frontiers | Frontiers in Signal Processing

TYPE Mini Review
PUBLISHED 10 May 2024
DOI 10.3389/frsip.2024.1356793

# Salient object detection: a mini review

Xiuwenxin Wang[1], Siyue Yu[1], Eng Gee Lim[1]* and M. L. Dennis Wong[2,3]*

[1]School of Advanced Technology, Xi'an Jiaotong Liverpool University, Suzhou, Jiangsu, China, [2]School of Engineering, Newcastle University, Newcastle upon Tyne, United Kingdom, [3]Newcastle University Medicine Malaysia, Iskandar Puteri, Malaysia



OPEN ACCESS

EDITED BY
Xiangyuan Lan,
Peng Cheng Laboratory, China

REVIEWED BY
Guangwei Gao,
Nanjing University of Posts and Telecommunications, China

*CORRESPONDENCE
M. L. Dennis Wong,
dennis.wong@newcastle.ac.uk
Eng Gee Lim,
enggee.lim@xjtlu.edu.cn

RECEIVED 16 December 2023
ACCEPTED 22 April 2024
PUBLISHED 10 May 2024

CITATION
Wang X, Yu S, Lim EG and Wong MLD (2024), Salient object detection: a mini review. *Front. Sig. Proc.* 4:1356793. doi: 10.3389/frsip.2024.1356793



This paper presents a mini-review of recent works in Salient Object Detection (SOD). First, We introduce SOD and its application in image processing tasks and applications. Following this, we discuss the conventional methods for SOD and present several recent works in this category. With the start of deep learning AI algorithms, SOD has also benefited from deep learning. Here, we present and discuss Deep learning-based SOD according to its training mechanism, i.e., fully supervised and weakly supervised. For the benefit of the readers, we have also included some standard data sets assembled for SOD research.

KEYWORDS

**computer vision, salient object detection, conventional salient object detection, deep learning, mini review**

## 1 Introduction

Salient Object Detection (SOD) aims to identify the most important regions in an image that capture human attention. These regions typically include objects like cars, dogs, and people. In Figure 1, the input and output images after significant object detection are visually represented. It is designed to mimic human attention to striking areas of the scene. Identifying salient areas in an image can facilitate subsequent advanced visual tasks, enhancing efficiency and resource management and improving performance (Gupta et al., 2020). Thus, SOD can help filter irrelevant backgrounds, and SOD plays a significant pre-processing role in computer vision applications, providing important basic processing for these applications, e.g., segmentation (Donoser et al., 2009; Qin et al., 2014; Noh et al., 2015; Fu et al., 2017; Shelhamer et al., 2017), classification (Borji and Itti, 2011; Joseph et al., 2019; Akila et al., 2021; Liu et al., 2021; Jia et al., 2022; Ma and Yang, 2023), tracking (Frintrop and Kessel, 2009; Su et al., 2014; Ma et al., 2017; Lee and Kim, 2018; Chen et al., 2019), etc.

Existing SOD approaches can be roughly divided into two classes: 1) conventional approaches; and 2) deep-learning-based approaches, as shown in Figure 2. Conventional approaches exploit low-level features and some heuristics to detect salient objects, containing local contrast-based, diffusion-based, Bayesian approach, objectness prior, and classical supervision. In addition, deep learning-based approaches can help extract comprehensive deep semantic features to improve performance. They can be further sub-categorised into fully supervised learning (Wang et al., 2015a; Lee et al., 2016a; Kim and Pavlovic, 2016; He et al., 2017a; Hou et al., 2017; Shelhamer et al., 2017; Su et al., 2019) and weakly supervised learning (Zhao et al., 2015a; Lee et al., 2016b; Zhang et al., 2018; Shen et al., 2018; Tang et al., 2018; Zhang et al., 2020a; Yu et al., 2021) based on the given labels. This paper will summarise and discuss several chosen methods according to the two

FIGURE 1
The input images and outputs of salient object detection (Cheng et al., 2013; Borji et al., 2014; Borji et al., 2015; Cheng et al., 2015).

categories. Beyond concluding, the paper will briefly present recent datasets commonly used for SOD for the interest of the readers.

## 2 Conventional methods

Since the Itti attention architecture has been put forward (Itti et al., 1998). in 1998, this research on visual salient object detection drew a high interest. Inspired by the human eye mechanism (Itti et al., 1998) and the proposed salient object features (e.g., sharp colours, strong contrasts, directional differences). Over the past 20 years, many salient object detection methods have been proposed. Most of the proposed works first identify significant subsets from the image by computing the significance graph and then combining them to form the segment of the substantial object. Depending on the chosen priors, one can further categorise the conventional methods into local contrast-based, diffusion-based, Bayesian approach-based, objectness prior-based, and classical supervised SOD methods.

### 2.1 Local contrast based SOD

In the early works on salient object detection, the estimation of element uniqueness was typically followed by pixel-level centre positioning. This process involved utilizing one or more low-level features to determine the orientation, colour, and contrast of image elements relative to the surrounding environment. The concept of local contrast, which measures the difference between a pixel and its neighbouring pixels, played a crucial role in these approaches. Local contrast is often calculated as a Gaussian-weighted sum of these differences. For example, Ma and Yang. (2003) calculated local contrast by considering the differences between a pixel and its local surrounding pixels. They employed a Gaussian weighting scheme to emphasize the contributions of nearby pixels in the contrast calculation. Similarly, in the work by Liu et al. (2007), local contrast was extended to image patches. The authors utilized local contrast as a feature for learning tasks related to image analysis and processing. In the study by Ma and Zhang (2003), they proposed color-quantized CIELuv images subdivided into pixels blocks, and differences were computed using local contrast. The fuzzy growth approach was then utilized to segment points of interest and regions to generate the saliency map. Rosin (2009) proposed a parameter-free method that employed simple point-wise operations, including edge detection, threshold decomposition, and distance transformation. Furthermore, Hu et al. (2005) introduced a linear subspace estimation approach that mapped a two-dimensional image to a one-dimensional linear subspace following a polar coordinate transformation. By projecting all data onto their corresponding subspace normal, this approach considered both feature contrast and geometric properties of the region. To enhance robustness, Liu et al. (2007) employed pyramids to adjust the contrast at a single scale and extended it to operate on multiple scales. Using linear combinations, they computed multi-scale contrast features at image pixels by combining contrasts from an L-layer Gaussian pyramid. In another work by Liu and Gleicher (2006), block/pixel-based multi-scale contrast features were integrated with regional information for object localization. However, one limitation of this method was its heavy reliance on image segmentation quality, which could impact its performance. Additionally, it was observed that pixel-based multi-contrast saliency maps tended to emphasize high-contrast edges more than the overall salient objects, which could be considered a drawback of the approach.

Additionally, Jiang et al. (2011) referenced as (Jiang et al., 2011) utilized an image segmentation algorithm to generate multi-scale segmentations, enabling the achievement of multi-scale local contrast in their work. The saliency of a region at a specific scale was determined by comparing its regional characteristics with those of its neighbouring areas. The regional saliency values were

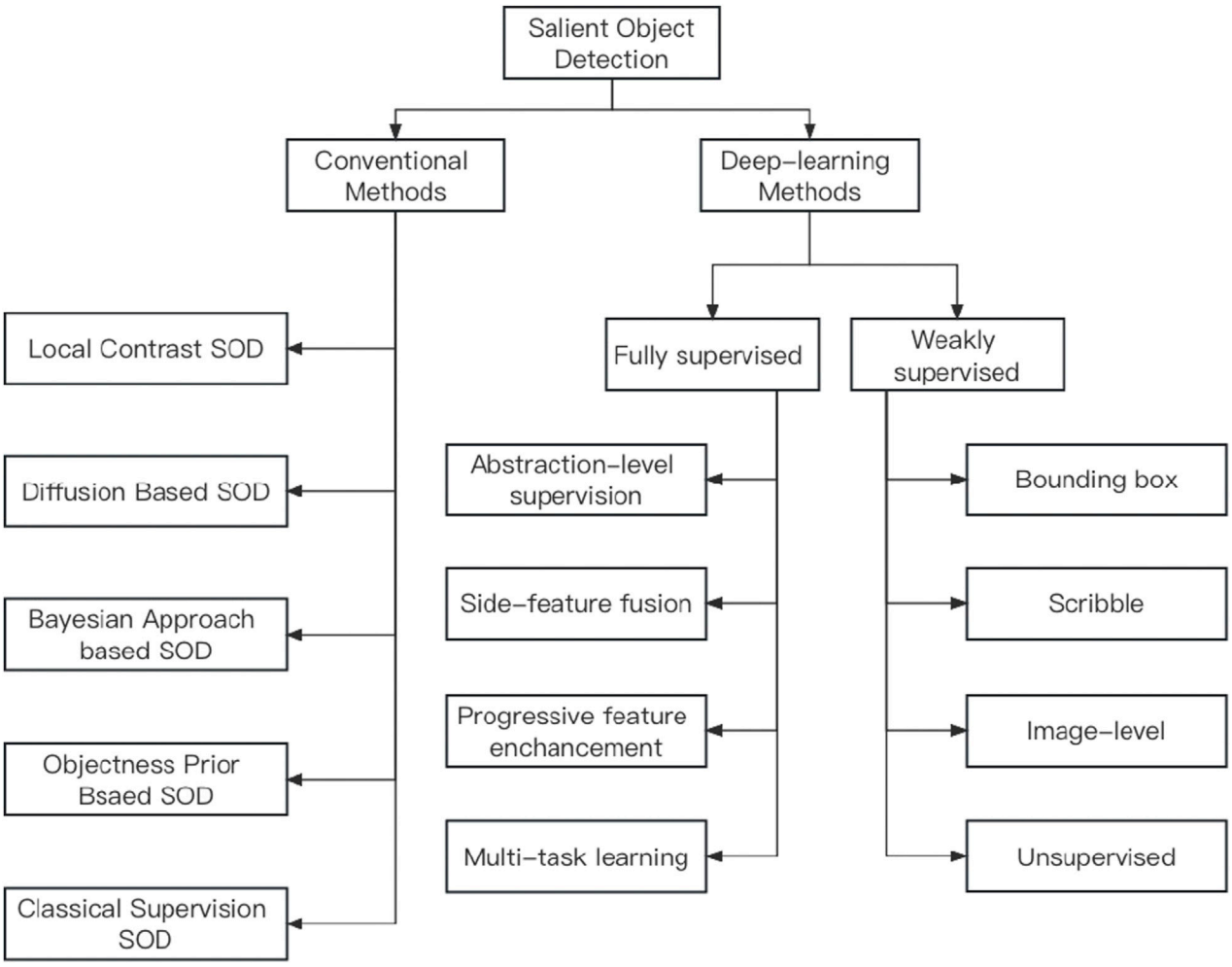


FIGURE 2
Existing salient object detection methods.

propagated to individual pixels across different scales to obtain a pixel-wise saliency map. Klein and Frintrop (2011), as referenced in (Klein and Frintrop, 2011), defined the saliency of image regions using Kullback-Leibler divergence (KLD). They designed specific scalable feature detectors to represent the distribution of feature channels, such as intensity, colour, and orientation. KLD was employed to quantify the difference between the central and surrounding feature statistics, thereby estimating the centre-surround contrast. Li et al. (2013), as referenced in (Li et al., 2013), conducted local contrast analysis to identify salient regions through imbalanced maximum edge learning. The regional context of a central rectangular patch encompassed all spatially overlapping surrounding patches. They utilized a trained cost-sensitive support vector machine (SVM) to obtain the inter-class separability between the central positive patch and all the surrounding negative patches.

## 2.2 Diffusion based SOD

The diffusion-based saliency object detection (SOD) models utilize a graph structure on the image and employ a diffusion matrix to propagate saliency values across the entire area. In a patch-based approach, Gopalakrishnan et al. (2010) leverage the equilibrium distribution of ergodic Markov chains on complete and k-regular graphs. This generates salient and background seeds as partial labels on the "pop-up" plot. Semi-supervised learning techniques are then used to infer labels for unlabeled nodes.

In the work by Yang et al. (2013), manifold ranking is incorporated as a saliency measure in a two-stage scheme. Regional saliency maps are computed in the first stage, which reflect the correlation of different sides of constituent superpixels with the pseudo-background. In the second stage, foreground nodes obtained from adaptive thresholding of the inverse initial saliency map are used as saliency queries. Manifold ranking is applied again to compute the final saliency score for each superpixel.

Zhang et al. (2017a) propose a method to effectively suppress distant background regions near the image centre using a transition probability matrix. Multiple sparse affinities with different feature layers from a pre-trained FCN network are computed, and the complete affinity matrix is inferred through iterative optimization.

Filali et al. (2016) extend the single-layer manifold ranking framework to a multi-layer saliency maps framework,

incorporating texture cues and colours to accurately detect the boundaries of salient objects. Sun et al. (2015) identify salient regions in the image by computing the Markov absorption probability, which represents the probability of a transient node being absorbed by an absorbing node. Ranking-based refinement is performed using adaptively thresholded salient nodes in the saliency map.

Furthermore, Jiang et al. (2019) propose a super-diffusion framework that integrates various diffusion matrices, salient features, and seed vectors to achieve robust and optimal performance.

## 2.3 Bayesian approach based SOD

The Bayesian approach for saliency object detection (SOD) involves estimating the probability that each pixel in an image is significant, given the input image. Xie et al. (2012) proposed a method that estimated the convex hull based on interest points. This convex hull was crucial for determining saliency priors and likelihood functions. They computed pixel-specific saliency priors over the intersection of surrounding clusters and the convex hull. Clustering techniques were employed to group superpixels into more significant regions, generating bounding clusters.

In a similar vein, Sun et al. (2012) developed a method where they computed a prior map resembling Xie et al.'s approach (Xie et al., 2012). However, they introduced a weighting scheme for the convex hull at the boundaries of superpixels, utilizing the probability of the boundary and the colour difference between the superpixel and the background region.

To enhance likelihood estimation, Wang et al. (2016a) proposed a geodesically weighted Bayesian framework, incorporating fully connected conditional random fields (CRFs). This framework, referenced as (Wang et al., 2016a), inferred more accurate initial saliencies through CRFs and used saliency maps from existing methods as prior distributions.

## 2.4 Objectness prior based SOD

Objective-based saliency object detection (SOD) methods leverage object proposal algorithms to identify potential object-containing image windows (Alexe et al., 2012). Chang et al. (2011) proposed a method (Chang et al., 2011) that jointly estimates latent object windows' objectness and regional saliency by iteratively minimizing an energy function.

Jia et al. (2013) developed a technique (Jia and Han, 2013) that utilizes object scores as saliency measures and gives greater weight to foreground pixels compared to background pixels when propagating saliency information using Gaussian Markov Random Fields (MRF).

Li et al. (2015) combined objective foreground labels with boundary cues in a co-transduction framework to generate improved saliency maps for complex images (Li et al., 2015). Additionally, Jiang et al. (2013) explored focusability and objectivity priors, integrating them nonlinearly with uniqueness cues at the pixel level to enhance SOD performance (Jiang et al., 2013). Regional objectivity scores were computed by averaging the objectivity scores of constituent pixels.

## 2.5 Classical supervised SOD

Supervised models based on classical machine learning (ML) algorithms have been proposed for saliency object detection (SOD). These approaches typically involve several steps, First, complex features, such as superpixels or blocks, are manually extracted from each image region. These features capture information like colour, location, size, and texture, which are used to create region descriptors. Next, the extracted features are used to generate region descriptors that represent the characteristics of each image region. Then, a trained ML regressor or classifier, which can be linear or nonlinear, is applied to predict saliency scores or confidence levels based on the input region descriptors. This step involves mapping the region descriptors to saliency scores. Finally, the saliency score of each region is assigned to the pixels it contains, resulting in an initial saliency map. This map highlights the salient areas of the image based on the predicted scores.

To provide a cohesive overview of different studies in this field, here are the descriptions of the approaches, organized more coherently. Mehrani and Veksler (2010) incorporated standard features such as colour, location, size, and texture to form region descriptors. They employed a trained boosted decision tree classifier and further refined the initial segmentation through binary graph cut optimization to achieve accurate boundaries.

Kim et al. (2014) employed a high-dimensional colour transformation to represent the saliency map as a linear combination of high-dimensional colour spaces. They estimated the initial saliency map using a random forest regressor proposed by Breiman (2001).

Wang et al. (2013) formulated SOD as a multiple instance learning (MIL) problem. They independently trained four MIL classifiers using regional feature descriptors, including low-level, mid-level, and boundary cues.

Jiang et al. (2013) utilized region descriptors and trained a random forest regressor to map region feature vectors to region saliency scores.

Yang and Yang (2017) developed a maximum margin method to jointly learn conditional random field (CRF) and discriminative dictionaries for SOD. Their hierarchical CRF model conditioned the target variable on an intermediate layer of sparse codes of image patches.

These studies demonstrate different approaches within the framework of supervised ML algorithms for SOD, showcasing variations in feature extraction, ML algorithms, and additional refinement techniques employed.

# 3 Deep-learning based SOD

While hand-crafted feature-based methods can achieve real-time salient object detection (SOD), they have limitations when it comes to complex scenes. However, the emergence of convolutional neural networks (CNNs) has provided new insights for SOD researchers, leveraging the multi-level and multi-scale features of CNNs to accurately capture salient regions without relying on prior knowledge. CNNs can effectively localize salient object boundaries, even with challenging factors like shadows or reflections. As a result, CNN-based SOD methods have outperformed conventional

approaches on various datasets and are now the preferred choice for SOD. Deep learning-based SOD models utilize the hierarchical nature of CNNs and introduce novel network architectures to generate representations that enable saliency detection. These models leverage the multi-layer features to automatically identify highly salient regions at coarse scales while utilizing shallower layers to capture detailed information about boundaries and delicate structures, aiding in localizing salient objects. The versatility of CNNs makes them convenient tools for designing and researching novel SOD models (Gupta et al., 2020). In the following subsections, systematically review deep learning-based SOD models. These models can be broadly categorized into fully supervised models, weakly supervised models, and adversarial models, depending on their level of supervision.

## 3.1 Fully supervised models

Fully supervised salient object detection (SOD) models are typically developed with the assumption that a sufficient amount of human-annotated training data, consisting of salient object masks, is available. Zhang et al. (2016) explored the use of multi-context deep features for SOD with abstraction-level supervision. They employed two structurally similar convolutional neural networks (CNNs) to model each image superpixel's global and local context independently. Each CNN took as input a fixed-scale window centred on the superpixel of interest. The outputs of the two CNNs were then combined and fed through a shared multi-layer perceptron (MLP) for regression, yielding the final saliency score.

Zhao et al. (2015b) investigated the use of multi-context deep features for SOD with abstraction-level supervision. They employed two convolutional neural networks (CNNs) to model each image superpixel's global and local context independently. Each CNN processed a fixed-scale window centred on the superpixel, defining the context range. The extracted multi-context features from the superpixels were combined and regressed using a shared multi-layer perceptron (MLP) to generate the final saliency score.

Lee et al. (2016a) developed feature descriptors for each superpixel by integrating encoded low-level distance maps (ELD-Map) with deep CNN features that exhibited more robust semantic representations. ELD-Map encoded the similarities and differences between the queried superpixel and other superpixels. A stack of hand-crafted feature distance maps captured these relationships, which were then processed using a simple CNN to generate the ELD maps.

Wang et al. (2015b) combined pixel-wise local estimation with object-aware global search to achieve robust saliency detection. They trained a deep CNN with patch inputs to assign saliency values to each pixel in an image. Candidate object regions were represented by vectors that combined global contrast, geometric information, and local saliency measurement features. These vectors were then processed using an MLP to obtain the final saliency score.

Zhang et al. (2016) proposed a maximum *a posteriori* (MAP)-based subset optimization approach to filter a set of scored bounding box proposals into a compact subset of detections. They utilized a CNN model to generate a fixed number of scoring location recommendations for an optimizer based on MAP.

Kim and Pavlovic (2016) employed a CNN as a multi-label classifier to estimate the proximity of region proposals for each predefined shape class using a fixed binary representation. The final saliency of an image pixel was computed by averaging the predictions of all region proposals containing that pixel.

Su et al. (2019) proposed a SOD framework that addressed the selectivity-invariant dilemma by incorporating three streams. The first stream employed an integrated continuous dilation module to achieve feature invariance within objects. The second stream used hierarchical multi-scale features and boundary ground-truth supervision for accurate salient edge localization. The third stream modelled the challenging transition zone between object boundaries and their interiors.

Zhao et al. (2019a) proposed a two-stage fusion scheme to exploit the complementarity between saliency and edges. In the first stage, a U-Net architecture with different kernel size convolutions and nonlinearity at the decoder was utilized to extract multi-layer saliency features at multiple scales. In the second stage, these layer-wise saliency features were fused with image-level edge features to generate a side-output feature set. The saliency maps obtained from this set were merged to get the final map.

To enhance the learning of semantic knowledge for SOD, Zeng et al. (2019) introduced a joint learning approach for weakly supervised semantic segmentation and SOD. Their architecture consisted of two subnetworks operating on shared backbone features. The first sub-network was trained to produce semantic segmentation using image-level supervision in the first stage. The obtained semantic segmentations were then used as pseudo-labels to supervise the second-stage training of semantic segmentation. The saliency aggregation subnetwork computed a weighted sum of segmentation masks for all classes, guided by saliency ground truth labels to generate a saliency map.

He et al. (2017a) employed subitizing as an auxiliary task to improve SOD performance. They connected a pre-trained subitizing subnet to the SOD subnet using an adaptive weight layer. The SOD subnet was based on the U-Net architecture with skip connections and hierarchical supervision. An adaptive weighting layer was inserted between the two-halves of the U-Net, with weights dynamically determined by the subnets. Both subnets were fine-tuned together in an end-to-end manner during network training.

Furthermore, Islam et al. (2018) introduced a skip connection strategy in their model for SOD. This strategy facilitated top-down progressive refinement of the coarsest feature map generated by the encoder. The refinement process was supervised by ground-truth masks designed for the subitization task and pixel-wise saliency annotations. High-level auxiliary features of one layer were obtained by gating its features with the coarse hierarchy features. Finally, a fusion layer combined multi-scale saliency predictions to generate the final saliency map.

## 3.2 Weakly supervised models

Fully supervised salient object detection (SOD) models rely on human-annotated training data, which is both labour-intensive and time-consuming. To mitigate these challenges, researchers have explored alternative approaches using weak supervision and generative adversarial networks (GANs) for SOD.

Weakly supervised SOD models utilize sparse annotations such as bounding boxes, scribbles, or image-level labels instead of pixel-level annotations. Wang et al. (2017a) proposed a weakly supervised SOD model that primarily uses image-level labels for supervision. They jointly trained classification and foreground feature inference networks (FIN) under image-level supervision. The FIN captures salient regions independent of specific object categories. In the second stage, a SOD subnet merges the FIN graph with deeper side features from the backbone network to generate initial saliency predictions. These predictions are refined using iterative conditional random fields (CRFs) for self-training of the SOD branch.

Another weakly supervised approach involves utilizing scribble annotations. Zhang et al. (2020b) developed a SOD model that incorporates scribble annotations. They employed edge detection subnets alongside SOD flows to address the limitations of boundary localization in SOD caused by the lack of fine details and structure in scribble annotations.

GANs have also been applied to SOD to enhance saliency boundaries and generate realistic saliency maps. Cai et al. (2020) introduced a dynamic matching module in the GAN framework to improve the accuracy of salient object boundaries. Their model integrates low-level colour and texture features using superpixel-based methods to refine regional saliency scores.

Tang and Wu (2019) proposed a cascaded CNN-based generator for SOD that implicitly enhances saliency boundaries through adversarial learning. They adopted the conditional GAN strategy, incorporating adversarial losses to enforce clear boundaries and spatial consistency. Local image patches were leveraged to capture the regional structure of salient regions.

Furthermore, Sym et al. (Zhu et al., 2018) incorporated a correlation layer in the discriminator of their GAN-based SOD model. This correlation layer enables local patch-based comparisons between synthetic saliency maps and corresponding saliency masks, enhancing the model's ability to capture salient object boundaries.

In contrast to other SOD models, Liu et al. (2019a) specifically tackled the issue of feature dilution through a progressive refinement approach. They recognized that in some SOD models, features can become diluted as they pass through multiple layers, leading to a degradation in performance. To overcome this, Liu et al. proposed a progressive refinement strategy that iteratively refines the saliency predictions, allowing the model to focus on more informative features and improve the overall accuracy.

Context extraction models, such as the one proposed by Liu et al. (2018a), have also made significant contributions to the field of SOD. These models employ extensive computational operations to capture contextual information and have achieved state-of-the-art results. Considering the context surrounding salient objects, these models can refine the saliency predictions and enhance the detection accuracy.

Furthermore, leveraging additional SOD learning information that is related to the task can be beneficial in a multi-task learning environment. For example, Zhao et al. (2019b) and He et al. (2017b) explored the use of additional information in SOD tasks, which can aid in improving the performance of both functions. By jointly learning related tasks, the model can leverage shared knowledge and enhance the performance of individual tasks.

These weakly supervised and GAN-based approaches offer alternatives to fully supervised SOD models, reducing the reliance on pixel-level annotations and addressing the labour-intensive and time-consuming nature of the annotation process.

## 4 Datasets

In this section, let us delve into some of the commonly used datasets in Salient Object Detection (SOD). These datasets have played a crucial role in advancing the field and evaluating the performance of different SOD algorithms.

One of the widely used datasets is the MSRA Dataset. Created by Liu et al. (2007), it is divided into two parts: MSRA-A and MSRA-B. This dataset provides a large-scale collection of images with salient object annotations in bounding boxes. Researchers often rely on this dataset to evaluate and benchmark SOD algorithms.

Another important dataset is BSD-SOD, derived from the Berkeley Segmentation Dataset (BSD) (Movahedi and Elder, 2010). BSD-SOD consists of 300 images with pixel-level annotations for salient objects. The dataset poses challenges due to low contrast between objects and the background and objects touching the image boundaries.

The Complex Scene Saliency Dataset (CSSD) and its extended version (ECSSD) (Yan et al., 2013a) are also widely used in SOD research. CSSD contains 200 images, while ECSSD consists of 1,000 images. These datasets focus on scenes that are both semantically meaningful and structurally complex, providing diverse challenges for evaluating SOD algorithms.

The PASCAL-S Dataset, proposed by Li et al. (2014), comprises 850 complex scene images extracted from the PASCAL VOC dataset (Everingham et al., 2010). This dataset offers a diverse range of scenes with salient objects, making it suitable for evaluating and comparing different SOD algorithms.

Lastly, the DUTS Dataset, introduced by Wang et al. (2017a), has gained popularity in recent years. It includes a training set with 10,553 images and a test set with 5,019 images. The DUTS dataset serves as a benchmark for evaluating SOD models and has significantly contributed to advancing the field.

These datasets serve as valuable resources for training, evaluating, and comparing SOD algorithms, allowing researchers to assess the performance and generalization capabilities of different approaches.

## 5 Comparison and analysis

Runtime Performance: To assess the runtime of various saliency detection models, we considered representative methods from different categories: traditional models [e.g., Significant Filter (SF) (Perazzi et al., 2012)], Manifold Ranking (MR) (Yang et al., 2013), Robust Background Detection (RBD) (Zhu et al., 2014), classical machine learning-based models (e.g., Discriminative Region Feature Integration (DRFI) (Jiang et al., 2013)], and deep learning-based

TABLE 1 Average running time of several salient object detection (SOD)models.

| Models | SF (82) | MR (45) | RBD (83) | DRFI (57) | MCDL (64) |
|---|---|---|---|---|---|
| Time(*s*) | 0.16 | 0.25 | 0.25 | 9 | 2.41 |
| GPU Support | NO | NO | NO | NO | Yes |
| Learning | NO | NO | NO | CML | DL |
| Code | C++ | Matlab | Matlab | Matlab | Caffe |
| **Models** | **PiCANet (87)** | **RAS (88)** | **PoolNet (89)** | **AFNet (90)** | **EGNet (66)** |
| Time(*s*) | 0.19 | 0.0291 | 0.033 | 0.023 | 0.11 |
| GPU Support | Yes | Yes | Yes | Yes | Yes |
| Learning | DL | DL | DL | DL | DL |
| Code | Caffe | Caffe | Caffe | pytorch | pytorch |
| **Models** | **AMULet (84)** | **UCF (86)** | **C2S-Net (93)** | **CPD (85)** | **BASNet (91)** |
| Time(*s*) | 0.07 | 0.046 | 0.034 | 0.016 | 0.014 |
| GPU Support | Yes | Yes | Yes | Yes | Yes |
| Learning | DL | DL | DL | DL | DL |
| Code | Caffe | Caffe | Caffe | pytorch | pytorch |

models [e.g., Abstraction-level Supervision (MCDL) (Zhao et al., 2015b), Side Feature Fusion, Context Extraction].

Traditional SOD models, such as Significant Filter (Perazzi et al., 2012), rely on low-level cues, while models like Manifold Ranking (Yang et al., 2013) and Robust Background Detection (Zhu et al., 2014) utilize background priors in different ways. Classical machine learning-based model Discriminative Region Feature Integration (Jiang et al., 2013) integrates heuristic region descriptors using classical machine learning-based techniques.

Deep learning-based SOD models fall into various subcategories. For example, Abstraction-level Supervision models focus on improving predictive performance, while Side Feature Fusion models [e.g., AMULet (Zhang et al., 2017b), EGNet (Zhao et al., 2019a), CPD (Wu et al., 2019a)] incorporate context extraction strategies to capture high-level contextual information. Simple encoder-decoder enhancement models [e.g., UCF (Zhang et al., 2017c)] aim to improve efficiency, and context extraction models [e.g., PiCANet (Liu et al., 2018b)] utilize LSTM or multi-scale convolutional kernels. Progressive feature refinement models [e.g., RAS (Chen et al., 2018), PoolNet (Liu et al., 2019b), AFNet (Feng et al., 2019), BASNet (Qin et al., 2019)] refine features at different scales, and multi-task models [e.g., SCRN (Wu et al., 2019b)] address multiple saliency-related tasks. Weakly supervised SOD models [e.g., C2S-Net (Li et al., 2018)] are also considered for evaluation.

Runtime evaluations were conducted on a workstation with an Intel Xeon(R) Bronze 3104 CPU@1.70$GHz$ × 12 and Nvidia Quadro-P5000 GPU (with 17 GB RAM). As shown in Table 1, traditional SOD models, without any accelerators, exhibited long runtimes. However, despite their popularity, these models, which rely on low-level features and saliency priors, struggle to capture the high-level contextual information necessary for accurate saliency detection. As a result, their performance on saliency metrics (e.g., MAE: above 0.163 and max $F_\beta$: below 0.685) is relatively low, and they generate subpar saliency maps for complex scenes. While some deep learning-based models explicitly address the runtime issue [e.g., RAS (Chen et al., 2018), CPD (Wu et al., 2019a)] and demonstrate high performance, others [e.g., MDCL (Zhao et al., 2015b), PiCANet (Liu et al., 2018b), EGNet (Zhao et al., 2019a)] have longer inference times due to their context extraction strategies. Models that reduce channel dimensions or discard high-resolution information [e.g., RAS (Chen et al., 2018), SCRN (Wu et al., 2019b), BASNet (Qin et al., 2019)] achieve a balance between efficiency and predictive performance. Improving model efficiency involves introducing novel techniques [e.g., RAS (Chen et al., 2018), AFNet (Feng et al., 2019)] to reduce computational complexity or discard certain information while maintaining satisfactory predictive performance.

Recent deep-learning-based saliency object detection (SOD) models, including MINet (Pang et al., 2020), SACNet (Hu et al., 2020), GateNet (Zhao et al., 2020), $U^2 - Net$ (Qin et al., 2020), LDF (Wei et al., 2020), DSRNet (Wang et al., 2020), EGNet (Zhao et al., 2019a), PoolNet-Edge (Liu et al., 2019b), AFNet (Feng et al., 2019), MLMS (Wu et al., 2019), PAGE (Wang et al., 2019), CPD (Wu et al., 2019a), BDPM (Zhang et al., 2018), JDF (Xu et al., 2019), RAS (Chen et al., 2018), PAGR (Zhang et al., 2018), C2S-Net (Li et al., 2018), PiCANet (Liu et al., 2018b), DSS (Hou et al., 2017), UCF (Zhang et al., 2017c), MSRNet (Li et al., 2017), ILS (Wang et al., 2017b), NLDF (Luo et al., 2017), AMULet (Zhang et al., 2017b), SCRN (Wu et al., 2019b), BANet (Qin et al., 2019), BASNet (Qin et al., 2019), CapSal (Zhang et al., 2019), DGRL (Wang et al., 2018), SRM (Wang et al., 2017), have been quantitatively evaluated using four evaluation metrics across SOD dataset ECSSD (Yan et al., 2013b). The evaluation metrics used are maximum F-measure (Achanta et al., 2009), S-measure (Fan et al., 2017), E-measure (Fan et al., 2018), and

TABLE 2 Quantitative Performance of recent state-of-the-art deep learning-based SOD methods on one popular dataset. Performance metrics of maximum F-measure, S-measure, E-measure and Mean Absolute Error (MAE) is represented by $maxF_\beta$, $S_m$, $E_m$, and MAE, respectively. Superscript in the first column: *"X"*, *"S"*, *"D"* ResNeXt-101, ResNeXt-101 and DenseNet backbone. ↑ and ↓indicate that the larger and smaller scores are better respectively.

| VGG | | | | |
|---|---|---|---|---|
| Model | $maxF_\beta$ ↑ | $S_m$ ↑ | $E_m$ ↑ | MAE ↓ |
| ILS (106) | 0.855 | 0.811 | 0.868 | 0.103 |
| MSRNet (105) | 0.911 | 0.895 | 0.918 | 0.054 |
| NLDF (107) | 0.905 | 0.875 | 0.912 | 0.063 |
| Amulet (84) | 0.915 | 0.894 | 0.912 | 0.059 |
| UCF (86) | 0.903 | 0.884 | 0.896 | 0.069 |
| DSS (18) | 0.899 | 0.873 | 0.907 | 0.068 |
| PiCANet (87) | 0.931 | 0.914 | 0.926 | 0.046 |
| RAS (88) | 0.921 | 0.893 | 0.922 | 0.056 |
| C2S-Net (93) | 0.910 | 0.893 | 0.914 | 0.054 |
| PAGR (104) | 0.927 | 0.889 | 0.917 | 0.061 |
| JDF (103) | 0.927 | 0.906 | 0.931 | 0.049 |
| BDMP (102) | 0.929 | 0.910 | 0.915 | 0.044 |
| CPD (85) | 0.936 | 0.910 | 0.943 | 0.040 |
| MLMS (100) | 0.928 | 0.911 | 0.916 | 0.045 |
| PAGE (101) | 0.931 | 0.912 | 0.943 | 0.042 |
| AFNet (90) | 0.935 | 0.912 | 0.940 | 0.042 |
| PoolNetEdge (89) | 0.941 | 0.917 | 0.942 | 0.041 |
| EGNet (66) | 0.942 | 0.918 | 0.941 | 0.041 |
| MINet (94) | 0.943 | 0.919 | 0.947 | 0.036 |
| ResNet-50/ResNet-101/DenseNet/ResNeXt-101/RSU | | | | |
| SRM (110) | 0.917 | 0.895 | 0.928 | 0.054 |
| DGRL (109) | 0.925 | 0.906 | 0.943 | 0.043 |
| BASNet (91) | 0.942 | 0.916 | 0.921 | 0.037 |
| *CapSal*$^{s}$ (108) | 0.862 | 0.826 | 0.866 | 0.074 |
| PoolNetEdge (89) | 0.949 | 0.926 | 0.948 | 0.035 |
| BANet (22) | 0.945 | 0.924 | 0.953 | 0.035 |
| SCRN (92) | 0.950 | 0.927 | 0.942 | 0.037 |
| *DSRNet*$^{D}$ (99) | 0.950 | 0.922 | 0.953 | 0.031 |
| LDF (98) | 0.950 | 0.923 | 0.950 | 0.034 |
| $U^2Net^{RSU}$ (97) | 0.951 | 0.928 | 0.925 | 0.032 |
| MINet (94) | 0.947 | 0.925 | 0.953 | 0.033 |
| *GateNet*$^{X}$ (96) | 0.952 | 0.929 | — | 0.035 |
| *SACNet*$^{S}$ (95) | 0.954 | 0.930 | 0.958 | 0.028 |

mean average error [MAE (Perazzi et al., 2012)]. Based on the evaluation results shown in Table 2, the more recent models such as SACNet (Hu et al., 2020), MINet (Pang et al., 2020), GateNet (Zhao et al., 2020), and EGNet (Zhao et al., 2019a) outperform others in terms of various evaluation metrics across the dataset.

## 6 Future recommendations

Future SOD networks should be able to achieve the primary goals of SOD in the most complex scenarios. This chapter discusses some future directions for SOD.

Feature aggregation: Many deep learning models need help extracting compelling features and aggregating them despite pre-training multi-scale, multi-level features of the CNN network. On the one hand, for feature aggregation, the crude method of combining all levels of features into the transfer layer (Zhang et al., 2017b) may introduce information redundancy and noisy feature interference into the model. On the other hand, excessive control over the exchange of information between stages (Zhang et al., 2018) may severely hinder the network's learning ability. These outstanding issues in feature aggregation suggest that when merging features from different layers, the focus should be placed on reducing aliasing effects and noise interference to generate useful features for saliency detection.

Inspiration from traditional models: Very few superoxide dismutase models based on deep learning use the saliency map of the traditional superoxide dismutase model as a saliency map to guide the saliency process (Wang et al., 2016; Chen et al., 2018). On the one hand, it (Wang et al., 2016), the saliency prior is used to initialize a recurrent framework, while in (88), the prior saliency map can replace the coarse saliency map for reverse attention. Optimization. On the other hand, literature (Feng et al., 2019) implements dilation and erosion operators through max-pooling to create turning attention maps. Leveraging different methods to integrate heuristic saliency priors or tools into deep SOD is expected to improve its training and inference.

Dataset-related issues: The availability of large datasets with less bias is crucial for developing SOD models. The bias present in the training data set affects the model's ability to generalize to salient targets in complex scenes. Existing SOD data sets can be quickly browsed to observe whether central bias and data selection bias exist. It is essential to develop datasets with more realistic scenarios and less bias while keeping the scale large. The proposed model performs better on some selected images than the ground truth. A more stringent annotation procedure should be developed to improve this situation, emphasizing acceptable annotation.

Real-time performance: Recently, DNN models (Chen et al., 2020; Qin et al., 2020) have been proposed for the needs of mobile and embedded applications. Achieving this through convolutional layers with fewer channels results in compact models and high efficiency. Qin et al. (Qin et al., 2020)designed a two-layer nested U-shaped structure lightweight network, which trained SOD from scratch. Recently, literature (Zhang et al., 2019) also proposed a pixel-wise saliency prediction based on knowledge distillation to solve the problem of a large memory footprint.

## 7 Conclusion

We have presented a comprehensive overview of Salient Object Detection (SOD), a computer vision task that aims to identify and

segment the most prominent objects or regions in images. We have discussed the evolution of SOD methods from traditional ones that rely on hand-crafted features or heuristic priors to deep learning-based ones that leverage robust neural networks and large-scale datasets. We have also introduced the main challenges and evaluation metrics of SOD, as well as the most influential and recent models in the field. We hope this mini-review can provide a valuable reference for researchers and practitioners who are new to SOD.

## Author contributions

XW: Conceptualization, Writing–original draft, Methodology, Visualization. SY: Funding acquisition, Supervision, Writing–original draft. EL: Funding acquisition, Supervision, Writing–review and editing. MW: Conceptualization, Funding acquisition, Writing–review and editing, Methodology.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work is partially supported by the XJTLU AI University Research Centre, Jiangsu Province Engineering Research Centre of Data Science and Cognitive Computation at XJTLU, and the SIP AI innovation platform (YZCXPT2022103).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

## Publisher's note



## References

Achanta, R., Hemami, S., Estrada, F., and Susstrunk, S. (2009). "Frequency-tuned salient region detection," in 2009 IEEE conference on computer vision and pattern recognition, Miami, FL, USA, 20-25 June 2009 (IEEE), 1597–1604.

Akila, K., Indra Priyadharshini, S., Ulaganathan, P., Prempriya, P., Yuvasri, B., Suriya Praba, T., et al. (2021). Ontology based multiobject segmentation and classification in sports videos. *J. Intelligent Fuzzy Syst.* 41, 5399–5409. doi:10.3233/jifs-189862

Alexe, B., Deselaers, T., and Ferrari, V. (2012). Measuring the objectness of image windows. *IEEE Trans. Pattern Analysis Mach. Intell.* 34, 2189–2202. doi:10.1109/TPAMI.2012.28

Borji, A., Cheng, M. M., Jiang, H., and Li, J. (2014). Salient object detection: a survey. *ArXiv e-prints*. doi:10.48550/arXiv.1411.5878

Borji, A., Cheng, M. M., Jiang, H., and Li, J. (2015). Salient object detection: a benchmark. *IEEE Tip.* 24, 5706–5722. doi:10.1109/TIP.2015.2487833

Borji, A., and Itti, L. (2011). "Scene classification with a sparse set of salient regions," in IEEE International Conference on Robotics and Automation, Shanghai, China, 09-13 May 2011 (IEEE), 1902–1908.

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi:10.1023/a:1010933404324

Cai, Y., Dai, L., Wang, H., Chen, L., and Li, Y. (2020). A novel saliency detection algorithm based on adversarial learning model. *IEEE Trans. Image Process.* 29, 4489–4504. doi:10.1109/TIP.2020.2972692

Chang, K. Y., Liu, T. L., Chen, H. T., and Lai, S. H. (2011). "Fusing generic objectness and visual saliency for salient object detection," in 2011 International Conference on Computer Vision, Kobe, Japan, 12-17 May 2009 (IEEE), 914–921.

Chen, B., Li, P., Sun, C., Wang, D., Yang, G., and Lu, H. (2019). Multi attention module for visual tracking. *Pattern Recognit.* 87, 80–93. doi:10.1016/j.patcog.2018.10.005

Chen, S., Tan, X., Wang, B., and Hu, X. (2018). "Reverse attention for salient object detection," in Proceedings of the European conference on computer vision (Europe: ECCV), 234–250.

Chen, S., Tan, X., Wang, B., Lu, H., Hu, X., and Fu, Y. (2020). Reverse attention-based residual network for salient object detection. *IEEE Trans. Image Process.* 29, 3763–3776. doi:10.1109/tip.2020.2965989

Cheng, M. M., Mitra, N. J., Huang, X., Torr, P. H. S., and Hu, S. M. (2015). Global contrast based salient region detection. *IEEE TPAMI* 37, 569–582. doi:10.1109/TPAMI.2014.2345401

Cheng, M. M., Warrell, J., Lin, W. Y., Zheng, S., Vineet, V., and Crook, N. (2013). Efficient salient region detection with soft image abstraction. *IEEE ICCV*, 1529–1536. doi:10.1109/iccv.2013.193

Donoser, M., Urschler, M., Hirzer, M., and Bischof, H. (2009). "Saliency driven total variation segmentation," in 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September 2009 - 02 October 2009 (IEEE), 817–824.

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* 88, 303–338. doi:10.1007/s11263-009-0275-4

Fan, D. P., Cheng, M. M., Liu, Y., Li, T., and Borji, A. (2017). "Structure-measure: a new way to evaluate foreground maps," in Proceedings of the IEEE international conference on computer vision, Venice, Italy, 22-29 October 2017, 4548–4557.

Fan, D. P., Gong, C., Cao, Y., Ren, B., Cheng, M. M., and Borji, A. (2018). Enhanced-alignment measure for binary foreground map evaluation. *arXiv Prepr. arXiv: 1805.10421*. doi:10.24963/ijcai.2018/97

Feng, M., Lu, H., and Ding, E. (2019). "Attentive feedback network for boundary-aware salient object detection," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Long Beach, CA, USA, 15-20 June 2019, 1623–1632.

Filali, I., Allili, M. S., and Benblidia, N. (2016). Multi-scale salient object detection using graph ranking and global–local saliency refinement. *Signal Process. Image Commun.* 47, 380–401. doi:10.1016/j.image.2016.07.007

Frintrop, S., and Kessel, M. (2009). "Most salient region tracking," in 2009 IEEE International Conference on Robotics and Automation, Kobe, Japan, 12-17 May 2009 (IEEE), 1869–1874.

Fu, H., Xu, D., and Lin, S. (2017). Object-based multiple foreground segmentation in rgbd video. *IEEE Trans. Image Process.* 26, 1418–1427. doi:10.1109/tip.2017.2651369

Gopalakrishnan, V., Hu, Y., and Rajan, D. (2010). Random walks on graphs for salient object detection in images. *IEEE Trans. Image Process.* 19, 3232–3242. doi:10.1109/tip.2010.2053940

Gupta, A. K., Seal, A., Prasad, M., and Khanna, P. (2020). Salient object detection techniques in computer vision—a survey. *Entropy* 22, 1174. doi:10.3390/e22101174

He, S., Jiao, J., Zhang, X., Han, G., and Lau, R. W. (2017a). Delving into salient object subitizing and detection. *Proc. IEEE Int. Conf. Comput. Vis.*, 1059–1067. doi:10.1109/iccv.2017.120

He, S., Jiao, J., Zhang, X., Han, G., and Lau, R. W. (2017b). "Delving into salient object subitizing and detection," in Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22-29 October 2017 (Venice, Italy: ICCV).

Hou, Q., Cheng, M. M., Hu, X., Borji, A., Tu, Z., and Torr, P. H. (2017). Deeply supervised salient object detection with short connections. *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 3203–3212. doi:10.1109/cvpr.2017.563

Hu, X., Fu, C. W., Zhu, L., Wang, T., and Heng, P. A. (2020). Sac-net: spatial attenuation context for salient object detection. *IEEE Trans. Circuits Syst. Video Technol.* 31, 1079–1090. doi:10.1109/tcsvt.2020.2995220

Hu, Y., Rajan, D., and Chia, L. T. (2005). Robust subspace analysis for detecting visual attention regions in images. *Proc. 13th Annu. ACM Int. Conf. Multimedia*, 716–724. doi:10.1145/1101149.1101306

Islam, M. A., Kalash, M., and Bruce, N. D. (2018). Revisiting salient object detection: simultaneous detection, ranking, and subitizing of multiple salient objects. *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 7142–7150. doi:10.1109/cvpr.2018.00746

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 1254–1259. doi:10.1109/34.730558

Jia, X., DongYe, C., and Peng, Y. (2022). Siatrans: siamese transformer network for rgb-d salient object detection with depth image classification. *Image and Vis. Comput.* 127, 104549. N.PAG. doi:10.1016/j.imavis.2022.104549

Jia, Y., and Han, M. (2013). Category-independent object-level saliency detection. *Proc. IEEE Int. Conf. Comput. Vis.*, 1761–1768. doi:10.1109/iccv.2013.221

Jiang, H., Wang, J., Yuan, Z., Liu, T., Zheng, N., and Li, S. (2011). Automatic salient object segmentation based on context and shape prior. *BMVC* 6 (9), 110. doi:10.5244/c.25.110

Jiang, P., Ling, H., Yu, J., and Peng, J. (2013). Salient region detection by ufo: uniqueness, focusness and objectness. *Proc. IEEE Int. Conf. Comput. Vis.*, 1976–1983. doi:10.1109/iccv.2013.248

Jiang, P., Pan, Z., Tu, C., Vasconcelos, N., Chen, B., and Peng, J. (2019). Super diffusion for salient object detection. *IEEE Trans. Image Process.* 29, 2903–2917. doi:10.1109/tip.2019.2954209

Joseph, S. I. T., Sasikala, J., and Sujitha Juliet, D. (2019). A novel vessel detection and classification algorithm using a deep learning neural network model with morphological processing (m-dlnn). *Soft Comput. - A Fusion Found. Methodol. Appl.* 23, 2693–2700. doi:10.1007/s00500-018-3645-4

Kim, J., Han, D., Tai, Y. W., Kim, J., Kim, T. W., Kim, K. P., et al. (2014). "S-1 plus oxaliplatin versus capecitabine plus oxaliplatin for the first-line treatment of patients with metastatic colorectal cancer: updated results from a phase 3 trial," in Proceedings of the IEEE conference on computer vision and pattern recognition, Columbus, OH, USA, 23-28 June 2014, 883–890.

Kim, J., and Pavlovic, V. (2016). "A shape-based approach for salient object detection using deep learning," in *Computer vision–ECCV 2016: 14th European conference, Amsterdam, The Netherlands, october 11–14, 2016, proceedings, Part IV 14*(Springer), 455–470.

Klein, D. A., and Frintrop, S. (2011). "Center-surround divergence of feature statistics for salient object detection," in 2011 International Conference on Computer Vision, Barcelona, Spain, 06-13 November 2011 (IEEE), 2214–2219.

Lee, G., Tai, Y. W., and Kim, J. (2016a). Deep saliency with encoded low level distance map and high level features. *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 660–668. doi:10.1109/cvpr.2016.78

Lee, G., Tai, Y. W., and Kim, J. (2016b). Deep saliency with encoded low level distance map and high level features. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. doi:10.1109/cvpr.2016.78

Lee, H., and Kim, D. (2018). "Salient region-based online object tracking," in 2018 IEEE Winter conference on applications of computer vision (WACV), Lake Tahoe, NV, USA, 12-15 March 2018 (IEEE), 1170–1177.

Li, G., Xie, Y., Lin, L., and Yu, Y. (2017). "Instance-level salient object segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, 21-26 July 2017, 2386–2395.

Li, H., Lu, H., Lin, Z., Shen, X., and Price, B. (2015). Inner and inter label propagation: salient object detection in the wild. *IEEE Trans. Image Process.* 24, 3176–3186. doi:10.1109/TIP.2015.2440174

Li, X., Li, Y., Shen, C., Dick, A., and Van Den Hengel, A. (2013). "Contextual hypergraph modeling for salient object detection," in Proceedings of the IEEE international conference on computer vision, Sydney, NSW, Australia, 01-08 December 2013, 3328–3335.

Li, X., Yang, F., Cheng, H., Liu, W., and Shen, D. (2018). Contour knowledge transfer for salient object detection. *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 355–370. doi:10.1007/978-3-030-01267-0_22

Li, Y., Hou, X., Koch, C., Rehg, J. M., and Yuille, A. L. (2014). "The secrets of salient object segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23-28 June 2014 (Columbus, OH, USA: CVPR).

Liu, F., and Gleicher, M. (2006). "Region enhanced scale-invariant saliency detection," in 2006 IEEE International Conference on Multimedia and Expo, Toronto, ON, Canada, 09-12 July 2006 (IEEE), 1477–1480.

Liu, J. J., Hou, Q., Cheng, M. M., Feng, J., and Jiang, J. (2019a). "A simple pooling-based design for real-time salient object detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15-20 June 2019 (Long Beach, CA, USA: CVPR).

Liu, J. J., Hou, Q., Cheng, M. M., Feng, J., and Jiang, J. (2019b). "A simple pooling-based design for real-time salient object detection," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Long Beach, CA, USA, 15-20 June 2019, 3917–3926.

Liu, N., Han, J., and Yang, M. H. (2018a). "Picanet: learning pixel-wise contextual attention for saliency detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (United States: CVPR).

Liu, N., Han, J., and Yang, M. H. (2018b). Picanet: learning pixel-wise contextual attention for saliency detection. *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 3089–3098. doi:10.1109/cvpr.2018.00326

Liu, T., Sun, J., Zheng, N., Tang, X., and Shum, H. (2007). "Learning to detect a salient object," in 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17-22 June 2007, 1–8.

Liu, Y., Wang, Y., and Kong, A. W. K. (2021). Pixel-wise ordinal classification for salient object grading. *Image and Vis. Comput.* 106, 104086. N.PAG. doi:10.1016/j.imavis.2020.104086

Luo, Z., Mishra, A., Achkar, A., Eichel, J., Li, S., and Jodoin, P. M. (2017). "Non-local deep features for salient object detection," in Proceedings of the IEEE Conference on computer vision and pattern recognition, Honolulu, HI, USA, 21-26 July 2017, 6609–6617.

Ma, C., Miao, Z., Zhang, X. P., and Li, M. (2017). A saliency prior context model for real-time object tracking. *IEEE Trans. Multimedia* 19, 2415–2424. doi:10.1109/tmm.2017.2694219

Ma, S., and Yang, J. J. (2023). Image-based vehicle classification by synergizing features from supervised and self-supervised learning paradigms. *Eng* 4, 444–456. doi:10.3390/eng4010027

Ma, Y. F., and Zhang, H. J. (2003). Contrast-based image attention analysis by using fuzzy growing. *Proc. eleventh ACM Int. Conf. Multimedia*, 374–381. doi:10.1145/957013.957094

Mehrani, P., and Veksler, O. (2010). Saliency segmentation based on learning and graph cut refinement. *BMVC (Citeseer)* 41, 1–12.

Movahedi, V., and Elder, J. H. (2010). "Design and perceptual validation of performance measures for salient object segmentation," in 2010 IEEE computer society conference on computer vision and pattern recognition-workshops (IEEE), 49–56.

Noh, H., Hong, S., and Han, B. (2015). Learning deconvolution network for semantic segmentation. *Proc. IEEE Int. Conf. Comput. Vis.*, 1520–1528. doi:10.1109/iccv.2015.178

Pang, Y., Zhao, X., Zhang, L., and Lu, H. (2020). "Multi-scale interactive network for salient object detection," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Seattle, WA, USA, 13-19 June 2020, 9413–9422.

Perazzi, F., Krähenbühl, P., Pritch, Y., and Hornung, A. (2012). "Saliency filters: contrast based filtering for salient region detection," in 2012 IEEE conference on computer vision and pattern recognition, Providence, RI, USA, 16-21 June 2012 (IEEE), 733–740.

Qin, C., Zhang, G., Zhou, Y., Tao, W., and Cao, Z. (2014). Integration of the saliency-based seed extraction and random walks for image segmentation. *Neurocomputing* 129, 378–391. doi:10.1016/j.neucom.2013.09.021

Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O. R., and Jagersand, M. (2020). U2-net: going deeper with nested u-structure for salient object detection. *Pattern Recognit.* 106, 107404. doi:10.1016/j.patcog.2020.107404

Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., and Jagersand, M. (2019). "Basnet: boundary-aware salient object detection," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Long Beach, CA, USA, 15-20 June 2019, 7479–7489.

Rosin, P. L. (2009). A simple method for detecting salient regions. *Pattern Recognit.* 42, 2363–2371. doi:10.1016/j.patcog.2009.04.021

Shelhamer, E., Long, J., and Darrell, T. (2017). Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Analysis Mach. Intell.* 39, 640–651. doi:10.1109/TPAMI.2016.2572683

Shen, Y., Ji, R., Zhang, S., Zuo, W., and Wang, Y. (2018). Generative adversarial learning towards fast weakly supervised detection. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 5764–5773. doi:10.1109/cvpr.2018.00604

Su, J., Li, J., Zhang, Y., Xia, C., and Tian, Y. (2019). Selectivity or invariance: boundary-aware salient object detection. *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 3799–3808. doi:10.1109/iccv.2019.00390

Su, Y., Zhao, Q., Zhao, L., and Gu, D. (2014). Abrupt motion tracking using a visual saliency embedded particle filter. *Pattern Recognit.* 47, 1826–1834. doi:10.1016/j.patcog.2013.11.028

Sun, J., Lu, H., and Li, S. (2012). "Saliency detection based on integration of boundary and soft-segmentation," in 2012 19th IEEE International Conference on Image Processing, Orlando, FL, USA, 30 September 2012 - 03 October 2012 (IEEE), 1085–1088.

Sun, J., Lu, H., and Liu, X. (2015). Saliency region detection based on markov absorption probabilities. *IEEE Trans. Image Process.* 24, 1639–1649. doi:10.1109/tip.2015.2403241

Tang, M., Djelouah, A., Perazzi, F., Boykov, Y., and Schroers, C. (2018). Normalized cut loss for weakly-supervised cnn segmentation. *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 1818–1827. doi:10.1109/cvpr.2018.00195

Tang, Y., and Wu, X. (2019). Salient object detection using cascaded convolutional neural networks and adversarial learning. *IEEE Trans. Multimedia* 21, 2237–2247. doi:10.1109/TMM.2019.2900908

Wang, L., Chen, R., Zhu, L., Xie, H., and Li, X. (2020). Deep sub-region network for salient object detection. *IEEE Trans. Circuits Syst. Video Technol.* 31, 728–741. doi:10.1109/tcsvt.2020.2988768

Wang, L., Lu, H., Ruan, X., and Yang, M. H. (2015b). "Deep networks for saliency detection via local estimation and global search," in Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, MA, USA, 07-12 June 2015, 3183–3192.

Wang, L., Lu, H., Wang, Y., Feng, M., Wang, D., Yin, B., et al. (2017a). "Learning to detect salient objects with image-level supervision," in Proceedings of the IEEE Conference on Computer Vision and Pattern, Honolulu, HI, USA, 21-26 July 2017 (Honolulu, HI, USA: CVPR).

Wang, L., Lu, H., Wang, Y., Feng, M., Wang, D., Yin, B., et al. (2017b). "Learning to detect salient objects with image-level supervision," in Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, 21-26 July 2017, 136–145.

Wang, L., Ouyang, W., Wang, X., and Lu, H. (2015a). Visual tracking with fully convolutional networks. *Proc. IEEE Int. Conf. Comput. Vis.*, 3119–3127. doi:10.1109/iccv.2015.357

Wang, L., Wang, L., Lu, H., Zhang, P., and Ruan, X. (2016b). "Saliency detection with recurrent fully convolutional networks," in *Computer vision–ECCV 2016: 14th European conference, Amsterdam, The Netherlands, october 11–14, 2016, proceedings, Part IV 14* (Springer), 825–841.

Wang, Q., Yuan, Y., Yan, P., and Li, X. (2013). Saliency detection by multiple-instance learning. *IEEE Trans. Cybern.* 43, 660–672. doi:10.1109/TSMCB.2012.2214210

Wang, T., Borji, A., Zhang, L., Zhang, P., and Lu, H. (2017c). "A stagewise refinement model for detecting salient objects in images," in Proceedings of the IEEE international conference on computer vision, Venice, Italy, 22-29 October 2017, 4019–4028.

Wang, T., Zhang, L., Wang, S., Lu, H., Yang, G., Ruan, X., et al. (2018). "Detect globally, refine locally: a novel approach to saliency detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, 18-23 June 2018, 3127–3135.

Wang, W., Zhao, S., Shen, J., Hoi, S. C., and Borji, A. (2019). "Salient object detection with pyramid attention and salient edges," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Long Beach, CA, USA, 15-20 June 2019, 1448–1457.

Wang, X., Ma, H., and Chen, X. (2016a). Geodesic weighted bayesian model for saliency optimization. *Pattern Recognit. Lett.* 75, 1–8. doi:10.1016/j.patrec.2016.02.008

Wei, J., Wang, S., Wu, Z., Su, C., Huang, Q., and Tian, Q. (2020). "Label decoupling framework for salient object detection," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Seattle, WA, USA, 13-19 June 2020, 13025–13034.

Wu, R., Feng, M., Guan, W., Wang, D., Lu, H., and Ding, E. (2019c). "A mutual learning method for salient object detection with intertwined multi-supervision," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Long Beach, CA, USA, 15-20 June 2019, 8150–8159.

Wu, Z., Su, L., and Huang, Q. (2019a). Cascaded partial decoder for fast and accurate salient object detection. *Proc. IEEE/CVF Conf. Comput. Vis. pattern Recognit.*, 3907–3916. doi:10.1109/CVPR.2019.00403

Wu, Z., Su, L., and Huang, Q. (2019b). "Stacked cross refinement network for edge-aware salient object detection," in Proceedings of the IEEE/CVF international conference on computer vision, Seoul, Korea (South), 27 October 2019 - 02 November 2019, 7264–7273.

Xie, Y., Lu, H., and Yang, M. H. (2012). Bayesian saliency via low and mid level cues. *IEEE Trans. Image Process.* 22, 1689–1698. doi:10.1109/TIP.2012.2216276

Xu, Y., Xu, D., Hong, X., Ouyang, W., Ji, R., Xu, M., et al. (2019). Structured modeling of joint deep feature and prediction refinement for salient object detection. *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 3789–3798. doi:10.1109/iccv.2019.00389

Yan, Q., Xu, L., Shi, J., and Jia, J. (2013a). "Hierarchical saliency detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23-28 June 2013 (Portland, OR, USA: CVPR).

Yan, Q., Xu, L., Shi, J., and Jia, J. (2013b). "Hierarchical saliency detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, Portland, OR, USA, 23-28 June 2013, 1155–1162.

Yang, C., Zhang, L., Lu, H., Ruan, X., and Yang, M. H. (2013). "Saliency detection via graph-based manifold ranking," in Proceedings of the IEEE conference on computer vision and pattern recognition, Portland, OR, USA, 23-28 June 2013, 3166–3173.

Yang, J., and Yang, M. H. (2017). Top-down visual saliency via joint crf and dictionary learning. *IEEE Trans. Pattern Analysis Mach. Intell.* 39, 576–588. doi:10.1109/TPAMI.2016.2547384

Yu, S., Zhang, B., Xiao, J., and Lim, E. G. (2021). Structure-consistent weakly supervised salient object detection with local saliency coherence. *Proc. AAAI Conf. Artif. Intell.* 35, 3234–3242. doi:10.1609/aaai.v35i4.16434

Zeng, Y., Zhuge, Y., Lu, H., and Zhang, L. (2019). "Joint learning of saliency detection and weakly supervised semantic segmentation," in Proceedings of the IEEE/CVF international conference on computer vision, Seoul, Korea (South), 27 October 2019 - 02 November 2019, 7223–7233.

Zhang, J., Sclaroff, S., Lin, Z., Shen, X., Price, B., and Mech, R. (2016). "Unconstrained salient object detection via proposal subset optimization," in Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 27-30 June 2016, 5733–5742.

Zhang, J., Yu, X., Li, A., Song, P., Liu, B., and Dai, Y. (2020a). Weakly-supervised salient object detection via scribble annotations. *Proc. IEEE/CVF Conf. Comput. Vis. pattern Recognit.*, 12546–12555. doi:10.1109/cvpr42600.2020.01256

Zhang, J., Yu, X., Li, A., Song, P., Liu, B., and Dai, Y. (2020b). "Weakly-supervised salient object detection via scribble annotations," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13-19 June 2020 (Seattle, WA, USA: CVPR).

Zhang, J., Zhang, T., Dai, Y., Harandi, M., and Hartley, R. (2018a). "Deep unsupervised saliency detection: a multiple noisy labeling perspective," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18-23 June 2018 (Salt Lake City, UT, USA: CVPR).

Zhang, L., Ai, J., Jiang, B., Lu, H., and Li, X. (2017a). Saliency detection via absorbing Markov chain with learnt transition probability. *IEEE Trans. image Process.* 27, 987–998. doi:10.1109/tip.2017.2766787

Zhang, L., Dai, J., Lu, H., He, Y., and Wang, G. (2018b). "A bi-directional message passing model for salient object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, 18-23 June 2018, 1741–1750.

Zhang, L., Zhang, J., Lin, Z., Lu, H., and He, Y. (2019a). "Capsal: leveraging captioning to boost semantics for salient object detection," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Long Beach, CA, USA, 15-20 June 2019, 6024–6033.

Zhang, P., Su, L., Li, L., Bao, B., Cosman, P., Li, G., et al. (2019b). Training efficient saliency prediction models with knowledge distillation. *Proc. 27th ACM Int. Conf. multimedia*, 512–520. doi:10.1145/3343031.3351089

Zhang, P., Wang, D., Lu, H., Wang, H., and Ruan, X. (2017b). "Amulet: aggregating multi-level convolutional features for salient object detection," in Proceedings of the IEEE international conference on computer vision, Venice, Italy, 22-29 October 2017, 202–211.

Zhang, P., Wang, D., Lu, H., Wang, H., and Yin, B. (2017c). "Learning uncertain convolutional features for accurate saliency detection," in Proceedings of the IEEE International Conference on computer vision, Venice, Italy, 22-29 October 2017, 212–221.

Zhang, X., Wang, T., Qi, J., Lu, H., and Wang, G. (2018c). "Progressive attention guided recurrent network for salient object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, 8-23 June 2018, 714–722.

Zhao, J. X., Liu, J. J., Fan, D. P., Cao, Y., Yang, J., and Cheng, M. M. (2019a). Egnet: edge guidance network for salient object detection. *Proc. IEEE/CVF Int. Conf. Comput. Vis*, 8779–8788. doi:10.1109/iccv.2019.00887

Zhao, J. X., Liu, J. J., Fan, D. P., Cao, Y., Yang, J., and Cheng, M. M. (2019b). "Egnet: edge guidance network for salient object detection," in Proceedings of the IEEE/CVF International Conference on Computer Vision (USA: ICCV).

Zhao, R., Ouyang, W., Li, H., and Wang, X. (2015a). "Saliency detection by multi-context deep learning," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 07-12 June 2015 (Boston, MA, USA: CVPR).

Zhao, R., Ouyang, W., Li, H., and Wang, X. (2015b). "Saliency detection by multi-context deep learning," in Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, MA, USA, 07-12 June 2015, 1265–1274.

Zhao, X., Pang, Y., Zhang, L., Lu, H., and Zhang, L. (2020). "Suppress and balance: a simple gated network for salient object detection," in *Computer vision–ECCV 2020: 16th European conference, glasgow, UK, august 23–28, 2020, proceedings, Part II 16* (Springer), 35–51.

Zhu, D., Dai, L., Luo, Y., Zhang, G., Shao, X., Itti, L., et al. (2018). Multi-scale adversarial feature learning for saliency detection. *Symmetry* 10, 457. doi:10.3390/sym10100457

Zhu, W., Liang, S., Wei, Y., and Sun, J. (2014). "Saliency optimization from robust background detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, Columbus, OH, USA, 23-28 June 2014, 2814–2821.