection. IEEE Transactions on Image Processing 22 (10):3766-3778
154. Mukherjee L, Singh V, Peng J Scale invariant cosegmentation for image groups. In: Proceedings/CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2011. NIH Public Access, p 1881
155. Kim G, Xing EP, Fei-Fei L, Kanade T Distributed cosegmentation via submodular optimization on anisotropic diffusion. In: 2011 International Conference on Computer Vision, 2011. IEEE, pp 169-176
156. Batra D, Kowdle A, Parikh D, Luo J, Chen T icoseg: Interactive co-segmentation with intelligent scribble guidance. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, 2010. IEEE, pp 3169-3176
157. Chang K-Y, Liu T-L, Lai S-H From co-saliency to co-segmentation: An efficient and fully unsupervised energy minimization model. In: Computer vision and pattern recognition (cvpr), 2011 ieee conference on, 2011. IEEE, pp 2129-2136
158. Tan Z, Wan L, Feng W, Pun C-M Image co-saliency detection by propagating superpixel affinities. In: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, 2013. IEEE, pp 2114-2118
159. Li L, Liu Z, Zou W, Zhang X, Le Meur O Co-saliency detection based on region-level fusion and pixel-level refinement. In: Multimedia and Expo (ICME), 2014 IEEE International Conference on, 2014. IEEE, pp 1-6
160. Ye L, Liu Z, Li J, Zhao W-L, Shen L (2015) Co-saliency detection via co-salient object discovery and recovery. IEEE Signal Processing Letters 22 (11):2073-2077
161. Li Y, Fu K, Liu Z, Yang J (2015) Efficient saliency-model-guided visual co-saliency detection. IEEE Signal Processing Letters 22 (5):588-592
162. Ge C, Fu K, Liu F, Bai L, Yang J (2016) Co-saliency detection via inter and intra saliency propagation. Signal Processing: Image Communication 44:69-83
163. Song H, Liu Z, Xie Y, Wu L, Huang M (2016) RGBD co-saliency detection via bagging-based clustering. IEEE Signal Processing Letters 23 (12):1722-1726
164. Huang R, Feng W, Sun J (2017) Color feature reinforcement for cosaliency detection without single saliency residuals. IEEE Signal Processing Letters 24 (5):569-573
165. Cong R, Lei J, Fu H, Huang Q, Cao X, Ling N (2018) HSCS: Hierarchical Sparsity Based Co-saliency Detection for RGBD Images. IEEE Transactions on Multimedia
166. Li T, Song H, Zhang K, Liu Q, Lian W (2019) Low-rank weighted co-saliency detection via efficient manifold ranking. Multimedia Tools Applications:1-16
167. Zhang D, Han J, Li C, Wang J Co-saliency detection via looking deep and wide. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. pp 2994-3002
168. Zhang D, Meng D, Li C, Jiang L, Zhao Q, Han J A self-paced multiple-instance learning framework for co-saliency detection. In: Proceedings of the IEEE International Conference on Computer Vision, 2015. pp 594-602
169. Wei L, Zhao S, Bourahla OEF, Li X, Wu F (2017) Group-wise deep co-saliency detection. arXiv preprint arXiv:07381
170. Han J, Cheng G, Li Z, Zhang D (2018) A unified metric learning-based framework for co-saliency detection. IEEE Transactions on Circuits Systems for Video Technology 28 (10):2473-2483
171. Zhang K, Li T, Liu B, Liu Q Co-Saliency Detection via Mask-Guided Fully Convolutional Networks With Multi-Scale Label Smoothing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019. pp 3095-3104
172. Fang Y, Lin W, Chen Z, Tsai C-M, Lin C-W (2014) A video saliency detection model in compressed domain. IEEE transactions on circuits systems for video technology 24 (1):27-38
173. Ren Z, Gao S, Rajan D, Chia L-T, Huang Y Spatiotemporal saliency detection via sparse representation. In: 2012 IEEE International Conference on Multimedia and Expo, 2012. IEEE, pp 158-163
174. Liu Z, Zhang X, Luo S, Le Meur O (2014) Superpixel-based spatiotemporal saliency detection IEEE transactions on circuits systems for video technology 24 (9):1522-1540
175. Xi T, Zhao W, Wang H, Lin W (2017) Salient object detection with spatiotemporal background priors for video. IEEE Transactions on Image Processing 26 (7):3425-3436
176. Chen C, Li S, Wang Y, Qin H, Hao A (2017) Video saliency detection via spatial-temporal fusion and low-rank coherency diffusion. IEEE Transactions on Image Processing 26 (7):3156-3170
177. Xue Y, Guo X, Cao X Motion saliency detection using low-rank and sparse decomposition. In: Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, 2012. IEEE, pp 1485-1488
178. Wang W, Shen J, Shao L (2015) Consistent video saliency using local gradient flow optimization and global refinement. IEEE Transactions on Image Processing 24 (11):4185-4196

179. Liu Z, Li J, Ye L, Sun G, Shen L (2017) Saliency detection for unconstrained videos using superpixel-level graph and spatiotemporal propagation. IEEE transactions on circuits systems for video technology 27 (12):2527-2542
180. Guo F, Wang W, Shen J, Shao L, Yang J, Tao D, Tang YYJIToC (2017) Video saliency detection using object proposals.
181. Kim H, Kim Y, Sim J-Y, Kim C-S (2015) Spatiotemporal saliency detection for video sequences based on random walk with restart. IEEE Transactions on Image Processing 24 (8):2552-2564
182. Wang W, Shen J, Porikli F Saliency-aware geodesic video object segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. pp 3395-3402
183. Wang W, Shen J, Yang R, Porikli F (2018) A Unified Spatiotemporal Prior based on Geodesic Distance for Video Object Segmentation. IEEE transactions on pattern analysis machine intelligence 40 (1):20-33
184. Le T-N, Sugimoto A SpatioTemporal utilization of deep features for video saliency detection. In: Multimedia & Expo Workshops (ICMEW), 2017 IEEE International Conference on, 2017. IEEE, pp 465-470
185. Wang W, Shen J, Shao L (2018) Video salient object detection via fully convolutional networks. IEEE Transactions on Image Processing 27 (1):38-49
186. Le T-N, Sugimoto A Deeply supervised 3D recurrent FCN for salient object detection in videos. In, 2017. BMVC,
187. Li J, Xia C, Chen X (2018) A benchmark dataset and saliency-guided stacked autoencoders for video-based salient object detection. IEEE Transactions on Image Processing 27 (1):349-364
188. Cong R, Lei J, Fu H, Porikli F, Huang Q, Hou C (2019) Video Saliency Detection via Sparsity-based Reconstruction and Propagation. IEEE Transactions on Image Processing
189. Guo F, Wang W, Shen J, Shao L, Yang J, Tao D, Tang YYJItoc (2017) Video saliency detection using object proposals. (99):1-12
190. Fang Y, Ding G, Li J, Fang Z (2019) Deep3DSaliency: Deep Stereoscopic Video Saliency Detection Model by 3D Convolutional Networks. IEEE Transactions on Image Processing 28 (5):2305-2318
191. Min K, Corso JJ TASED-Net: Temporally-Aggregating Spatial Encoder-Decoder Network for Video Saliency Detection. In: Proceedings of the IEEE International Conference on Computer Vision, 2019. pp 2394-2403
192. Liu T, Sun J, Zheng N, Tang X, Shum H Learning to detect a salient object, 2007. In. CVPR,
193. Judd-a dataset. http://ilab.usc.edu/borji.
194. Koehler K, Guo F, Zhang S, Eckstein MP (2014) What do saliency models predict? Journal of vision 14 (3):14-14
195. Xu J, Jiang M, Wang S, Kankanhalli MS, Zhao QJJov (2014) Predicting human gaze beyond pixels. 14 (1):28-28
196. MSRA10K. http://mmcheng.net/gsal/.
197. Li Y, Hou X, Koch C, Rehg JM, Yuille AL The secrets of salient object segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014. pp 280-287
198. Xia C, Li J, Chen X, Zheng A, Zhang Y What is and what is not a salient object? learning salient object detector by ensembling linear exemplar regressors. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. IEEE, pp 4399-4407
199. Cheng Y, Fu H, Wei X, Xiao J, Cao X Depth enhanced saliency detection method. In: Proceedings of international conference on internet multimedia computing and service, 2014. ACM, p 23
200. Winn J, Criminisi A, Minka T Object categorization by learned universal visual dictionary. In: Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, 2005. IEEE, pp 1800-1807
201. Li H, Ngan KN (2011) A co-saliency model of image pairs. IEEE Transactions on Image Processing 20 (12):3365-3375
202. Zhang D, Han J, Li C, Wang J, Li X (2016) Detection of co-salient objects by looking deep and wide. International Journal of Computer Vision 120 (2):215-232
203. Li K, Zhang J, Tao W (2016) Unsupervised co-segmentation for indefinite number of common foreground objects. IEEE Transactions on Image Processing 25 (4):1898-1909
204. Fu H, Xu D, Lin S, Liu J Object-based RGBD image co-segmentation with mutex constraint. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015. pp 4428-4436
205. Cong R, Lei J, Fu H, Huang Q, Cao X, Hou C (2018) Co-saliency detection for RGBD images based on multi-constraint feature matching and cross label propagation. IEEE Transactions on Image Processing 27 (2):568-579
206. Tsai D, Flagg M, Nakazawa A, Rehg JM (2012) Motion coherent tracking using multi-label MRF optimization. International journal of computer vision 100 (2):190-202
207. Li F, Kim T, Humayun A, Tsai D, Rehg JM Video segmentation by tracking many figure-ground segments. In: Proceedings of the IEEE International Conference on Computer Vision, 2013. pp 2192-2199
208. Perazzi F, Pont-Tuset J, McWilliams B, Van Gool L, Gross M, Sorkine-Hornung A A benchmark dataset

and evaluation methodology for video object segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. pp 724-732
209. Torralba A, Efros AA Unbiased look at dataset bias. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, 2011. IEEE, pp 1521-1528
210. Chen T, Cheng M-M, Tan P, Shamir A, Hu S-M Sketch2photo: Internet image montage. In: ACM Transactions on Graphics (TOG), 2009. vol 5. ACM, p 124
211. Jian M, Dong J, Ma J (2011) Image retrieval using wavelet-based salient regions. The Imaging Science Journal 59 (4):219-231
212. Ko BC, Nam J-Y (2006) Object-of-interest image segmentation based on human attention and semantic region clustering. JOSA A 23 (10):2462-2470
213. Zhu J-Y, Wu J, Xu Y, Chang E, Tu Z (2015) Unsupervised object class discovery via saliency-guided multiple class learning. IEEE transactions on pattern analysis machine intelligence 37 (4):862-875
214. Frintrop S, García GM, Cremers AB A cognitive approach for object discovery. In: Pattern Recognition (ICPR), 2014 22nd International Conference on, 2014. IEEE, pp 2329-2334
215. Moosmann F, Larlus D, Jurie F Learning saliency maps for object categorization. In: International Workshop on The Representation and Use of Prior Knowledge in Vision (in ECCV'06), 2006.
216. Borji A, Ahmadabadi MN, Araabi BN, Applications (2011) Cost-sensitive learning of top-down modulation for attentional control. Machine Vision 22 (1):61-76
217. Borji A, Itti L Scene classification with a sparse set of salient regions. In: Robotics and Automation (ICRA), 2011 IEEE International Conference on, 2011. IEEE, pp 1902-1908
218. Shen H, Li S, Zhu C, Chang H, Zhang J (2013) Moving object detection in aerial video based on spatiotemporal saliency. Chinese Journal of Aeronautics 26 (5):1211-1217
219. Borji A, Cheng M-M, Jiang H, Li J (2014) Salient object detection: A survey. arXiv preprint. arXiv preprint arXiv: 2 (4)
220. Ma Y-F, Lu L, Zhang H-J, Li M A user attention model for video summarization. In: Proceedings of the tenth ACM international conference on Multimedia, 2002. ACM, pp 533-542
221. Ma Y-F, Zhang H-J A model of motion attention for video skimming. In: Image Processing. 2002. Proceedings. 2002 International Conference on, 2002. IEEE, pp I-I
222. Christopoulos C, Skodras A, Ebrahimi T (2000) The JPEG2000 still image coding system: an overview. IEEE transactions on consumer electronics 46 (4):1103-1127
223. Avidan S, Shamir A Seam carving for content-aware image resizing. In: ACM Transactions on graphics (TOG), 2007. vol 3. ACM, p 10
224. Feng S, Xu D, Yang X (2010) Attention-driven salient edge (s) and region (s) extraction with application to CBIR. Signal Processing 90 (1):1-15
225. Li L, Wu Z, Huang Q, Jiang S, Zha Z (2013) Partial-duplicate image retrieval via saliency-guided visually matching. IEEE Multimedia 99 (1):1
226. Jian M, Yin Y, Dong J, Lam K-M (2018) Content-based image retrieval via a hierarchical-local-feature extraction scheme. Multimedia Tools Applications:1-19
227. Huang H, Zhang L, Zhang H-C Arcimboldo-like collage using internet images. In: ACM transactions on graphics (TOG), 2011. vol 6. ACM, p 155
228. Goldberg C, Chen T, Zhang FL, Shamir A, Hu SM Data-driven object manipulation in images. In: Computer Graphics Forum, 2012. vol 2pt1. Wiley Online Library, pp 265-274
229. Chia AY-S, Zhuo S, Gupta RK, Tai Y-W, Cho S-Y, Tan P, Lin S Semantic colorization with internet images. In: ACM Transactions on Graphics (TOG), 2011. vol 6. ACM, p 156
230. Meger D, Forssén P-E, Lai K, Helmer S, McCann S, Southey T, Baumann M, Little JJ, Lowe DG (2008) Curious george: An attentive semantic robot. Robotics Autonomous Systems 56 (6):503-511
231. Sugano Y, Matsushita Y, Sato Y Calibration-free gaze sensing using saliency maps. In: Computer vision and pattern recognition (cvpr), 2010 ieee conference on, 2010. IEEE, pp 2667-2674
232. Jian M, Qi Q, Dong J, Yin Y, Lam K-M (2018) Integrating QDWD with pattern distinctness and local contrast for underwater saliency detection. Journal of visual communication image representation 53:31-41
233. Jian M, Qi Q, Yu H, Dong J, Cui C, Nie X, Zhang H, Yin Y, Lam K-M (2019) The extended marine underwater environment database and baseline evaluations. Applied Soft Computing 80:425-437
234. Griffin G, Holub A, Perona P (2007) Caltech-256 object category dataset.
235. Rother C, Kolmogorov V, Blake A Grabcut: Interactive foreground extraction using iterated graph cuts. In: ACM transactions on graphics (TOG), 2004. vol 3. ACM, pp 309-314

Inam Ullah is currently pursuing the Ph.D. degree from the School of Software Engineering, Shandong University, Jinan, China. He received his bachelor's degree from the University of Peshawar, Pakistan and his Master's degree from International Islamic University Islamabad, Pakistan. He was a Lecturer in the Department of Computer Science, University of Peshawar, Pakistan from 2016 to 2017. He was a visiting lecturer in the department of Computer science, Islamia college university, Peshawar, Pakistan. His current research interests include image processing, Computer vision, Machine Learning, deep learning, and biometric recognition.

Muwei Jian received a Ph.D. degree from the Department of Electronics and Information Engineering, The Hong Kong Polytechnic University, in October 2014. He was a Lecturer with the Department of Computer Science and Technology, Ocean University of China, from 2015 to 2017. Currently, Dr. Jian is a Professor and Ph.D. Supervisor at the School of Computer Science and Technology, Shandong University of Finance and Economics.

His current research interests include human face recognition, image and video processing, machine learning and computer vision. Prof. Jian was actively involved in professional activities. He has been a member of the Program Committee and Special Session Chair of several international conferences, such as *SNPD 2007, ICIS 2008, APSIPA 2015, EEECS 2016, ICTAI2016, ICGIP 2016 and ICTAI 2017*. Dr. Jian has also served as a reviewer for several international SCI-indexed journals, including *IEEE Trans., Pattern Recognition, Information Sciences, Computers in Industry, Machine Vision and Applications, Machine Learning and Cybernetics, The Imaging Science Journal, and Multimedia Tools and Applications*.

Prof. Jian holds 3 granted national patents and has published over 70 papers in refereed international leading journals/conferences such as *IEEE Trans. on Cybernetics, IEEE Trans. on Circuits and Systems for Video Technology, Pattern Recognition, Information Sciences, Signal Processing, ISCAS, ICME, and ICIP*.

Sumaira Hussain is currently pursuing the Ph.D. degree from the School of Software Engineering, Shandong University, Jinan, China. She received her Bachelor's and Master's degree from the Virtual University of Pakistan. She is a Lecturer in the Department of Computer Science, Sindh Madressatul Islam University, Karachi, Pakistan.

Her current research interests include Image Processing, Deep Learning, Computer vision, Machine Learning, and Artificial Intelligence.

Jie Guo received the M.S. degree from the School of Information Science and Engineering at Shandong Normal University. She is studying for her Ph.D. degree in Shandong University. Her research interests include Machine Learning, Pattern Recognition, and multimedia retrieval.

Hui Yu is a Professor at the University of Portsmouth, UK. His research interests include vision, computer graphics, and application of machine learning and AI to the above areas, particularly in human-machine interaction, image processing, and recognition, Virtual and Augmented reality, 3D reconstruction, robotics and geometric processing of facial performances. He serves as an

Associate Editor of IEEE Transactions on Human-Machine Systems and the Neurocomputing journal.

Xing Wang (M'18) received a bachelor's degree in computer science from Linyi Normal University, and a Ph.D. degree in computer application technology from Northeastern University in 2006 and 2011, respectively. He has been with Liaoning Technical University since 2011, has been an associate professor since 2013, has been a Ph.D. Student Supervisor since 2017, and has been the Vice Dean of School of Electronics and Information Engineering since 2018. His research interests include knowledge graph, image processing, and machine learning.