Neural Computing and Applications (2023) 35:23387–23399
https://doi.org/10.1007/s00521-021-06395-8

S.I.: HUMAN-IN-THE-LOOP MACHINE LEARNING AND ITS APPLICATIONS

Active learning with sampling by joint global-local uncertainty for salient object detection

Longfei Li[1] · Haidong Fu[1,2] · Xin Xu[1,2]

Received: 2 February 2021 / Accepted: 26 July 2021 / Published online: 6 August 2021

Abstract
The training of the SOD model relies on abundant annotated data, which needs laborious and expensive manual labeling. The generated pseudo-labels for reducing the annotation of the salient object will inevitably introduce noise, which will degrade the performance of the model and cannot fully represent the ground truth of manual labeling. To address this issue, we propose a novel active sampling strategy for salient object detection. The method is made up of two parts: a prediction module and an active learning module. The prediction module predicts the saliency of the image and provides the saliency prediction map for the active learning module. Then, the active learning module measures the global uncertainty and local uncertainty of the prediction map, aiming to select the most informative samples for the model. The selected samples are manually annotated and added to the training set to retrain the prediction model. Experimental results on DUTS dataset indicate that the amount of data can be reduced by 48.3% with competitive performance compared with the state-of-the-art SOD model.

Keywords Salient object detection · Active learning · Measure of uncertainty

1 Introduction

Human visual attention almost effortlessly directs the observer to focus on salient objects. In the fields of neuroscience and computer vision, imitating this selective scanning has become a hot subject. In the past, in order to facilitate extensive visual applications, the results generated by various salient object detection (SOD) models have been directly applied to image segmentation editing [1–4] as well as manipulation [5, 6], visual tracking [7, 8] and user interface optimization [9]. These applications have achieved unprecedented breakthroughs, largely due to the disclosure of massive pixel-level annotation data sets [10, 11]. However, the training of saliency models are limited by the high cost of sample annotation. Hence, one of the most pressing issues is lowering the cost of salient region marking.

For the past few years, the enormous success of deep convolutional neural networks (deep CNNs) has fueled a flood of efforts to train CNNs for saliency detection [12–15]. CNN-based methods normally involve a huge quantity of data and pixel-level annotations for training. However, annotating images with pixel-level ground truth is very costly. Some scholars attempt to train models using pseudo-labels created by models rather than manually annotated images in order to minimize annotation costs. Methods for generating pseudo-labels of salient regions can be divided into weakly supervised learning [16, 17], semi-supervised learning [18], and unsupervised learning [19]. Weakly supervised learning methods train the saliency model by replacing pixel-level labeling with image-level labeling or higher-level labeling. Compared with pixel-level labeling, image-level labeling only needs to annotate whether there is a salient object in the image, minimizing the complexity and expense of labeling. Such methods

✉ Haidong Fu
drfu@163.com

✉ Xin Xu
xuxin@wust.edu.cn

[1] School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China

[2] Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan University of Science and Technology, Wuhan 430065, China

[16, 17] rely excessively on semantic information, which leads to many salient regions being unrecognizable. Meanwhile, it weakens the supervision and cannot to ensure the generation ability of the model. Semi-supervised learning methods [18] train the classifier with a small amount of pixel-level labeled samples and then generate more detailed ground-truth through the classifier. The ground truth generated by this kind of method depends on the performance of the classifier, which leads to the annotation quality cannot be guaranteed. Unsupervised learning methods [19] use traditional manual features to refine learning and generate noisy pseudo-labels, which are then used to train SOD models. This approach has limited applicability. For those with little difference between foreground and background (low contrast), with small foreground objects, and with multiple salient objects, the ground-truth with high credibility cannot be well generated. Others [20] manually annotate the samples after active learning and selecting, and then add well-label data to the training set for retraining of the model. The annotated data generated by this kind of method will not introduce noise while reducing the cost of annotation by reducing the data volume and achieving good results.

However, the active strategy [20] is applied to the traditional SOD models based on machine learning to select images by measuring the uncertainty of the foreground of the region proposal generated by the input image. The proposed active learning strategy [20] only measures the local uncertainty of the image and ignores the uncertainty measurement of the background, which cannot reflect the image's overall uncertainty. Furthermore, all of the mainstream SOD models are built on a deep learning framework. Therefore, we propose a joint global-local uncertainty active learning strategy suitable for deep learning models, which is applied to deep learning models and achieves good results. This strategy starts from the uncertainty measurement of the whole image. The input images are subjected to local and global uncertainty measurements, and the images' uncertainty measurements are then synthetically evaluated by associating them with the intersection relation in mathematics. Eventually, the combined uncertainty measure is used to select the images for manual annotations by the oracle, which are then added to the training set. The saliency prediction model is retrained with the updated training set until the convergence condition is reached.

To summarize, the following are our major contributions:

- To address the problem of training data collection for salient object detection, a new active learning framework is proposed to combine with hard sample mining to alleviate the labeling work.
- To reduce the labeling cost, we design a global-local joint uncertainty measure to select the hard salient samples and integrate the active learning strategy into the deep learning-based SOD model. Thus, the optimal performance of the model can be obtained directly at a lower annotation cost.
- We evaluate the proposed approach comprehensively and compare 10 advanced SOD methods on 6 public datasets. The results indicate that our proposed strategy can effectively reduce 48.3% of labeling and close to the performance of full well-label data.

2 Related work

Currently, more and more saliency detection models have achieved great performance [21], owing to the availability of well-labeled SOD datasets [10, 11]. Yet, the main challenge is the limited amount of labeled samples with expensive labeling costs. Regarding the data collection in SOD, it is largely limited to manual annotation, which is a cumbersome process. The scale of labeled datasets in SOD is very limited as compared to other computer vision tasks (such as person re-identification [22], object detection [23], etc.). According to the manner of data utilization, SOD model training methods can be divided into three categories: full annotation learning, weak annotation learning, and active learning. We'll introduce them in detail as follows.

2.1 Learning from full annotations

The aim of salient object detection is to find the most salient object regions in images. Early SOD researches mainly focus on handcrafted features and heuristic priors, such as central priors [24] and boundary background priors [25]. Subsequently, motivated by the great success of deep CNNs in a variety of visual tasks [26, 27], much deep SOD methods [15, 28–31] have been proposed. These methods are based on single-scale [30] or multi-scale [15, 28, 29, 31] feature extraction, which divides the pixels or superpixels in the image into salient or non-salient categories to classify salient regions. However, the output of these approaches is often coarse because of the loss of spatial information in the full connection layer.

Salient object detection methods based on fully convolutional neural network (FCN) [32, 33] outperformed those based on deep CNNs, likely due to FCN's ability to capture multi-scale details and richer spatial information. To create more accurate maps, Hou et al. [12] created short connections on the full convolutional neural network and merged features from different levels. Chen et al. [34]

introduced a reverse attention network that would erase the currently predicted salient region and mine the missing part. By integrating the features of deep and shallow layers, Deng et al. [35] formulated an iterative technique for learning the residual map between predicted outcomes and ground truth. Liu et al. [36] suggested a U-Net liked pixel-level context attention network [37] framework that merged global and local context to learn the context of each pixel for saliency prediction. Wu et al. [38] introduced a mutual learning method that integrates saliency detection, edge detection, and foreground contour detection tasks into an end-to-end network. Wang et al. [39] designed a recurrent FCN for saliency detection through iterative correction of prediction errors. Zhao and Wu et al. [40] proposed a PFA network consisting of a spatial attention module, a context-aware pyramid feature extraction module, and a channel module for saliency detection and training under the supervision of ensuring accurate edges. Qin et al. [41] proposed a boundary-aware network. To improve the edge quality, the network introduced a three-level single loss joint supervised saliency prediction model and then used the residual refinement module to refine the edges of the saliency object. Wei et al. [42] suggested a label decoupling framework (LDF) made up of a label decoupling module (LD) and a feature interaction network (FIN). The original salient map was directly decomposed by LD into a body map and a detailed map, with the body map concentrated in the object's central region and the detailed map concentrated in the edge region.

While these methods achieve excellent performance, they all require large quantities of costly pixel-level annotation data for training, and data collection is especially tedious, often involving thousands of hours of human effort. Therefore, our work mainly studies how to reduce the workload of labeling in the full supervision.

2.2 Learning from weak annotations

Learning from weak annotations is an effective way to reduce SOD annotation workload. In [10, 43], SOD models learn from the annotations of whether there is an object class in the image. Cholakkal et al. [32] treated saliency object detection as a weakly supervised learning problem and trained the model with image-level labels. Image-level labeling refers to whether a saliency object exists in the image and is easier to implement than pixel-level labeling. Hsu et al. [16] improved the work of Cholakkal et al. [32] by replacing handcrafted features with features learned based on the CNN framework to detect salience objects. There are also some SOD methods [10, 17, 44] that attempt to predict saliency by using the category-level label or the higher-level label to train the model. Although this kind of method reduces the difficulty of pixel-level annotation to some extent, it also weakens the generalization performance of the model. Another common approach is to use semi-supervised learning. For example, Huo et al. [18] used prior knowledge to conduct an initial estimation of foreground saliency and background possibility for the input images and then used the samples labeled with the saliency of the original estimate probability and a small amount of well-labeled training samples for semi-supervised training classifier. The trained classifier was used to refine the initial saliency map to generate the final saliency map, thus avoiding the annotation of some images. Such methods make use of unlabeled data to a certain extent, but fail to fully consider the impact of noise samples on the classifier, resulting in slow performance improvement of the model and even biased by noise samples. In addition, some methods [45, 46] generated pseudo-pixel-level saliency labels (pseudo-labels) in unsupervised learning, or automatically generate noisy saliency maps through contour information [47]. These saliency maps are gradually refined and used to provide more detailed pixel-level supervision, so as to train more effective deep SOD models. In [45], Zhang et al. generated saliency prediction through the fusion process and generated pseudo labels by fusion of intra-image levels [48] and inter-image levels [49] weak saliency maps generated by several classical unsupervised salient object detectors [33]. Zhang et al [46]. jointly learned the potential saliency and noise patterns from the noisy salient maps produced by several conventional unsupervised SOD methods [33, 50–52] and generated a more refined salient map for the next iteration of the training. However, compared with the supervised learning methods, the pseudo-labels generated by unsupervised learning introduce noise, which will degrade the performance of the model.

Many of these studies assumed that the labeled data are predetermined. Hence, the focus of our research is on how to integrate active learning with the supervised deep learning model. It aims to minimize the amount of data needed by the model and the annotation work by using active learning to pick data without affecting the model's final performance.

2.3 Active learning

Active learning is commonly used in computer vision to address data collection issues in different tasks, such as recognition [53], object detection [54], and classification [55]. A great number of heuristic queries [56–58] have been proposed to calculate the availability of unlabeled samples in recent years. A general heuristic query contains entropy [59], reducing the classifier's expected error [60], maximizing the diversity of samples selected [61], or maximizing the change in the expected labels [62].

The data collection methods for active learning are primarily categorized into three scenarios according to the various ways of choosing instances: The synthesis of the member query, stream-based, and pool-based selective sampling. Augluin [63] is the first one to introduce member query synthesis. In this scenario, the unlabeled data for the query is generated by the model itself. The performance of the model can be improved through repeated training. In specific fields such as [64] and [65], especially in the case that labels are labeled experimentally rather than manually, this method has been proved to be very effective and reliable. In this scenario, however, there are some drawbacks. For example, Baum and Lang [66] pointed out that when manual annotation is needed, the effect of query learning is poor, since, without semantic sense, the model could generate some unrecognized symbols. Subsequently, researchers introduced certain selective sampling, such as stream-based or pool-based sampling [67, 68], to address the deficiencies of the above scenarios. In a stream-based or continuum selection strategy, unlabeled samples are queried one by one on the model to determine whether to query the sample for labeling. Dagan and Engelson [69] introduce an information assessment methodology to make this determination, in which samples are measured and samples with more information are more likely to be queried. Cohn et al. [68] proposed an alternative approach. They identified a sample range as an uncertainty region and only queried the samples falling within that range. This uncertainty region is determined by the data distribution of the current training data [70] or the minimum threshold of information measure. What's more, pool-based sampling is commonly used because it is easy to rapidly gather vast quantities of unlabeled image in real-time. Unlike steam-based sampling, in pool-based sampling, query decisions do not have to be made individually and continuously for each sample. Because for a set of labeled samples and a set of unlabeled samples, the pool-based approach will evaluate and sort the entire unlabeled sample pool after learning the features of a few labeled data to select the sample set with the most informative to query and annotate together. Compared with weak annotation learning, manual annotation is used after selecting samples by active learning and no noise information is introduced. Therefore, the pool-based active learning strategy is more suitable for the SOD data collection task. At present, some studies have tried to use the active learning method to alleviate SOD data annotation. In [20], images are screened through the uncertainty measurement of the object region proposal in input images, and then the images are mapped by learning a feature mapping matrix. Afterward, images are clustered to pick out redundant samples. However, the measurement of background uncertainty is ignored in the process of model training. Therefore, our work proposes a pool-based active learning framework aimed at minimizing labeling effort and maximizing SOD model performance, in which the uncertainty of the overall measurement image is used as the basis for data selection.

We have previously explored the use of active learning strategies [22] to select hard samples in the person re-identification field. The samples with the most information were selected for retraining of the model based on a combination of the two measurement strategies of sample uncertainty and intra-diversity, to train a model with a limited amount of data and achieve good performance. Considering the non-categorization of data in the field of salient object detection. In this work, we will conduct a comprehensive assessment of the information content of saliency samples with uncertainty. Not only the local uncertainty measurement for the foreground region but also the global uncertainty measurement for the whole image. The proposed active learning structure is shown in Fig. 1.

3 Methodology

In this part, we'll go through the active learning approach as well as the saliency prediction model in detail. The salient prediction model is then designed in detail and retrained through these actively chosen images. Eventually, we defined the acquisition function's design, which actively decides training images.

3.1 Pool-based active learning framework

We use a pool-based framework to train an active sampling learner for salient detection in this work. Figure 1 depicts

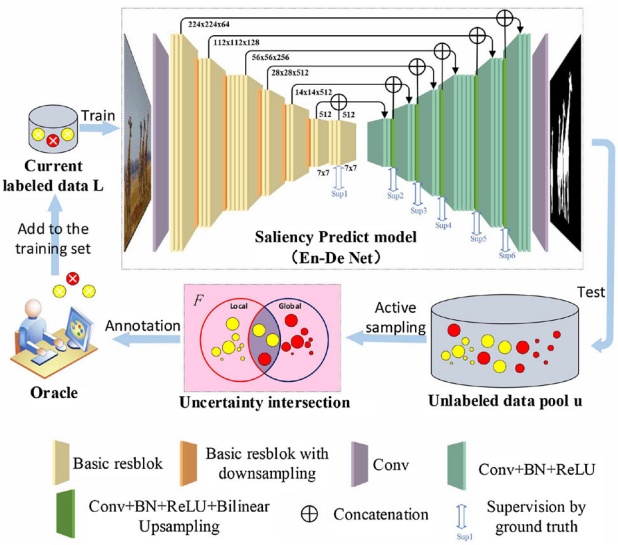

Fig. 1 Overview of the proposed SOD model based on active learning strategy

an engaged learner's overall workflow. Three components consist of this active sampling learner, which are the saliency prediction model $\mathcal{S}$, the acquisition function F based on global-local uncertainty sampling, and the current labeled data set $\mathcal{L}$, where $\mathcal{L} = \{(I_i, G_i)\}_{i=1}^{N}$ contains N training samples, and G_i is the pixel-wise ground truth of sample I_i.

The active learning process generally consists of the following steps. An unlabeled samples pool $\mathcal{U} = \{U_i\}_{i=1}^{M}$ with M unlabeled images is provided at the start. The saliency prediction model $\mathcal{S}$ is trained on the set of labeled samples $\mathcal{L}$, which is a set of randomly chosen labeled data, in the first iteration. Then, based on the designed acquisition function F, the learner systematically selects the most informative unlabeled samples from the pool $\mathcal{U}$. By adding those selected images, the set $\mathcal{L}$ is updated and the $\mathcal{S}$ is retrained on the updated labeled samples pool $\mathcal{L}$. This process is repeated until the termination condition is reached. The entire process is depicted in Algorithm 1.

Algorithm 1 Pool-based Active Learning Framework

Input: $\mathcal{L}$ (a number of labeled samples); $\mathcal{U}$ (unlabeled sample pool);
Output: salient predict net $\mathcal{S}$;

```
for iteration stopping criterion is reached do
    Train S on L;
    Unlabeled data pool U be predicted by SOD net S;
    Gain dataset with predicted values U' by step 3;
    Pick the set P from U' by F for annotation;
    Query the oracle for manual labelling on P ;
    Update L and U:L = L ∪ P, U = U \ P;
```

3.2 Salient predict model

It is commonly acknowledged that training deep neural networks requires a large amount of labeled data. To combine the active learning strategy proposed in this work to reduce the annotation cost, we use the Encoder-Decoder network as our saliency prediction model. A remarkable improvement in the structure is that it also has a large number of feature channels in the upsampling part, which allows the network to propagate context information to a higher convolution layer. This means it can simultaneously acquire low-level details information and high-level global contexts. Inspired by HED [71], to minimize overfitting, the final layer of each decoder process is supervised by the mark of ground truth (see Fig. 1). Inspired by BasNet [41], it starts with an input convolution layer in the encoder portion, and six phases consisting of basic res-blocks. ResNet-34 [72] is used in the input convolution layer and the subsequent four phases, in which the input layer contains 64 convolution kernels with the size of 3 x 3 and stride of 1. Furthermore, there is no pooling process after the input of the convolutional layer. That is, before the second step, the feature maps have the same spatial resolution as the input images. After the fourth phase of ResNet-34 [72], it adds two more phases to reach the same receptive field as ResNet-34. Following a non-overlapping max-pooling layer of size 2, both phases are made up of three basic res-blocks of 512 filters. We introduce a bridging step between the encoder and the decoder to better extract global information. It has three convolution layers, each with 512 dilated (dilation=2) [73] 3 x 3 filters. A batch normalization [74] and a ReLU activation function [75] are followed by both of these convolution layers.

For the encoder, the decoder is nearly symmetrical. Each phase includes three convolution layers, batch normalization, and a ReLU activation function. The input of each phase of the decoder is the concatenated feature maps of the upsampled output from its previous phases and their corresponding phases of the encoder. To obtain the side output saliency prediction results, the outputs of both the bridge phase and each decoder phase are transferred to a simple 3 3 convolution layer, then bilinear upsampling and a sigmoid function are performed. Therefore, our predicting module will create seven salient prediction results in the training process by inputting an image. The seven outputs, though, are supervised by the ground truth, and the last one has the greatest precision, where we calculate the input image's uncertainty. The training loss at each iteration can be defined by:

$$\mathcal{T} = \sum_{i=1}^{N} \alpha_i \ell^{(i)} \tag{1}$$

where $\ell^{(i)}$ represents the loss of the i-th iteration, N denotes the total number of iterations and the weight of each iteration is represented by α_i. Meanwhile, we set $\alpha_i = 1$ to treat all iterations equally to simplify the problem. We are going to get three outputs (i.e., bce, iou, and ssim) for each iteration and each of them corresponds to one loss. As described above, the seven outputs of the SOD model are all deeply supervised by $\mathcal{T}$.

In order to get a high-quality regional segmentation, we also define $\ell^{(i)}$ as mixing loss according to the suggestion of [41] :

$$\ell^{(i)} = \ell_{iou}^{(i)} + \ell_{bce}^{(i)} + \ell_{ssim}^{(i)} \tag{2}$$

where $\ell_{iou}^{(i)}$, $\ell_{bce}^{(i)}$ and $\ell_{ssim}^{(i)}$ denote IoU loss [76], BCE loss [77] and SSIM loss [78], respectively. The most commonly used loss in salient object detection and image segmentation is the binary cross-entropy (BCE) loss, which is described as:

$$\ell_{bce}^{(i)} = -\sum_{(i,j)} [(1 - G_{i,j}) * \log(1 - S_{i,j}) + G_{i,j} * \log(S_{i,j})] \tag{3}$$

where $S_{i,j}$ represents the pixel (i, j) predicted probability of being saliency region and $G_{i,j} \in \{0, 1\}$ represents ground truth mark of the pixel (i, j).

Initially, intersection over union(IoU) was proposed to measure the similarity between two sets [79] and then used for object detection and segmentation as a standard evaluation measure. Since the BCE separately measures the loss per pixel and lacks the image's global structure. To overcome this problem, we apply the IoU loss to $\ell^{(i)}$ as indicated by [41], which can calculate the resemblance of two images as a whole rather than a single pixel. It's defined as follows:

$$\ell_{iou}^{(i)} = 1 - \frac{\sum_{(i,j)}[G_{i,j} * S_{i,j}]}{\sum_{(i,j)}[G_{i,j} + S_{i,j} - G_{i,j} * S_{i,j}]} \tag{4}$$

where $S_{i,j}$ represents the pixel (i, j) predicted probability of being saliency region and $G_{i,j} \in \{0, 1\}$ represents the ground truth mark of the pixel (i, j).

The Structural Similarity Index (SSIM) is used for assessing image quality [78]. Because of its ability to extract the image's structural detail. To learn the structural information of the ground truth, we should incorporate it in our training loss. The SSIM loss is described as:

$$\ell_{ssim}^{(i)} = 1 - \frac{(2\mu_m\mu_n + C_1)(2\sigma_{mn} + C_2)}{(\mu_m^2 + \mu_n^2 + C_1)(\sigma_m^2 + \sigma_n^2 + C_2)} \tag{5}$$

where σ_m, σ_n and μ_m, μ_n are the standard deviations and mean of m and n, respectively. σ_{mn} is their covariance and g is the binary mask of ground truth. To avoid dividing by zero, we set $C_1 = 0.01^2$ and $C_2 = 0.03^2$. Meanwhile, let $m = \{m_i : i = 1, \ldots, N^2\}$ and $n = \{n_i : i = 1, \ldots, N^2\}$ be the pixel values of two related areas (size: N N) cropped from the saliency result map s.

3.3 Active acquisition algorithm based on joint global-local uncertainty

The acquisition function F is intended to measure the informativeness of an unlabeled sample by two perspectives: the image's local and global uncertainty.

3.3.1 Local uncertainty sampling

We introduce a local uncertainty metric α to score each sample based on the final output of the saliency prediction model. The final layer of the output is a two-dimensional probability matrix to show the saliency of each pixel in the image. In this work, we use the probability matrix $\mathcal{X}$ to query high-uncertainty images with a bigger proportion of "paradoxical pixels." The pixels closest to the decision boundary of model $\mathcal{S}$ are referred to as "Paradoxical pixels."

Particularly, the following is a description of the local uncertainty sampling process: The current SOD model is used to evaluate each unlabeled sample $\mathbf{U}$ from the current unlabeled sample pool $\mathcal{U}$, which computes a probability score s_i for each pixel $p_i \in \mathcal{X}$ of sample $\mathbf{U}$. The higher the pixel score is, the easier it is to be in the foreground, while the lower the pixel score is, the easier (it is to be) in the background. The middle part is known as "paradoxical," which provides more information, helps model training, and performs the minimum-maximum normalization for all pixel scores, as shown below:

$$s_i^{norm} = \frac{s_i - s_{min}}{s_{max} - s_{min}} \tag{6}$$

where s_{min} and s_{max} are the lowest and highest probability scores, respectively, in $\mathcal{X}$. The "paradoxical" pixel X_p is described as a pixel with a normalized score s^{norm} that falls between 0.4 and 0.6. The "paradoxical" pixels are classified as follows:

$$X_p = \{p_i \in \mathcal{X} \mid \forall i : 0.4 < s_i^{norm} < 0.6\} \tag{7}$$

In addition, we use the local uncertainty score α to measure the proportion of "paradoxical" pixels, which is:

$$\alpha = \frac{Sum(X_p)}{Sum(\mathcal{X})} \tag{8}$$

where Sum represents the sum of pixels. Assume that $\mathcal{A}$ is the set of local uncertainty fractions of the unlabeled sample $\mathbf{U}$, and we choose the sample with high local uncertainty as to the candidate set. Specifically, the selected candidate set consists of the following sets:

$$\mathcal{A} = \{\mathbf{U}_i \in \mathcal{U} \mid \forall i : \alpha_i > \mu + \rho\sigma\} \tag{9}$$

where μ denotes the average value of $\mathcal{A}$, ρ denotes a trade-off parameter, and σ denotes the standard deviation of $\mathcal{A}$.

3.3.2 Global uncertainty sampling

Since the local uncertainty can only roughly measure the uncertainty of the sample, we introduce information entropy to measure the global uncertainty of the sample, so as to further comprehensively evaluate the uncertainty of the sample. The global uncertainty metric β to measure the unlabeled sample pool $\mathbf{U}$ can be defined as:

$$\beta = -\sum_{s \in \mathcal{X}} p(s) \log p(s) \tag{10}$$

where $p(s)$ represents the probability scores of each pixel of matrix $\mathcal{X}$. Assume that $\mathcal{B}$ is the set of global uncertainty fractions for the unlabeled sample $\mathbf{U}$ and that the candidate

set is the sample with the high global uncertainty. The selected candidate set is comprised of the following sets:

$$\mathcal{B} = \{\mathbf{U}_i \in \mathcal{U} \,|\, \forall i : \beta_i > \mu' + \rho\sigma'\} \tag{11}$$

where μ' denotes the average value of $\mathcal{B}$, ρ denotes a trade-off parameter, and σ' denotes the standard deviation of $\mathcal{B}$. Eventually, the intersect of the candidate set α and β is collected by the set $\mathcal{P}$, and the set $\mathcal{P}$ to ask an external oracle for labeling. Algorithm 2 illustrates the whole process. Furthermore, the suggested active acquisition algorithm uses each image pixel to formulate both local and global uncertainty. This idea can be quickly and naturally extended to deep learning-based computer vision tasks.

Algorithm 2 Active Acquisition Algorithm

Input: $\mathcal{U}$ (unlabeled data pool);ρ (trade-off parameter);
Output: Query set $\mathcal{P}$ for annotation.

for each sample $\mathbf{U}$ in the unlabeled sample $\mathcal{U}$ **do**
  Generate a probability matrix $\mathcal{X}$ by SOD model ;
  $X_p = 0, Temp = 0$;
  for each probability score s_i in $\mathcal{X}$ **do**
    Normalize for s_i by Equ. 6;
    if (accord with Equ. 7) **then**
      X_p++;
    else
      continue;
    Compute the informativeness of pixel q_i' by Equ. 10
    $Temp \; += q_i'$;
  end
  Compute α of sample $\mathbf{U}$ by Equ. 8,9;
  Compute β of sample $\mathbf{U}$ by value of Temp,Equ. 11;
  $\mathcal{P} = \alpha \cap \beta$;
end

4 Experiments

4.1 Datasets and evaluation metrics

Six widespread public datasets, including PASCAL-S [80] with 850 samples, ECSSD [33] with 1000 samples, HKU-IS [29] with 4447 samples, DUT-OMRON [25] with 5168 samples, DUTS [10] with 15572 samples, and THUR15K [81] with 6232 samples, are used to test the proposed method. Among them, the largest saliency detection benchmark is DUTS, which contains 5019 testing samples (DUTS-TE) and 10,553 training samples (DUTS-TR). The SOD model was incrementally trained by selecting samples from DUTS-TR through active learning strategy, other datasets for evaluation. Six metrics will be used to assess the performance of our model and current mainstream methods. The mean absolute error (MAE) [82] is the first metric, which is defined as the mean absolute per-pixel difference between the mark of ground truth and its a predicted saliency map, as seen in Eq. 12. E-measure (E_m) [83], F-measure (F_m) , S-measure (S_m) [84], F-measure curves and precision-recall curves (PR curves) are also extensively evaluated saliency maps. Furthermore, the overall performance measurement computed by the weighted harmonic of precision and recall is known as the F-measure, as shown in Eq. 13.

$$MAE = \frac{1}{\mathcal{H} \times \mathcal{W}} \sum_{x=1}^{\mathcal{H}} \sum_{y=1}^{\mathcal{W}} |\mathcal{P}(x,y) - \mathcal{G}(x,y)| \tag{12}$$

where $\mathcal{G}$ represents the mark of ground-truth and $\mathcal{P}$ represents the predicted saliency map. $\mathcal{W}$ and $\mathcal{H}$ represent the width and height of $\mathcal{P}$, respectively.

$$F = \frac{(1+\delta^2) \times Precision \times Recall}{\delta^2 Precision + Recall} \tag{13}$$

where δ^2 is set to 0.3 to put a greater focus on accuracy, as indicated by [85]. Precision value is the proportion of the predicted saliency pixels that are correctly allocated to the saliency region, while Recall value is the proportion of the detected saliency pixels to the mark of ground-truth.

4.2 Implementation details

The SOD model is trained on DUTS-TR and tested on the above-mentioned six datasets. Our active learning strategy was first tested on the DUTS-TR dataset, and 1000 samples were randomly selected for manual labeling as the labeled training set, so as to initialize the performance of the SOD model. Each sample is resized to (256×256) pixels, then arbitrarily cropped to (224×224) pixels during training. We use the Adam optimizer [86] to train our saliency prediction model, with the hyper parameters set to default values: eps=1e-7, weight decay=0, betas=(0.9, 0.999), and the learning rate lrate=1e-4. Some of the encoder parameters are initialized using the ResNet-34 model [72].The rest of the parameters are initialized by Xavier [87]. Subsequently, local and global uncertainty measurements were carried out for the remaining 9553 samples, and their intersections were screened out, which were handed over to Oracle for manual annotation and then added to the labeled training set for the next round of training. Once the performance of the model appears tendency of decreasing or rising gently, the training process will be ended. According to [20], we defined the trade-off parameter ρ to be 1.145 in Equation 9,11 to calculate the acquisition function F. During the test, the input sample image is adjusted to the right size (256×256) and then sent to the SOD model to get the saliency map, which is then converted to the original size for comparison with the input sample image. Bilinear interpolation is used in all resizing procedures.

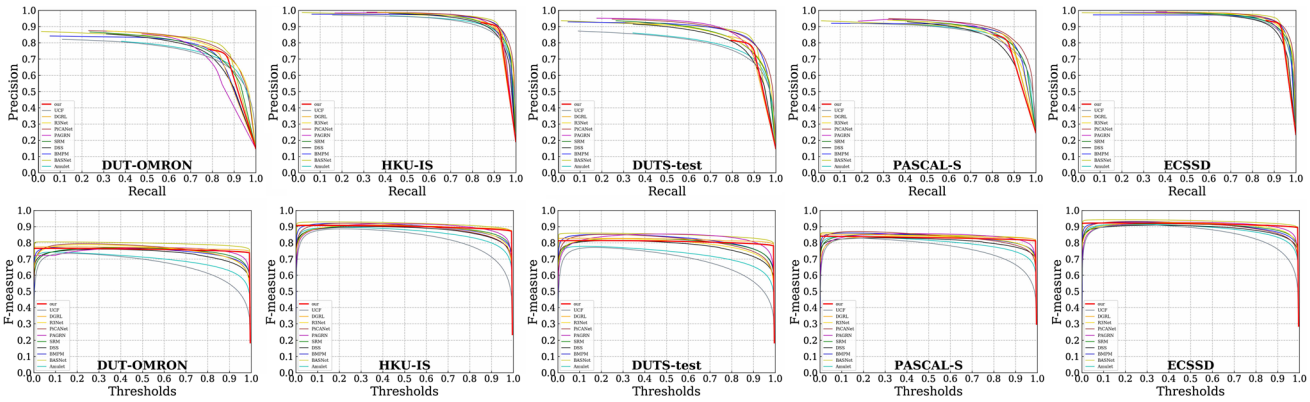

Fig. 2 PR curves (first row) and F-measure curves (second row) are shown on the five largest datasets

Table 1 Performance of the model trained by our proposed active sampling strategy on 6 datasets competes to that of the other ten mainstream methods in terms of the F_m, the S_m, the E_m and the MAE

Method		Our	Amulet	BMPM	DGRL	DSS	PAGRN	PiCANet	R3Net	SRM	BASNet	UCF
Train set	Name	DT	MK	DT	DT	MB	DT	DT	MK	DT	DT	MK
	Num	5457	10,000	10,553	10,553	2500	10,553	10,553	10,000	10,553	10,553	10,000
	F_m↑	*.838*	.772	**.829**	***.837***	.810	.790	.825	.818	.808	***.837***	.754
	S_m↑	**.777**	.754	***.790***	.774	.764	.720	*.793*	.738	.745	.772	.754
	E_m↑	.779	.773	***.803***	.788	*.804*	.779	**.802**	.784	.799	.779	.752
SOD	MAE↓	.109	.144	**.108**	***.106***	.124	.145	*.104*	.125	.128	.114	.148
	F_m↑	***.913***	.883	.901	***.913***	.884	**.904**	.900	.903	.897	*.927*	.853
	S_m↑	.905	.894	**.911**	.903	.883	.889	***.914***	.903	.895	*.916*	.883
	E_m↑	.914	.901	.914	**.917**	.908	.914	.910	***.920***	**.917**	*.921*	.879
ECSSD	MAE↓	**.041**	.059	.045	.042	.056	.061	.046	***.040***	.054	*.037*	.069
	F_m↑	**.896**	.857	.889	***.900***	.871	.890	.884	.881	.882	*.928*	.886
	S_m↑	.897	.883	***.907***	.894	.881	.887	**.906**	.892	.887	*.909*	.866
	E_m↑	.933	.910	.937	***.943***	.925	**.939**	.934	.928	.936	*.946*	.887
HKU-IS	MAE↓	**.037**	.052	.039	**.037**	.045	.048	.043	***.036***	.046	*.032*	.062
	F_m↑	.814	.727	.814	**.818**	.776	***.823***	.807	.787	.797	*.842*	.688
	S_m↑	.843	.803	*.867*	.842	.826	.838	**.861**	.836	.836	***.866***	.777
	E_m↑	.854	.789	.861	**.879**	.842	***.880***	.852	.841	.861	*.884*	.757
DUTS-TE	MAE↓	.059	.084	***.048***	.051	.059	.055	**.054**	.058	.058	*.047*	.112
	F_m↑	.827	.799	**.830**	*.841*	.807	***.832***	**.830**	.820	.827	*.841*	.762
	S_m↑	.822	.819	***.844***	.836	.803	.821	*.848*	.811	.834	**.838**	.803
	E_m↑	.829	.802	.842	**.847**	.831	*.853*	.833	.832	.846	***.852***	.770
PASCAL-S	MAE↓	.083	.098	*.074*	*.074*	.101	.089	**.081**	.092	.084	***.076***	.115
	F_m↑	**.757**	.694	.745	***.766***	.730	.750	**.757**	.753	.745	*.791*	.660
	S_m↑	**.819**	.780	.809	.806	.789	.775	***.826***	.818	.798	*.836*	.758
	E_m↑	.833	.778	.837	***.848***	.819	**.842**	.834	.824	.840	*.869*	.755
DUT-OMRON	MAE↓	.071	.098	.064	**.063**	***.062***	.071	.065	**.063**	.069	*.056*	.120

italic, bolditalic, bold and underline indicate the best, second best, third best and fourth performance. "MK" and "DT" represent training dataset MSRA10K and DUTS-TR, respectively

Table 2 Quantitative comparisons of different learning iteration on ECSSD dataset

Round	Size of training set	F-measure↑
1	1000	0.828
2	1954	0.854
3	2794	0.871
4	3470	0.883
5	4082	0.893
6	4591	0.899
7	5059	0.907
8	5457	0.913

4.3 Comparisons with other models

We compare our method with 10 state-of-the-art models, PiCANet [36], BMPM [88], R^3Net [35], PAGRN [89], DGRL [90], BASNet [41], DSS [12], SRM [91], Amulet [92], UCF [93]. We either use the authors' saliency maps or run their reported models for a fair comparison.

4.3.1 Quantitative evaluation

The F-measure curves and Precision–Recall (PR) curves of the SOD model tested on the five largest datasets based on our proposed active sampling strategy are shown in Fig. 2 to assess the quality segmentation of saliency object. In addition, for all testing datasets, Table 1 shows the E-measure (E_m) , the F-measure (F_m) , the S-measure (S_m), and the MAE measure. As shown, our SOD model differs little from the state-of-the-art with the four evaluation indexes. Even more to the point, only 51.7% of the largest training dataset DUTS are used to train the SOD model through the active selection strategy proposed in this work. Meanwhile, we also do research on SOD performance improvement corresponding to each iteration on the ECSSD dataset. The results are shown in Table 2. It is not difficult to find that through the mining of hard samples, the performance of the model can be significantly improved. Compared with the training of the whole dataset, this method not only reduces the learning of the model for useless (noise samples) or helpless samples (simple samples) but also reduces the amount of data that needs to be annotated manually.

4.3.2 Qualitative evaluation

Figure 3 illustrates several prediction instances from the prediction model as well as other state-of-the-art methods. It is observed that the SOD model trained by active learning has a similar performance to the current mainstream saliency model trained by full data. This method performs well in clearly highlighting the saliency object and restraining background noise. The model based on active sampling is also robust in dealing with a variety of

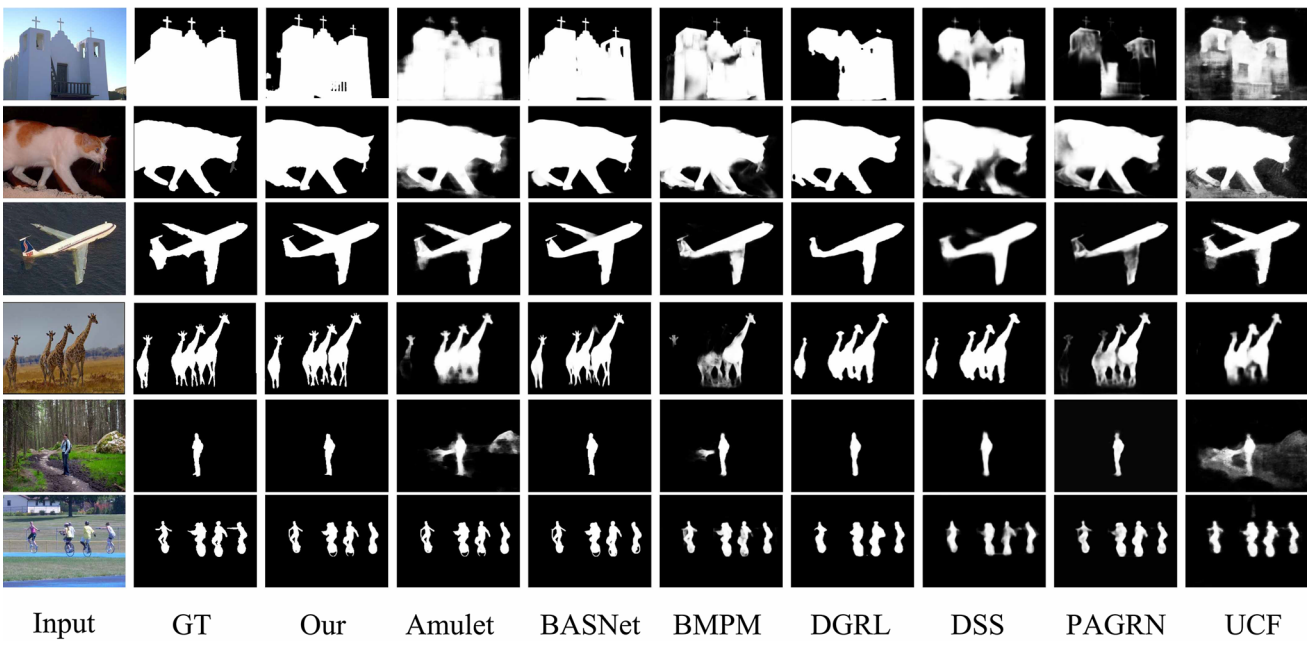

Fig. 3 Visual comparison of different models. Each row represents one sample and corresponding saliency maps. Each column represents the predictions of one model. Apparently, our model is equally good at dealing with messy backgrounds and producing more accurate and clear saliency maps

Table 3 Ablation study on different Training Mode, the experiments were carried out for three consecutive times to take the average value

Training mode	Size of training set	F-measure↑
Random select	5457	0.879
Active sampling	5457	0.913
Select all	10,553	0.922

Table 4 Ablation studies were carried out on different training sets, and the SOD model was trained on MSRA10K and tested on ECSSD

Round	Size of training set	F-measure↑	MAE↓
1	1000	0.823	0.085
2	1927	0.834	0.084
3	2694	0.858	0.074
4	3333	0.861	0.071
5	3891	0.869	0.069
6	4419	0.871	0.065
	10,000	0.889	0.061

challenging scenarios, including messy backgrounds, human structures, and low-contrast foreground objects.

4.4 Ablation study

To prove the effectiveness of the active selection strategy, we made corresponding ablation experiments, without using our proposed strategy, by randomly selected samples corresponding to salient prediction model for training, the results show that, under uncertainty based on active learning strategy, the SOD model not only improved convergence speed, and performance increase amplitude was relatively large. Meanwhile, we also used the whole number of DUT-TR for training, and then compared the performance. The results are shown in Table 3.

To demonstrate the applicability of our active sampling strategy on other training datasets, we trained the SOD model on another large MSRA10K training set and tested it on the ECSSD testing set. As shown in Table 4, the data volume was also reduced by 55.8%, and the performance was only 1.8% less different from the training of the whole MSRA10K dataset.

5 Conclusion

In this paper, we propose an active sampling approach focused on global-local uncertainty to minimize saliency labeling. The SOD model, which is based on active learning with local-global uncertainty, can not only significantly reduce the amount of data needed but also achieve good performance. Numerous experiments are conducted on six public datasets, and the results show that the proposed strategy can reduce the amount of data significantly and accelerate the efficiency of model training at the same time.

Acknowledgements This work was supported by the Natural Science Foundation of China (U1803262, 62006165).

Compliance with ethical standards

Conflict of interest No conflict of interest exits in this manuscript.

References

1. Xuebin Q, Shida H, Xiucheng Y, Masood D, Qiming Q, Jagersand M (2018) Accurate outline extraction of individual building from very high-resolution optical images. IEEE Geosci Remote Sens Lett 15(11):1775–1779
2. Timor K, Michael B (2001) Saliency, scale and image description. Int J Comput Vis 45(2):83–105
3. Xin X, Jinshan T, Xiaoming L, Xiaolong Z (2010) Human behavior understanding for video surveillance: recent advance. In: 2010 IEEE international conference on systems, man and cybernetics. IEEE, pp 3867–3873
4. Xuebin Q, Shida H, Zichen Z, Masood D, Martin J (2018) Bylabel: a boundary based semi-automatic image annotation tool. In: IEEE winter conference on applications of computer vision (WACV). IEEE, pp 1804–1813
5. Martin J (1995) Saliency maps and attention selection in scale and spatial coordinates: An information theoretic approach. In: Proceedings of IEEE international conference on computer vision. IEEE, pp 195–202
6. Roey M, Eli S, Lihi Z-M (2019) Saliency driven image manipulation. Mach Vis Appl 30(2):189–202
7. Hyemin L, Daijin K (2018) Salient region-based online object tracking. In: IEEE winter conference on applications of computer vision (WACV). IEEE, pp 1170–1177
8. Xuebin Q, Shida H, Camilo Perez Q, Abhineet S, Masood D, Martin J (2017) Real-time salient closed boundary tracking via line segments perceptual grouping. In: IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 4284–4289
9. Prakhar G, Shubh G, Ajaykrishnan J, Sourav P, Ritwik S (2018) Saliency prediction for mobile user interfaces. In: IEEE Winter conference on applications of computer vision (WACV). IEEE, pp 1529–1538
10. Lijun W, Huchuan L, Yifan W, Mengyang F, Dong W, Baocai Y, Xiang R (2017) Learning to detect salient objects with image-level supervision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 136–145
11. Tie L, Zejian Y, Jian S, Jingdong W, Nanning Z, Xiaoou T, Heung-Yeung S (2010) Learning to detect a salient object. IEEE Trans Pattern Anal Mach Intell 33(2):353–367
12. Qibin H, Ming-Ming C, Xiaowei H, Ali B, Zhuowen T, Philip HST (2017) Deeply supervised salient object detection with short connections. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 3203–3212
13. Xin X, Nan M, Xiaolong Z, Bo L (2016) Covariance descriptor based convolution neural network for saliency computation in

low contrast images. In: 2016 international joint conference on neural networks (IJCNN). IEEE, pp 616–623
14. Linzhao W, Lijun W, Huchuan L, Pingping Z, Xiang R (2016) Saliency detection with recurrent fully convolutional networks. In: European conference on computer vision. Springer, pp 825–841
15. Nan M, Xin X, Xiaolong Z, Hong Z (2018) Salient object detection using a covariance-based cnn model in low-contrast images. Neural Comput Appl 29(8):181–192
16. Kuang-Jui H, Yen-Yu L, Yung-Yu C (2019) Weakly supervised salient object detection by learning a classifier-driven map generator. IEEE Trans Image Process 28(11):5435–5449
17. Yu Z, Yunzhi Z, Huchuan L, Lihe Z, Mingyang Q, Yizhou Y (2019) Multi-source weak supervision for saliency detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 6074–6083
18. Shuwei H, Yuan Z, Sun-Yuan K (2017) Semi-supervised saliency classifier based on a linear feedback control system model. In: International joint conference on neural networks (IJCNN). IEEE, pp 3130–3137
19. Duc Tam N, Maximilian D, Chaithanya Kumar M, Thi Phuong Nhung N, Thi Hoai Phuong N, Zhongyu L, Thomas B (2019) Deepusps: deep robust unsupervised saliency prediction with self-supervision. arXiv preprint arXiv:1909.13055
20. Lihe Z, Jiayu S, Tiantian W, Yifan M, Huchuan L (2019) Visual saliency detection via kernelized subspace ranking with active learning. IEEE Trans Image Process 29:2258–2270
21. Wenguan W, Qiuxia L, Huazhu F, Jianbing S, Haibin L, Ruigang Y (2019) Salient object detection in the deep learning era: an in-depth survey. arXiv preprint arXiv:1904.09146
22. Xin X, Lei L, Xiaolong Z, Weili G, Ruimin H (2021) Rethinking data collection for person re-identification: active redundancy reduction. Pattern Recognition, p 107827
23. Shuai S, Zeming L, Tianyuan Z, Chao P, Gang Y, Xiangyu Z, Jing L, Jian S (2019) Objects365: a large-scale, high-quality dataset for object detection. In: Proceedings of the IEEE/CVF international conference on computer vision. pp 8430–8439
24. Zhuolin J, Larry SD (2013) Submodular salient region detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2043–2050
25. Chuan Y, Lihe Z, Huchuan L, Xiang R, Ming-Hsuan Y (2013) Saliency detection via graph-based manifold ranking. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3166–3173
26. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25:1097–1105
27. Karen S, Andrew Z (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
28. Xin X, Shiqin W, Zheng W, Xiaolong Z, Ruimin H (2021) Exploring image enhancement for salient object detection in low light images. ACM Trans Multimedia Comput Commun Appl 17(1):1–19
29. Guanbin L, Yizhou Y (2015) Visual saliency based on multiscale deep features. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 5455–5463
30. Nian L, Junwei H, Dingwen Z, Shifeng W, Tianming L (2015) Predicting eye fixations using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 362–370
31. Rui Z, Wanli O, Hongsheng L, Xiaogang W (2015) Saliency detection by multi-context deep learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 1265–1274
32. Hisham C, Jubin J, Deepu R (2016) Backtracking scspm image classifier for weakly supervised top-down saliency. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5278–5287
33. Qiong Y, Li X, Jianping S, Jiaya J (2013) Hierarchical saliency detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 1155–1162
34. Shuhan C, Xiuli T, Ben W, Xuelong H (2018) Reverse attention for salient object detection. In: Proceedings of the European conference on computer vision (ECCV). pp 234–250
35. Zijun D, Xiaowei H, Lei Z, Xuemiao X, Jing Q, Guoqiang H, Pheng-Ann H (2018) R3net: Recurrent residual refinement network for saliency detection. In: Proceedings of the 27th international joint conference on artificial intelligence. AAAI Press, pp 684–690
36. Nian L, Junwei H, Ming-Hsuan Y (2018) Picanet: Learning pixel-wise contextual attention for saliency detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 3089–3098
37. Olaf R, Philipp F, Thomas B (2015) U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 234–241
38. Runmin W, Mengyang F, Wenlong G, Dong W, Huchuan L, Errui D (2019) 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. pp 8150–8159
39. Linzhao W, Lijun W, Huchuan L, Pingping Z, Xiang R (2018) Salient object detection with recurrent fully convolutional networks. IEEE Trans Pattern Anal Mach Intell 41(7):1734–1746
40. Ting Z, Xiangqian W (2019) Pyramid feature attention network for saliency detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 3085–3094
41. Xuebin Q, Zichen Z, Chenyang H, Chao G, Masood D, Martin J (2019) Basnet: Boundary-aware salient object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 7479–7489
42. Jun W, Shuhui W, Zhe W, Chi S, Qingming H, Qi T (2020) Label decoupling framework for salient object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 13025–13034
43. Chunshui C, Yongzhen H, Zilei W, Liang W, Ninglong X, Tieniu T (2018) Lateral inhibition-inspired convolutional neural network for visual attention and saliency detection. In: Proceedings of the AAAI conference on artificial intelligence. vol 32
44. Guanbin L, Yuan X, Liang L (2018) Weakly supervised salient object detection using image labels. In: Proceedings of the AAAI conference on artificial intelligence. vol 32
45. Dingwen Z, Junwei H, Yu Z (2017) Supervision by fusion: towards unsupervised learning of deep salient object detector. In: Proceedings of the IEEE international conference on computer vision. pp 4048–4056
46. Jing Z, Tong Z, Yuchao D, Mehrtash H, Richard H (2018) Deep unsupervised saliency detection: a multiple noisy labeling perspective. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 9029–9038
47. Xin L, Fan Y, Hong C, Wei L, Dinggang S (2018) Contour knowledge transfer for salient object detection. In: Proceedings of the European conference on computer vision (ECCV). pp 355–370
48. Jianming Z, Stan S, Zhe L, Xiaohui S, Brian P, Radomir M (2015) Minimum barrier salient object detection at 80 fps. In: Proceedings of the IEEE international conference on computer vision. pp 1404–1412

49. Jianming Z, Stan S (2015) Exploiting surroundedness for saliency detection: a boolean map approach. IEEE Trans Pattern Anal Mach Intell 38(5):889–902
50. Wangjiang Z, Shuang L, Yichen W, Jian S (2014) Saliency optimization from robust background detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 2814–2821
51. Bowen J, Lihe Z, Huchuan L, Chuan Y, Ming-Hsuan Y (2013) Saliency detection via absorbing markov chain. In: Proceedings of the IEEE international conference on computer vision. pp 1665–1672
52. Xiaohui L, Huchuan L, Lihe Z, Xiang R, Ming-Hsuan Y (2013) Saliency detection via dense and sparse reconstruction. In: Proceedings of the IEEE international conference on computer vision. pp 2976–2983
53. Liang L, Keze W, Deyu M, Wangmeng Z, Lei Z (2017) Active self-paced learning for cost-effective and progressive face identification. IEEE Trans Pattern Anal Mach Intell 40(1):7–19
54. Keze W, Xiaopeng Y, Dongyu Z, Lei Z, Liang L (2018) Towards human-machine cooperation: Self-supervised sample mining for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 1605–1613
55. Xin L, Yuhong G (2014) Multi-level adaptive active learning for scene classification. In: European conference on computer vision. Springer, pp 234–249
56. Keze W, Dongyu Z, Ya L, Ruimao Z, Liang L (2016) Cost-effective active learning for deep image classification. IEEE Trans Circuits Syst Video Technol 27(12):2591–2600
57. Yarin G, Riashat I, Zoubin G (2017) Deep bayesian active learning with image data. In: International conference on machine learning. PMLR, pp 1183–1192
58. William HB, Tim G, Andreas N, Jan MK (2018) The power of ensembles for active learning in image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 9368–9377
59. David DL, Jason C (1994) Heterogeneous uncertainty sampling for supervised learning. In: Machine learning proceedings. Elsevier, pp 148–156
60. Sudheendra V, Ashish K (2010) Visual recognition and detection under bounded computational resources. In: IEEE computer society conference on computer vision and pattern recognition. IEEE, pp 1006–1013
61. Hoi SCH, Jin R, Zhu J, Lyu MR (2009) Semisupervised svm batch mode active learning with applications to image retrieval. ACM Trans Inf Syst 27(3):1–29
62. Alexander V, Vittorio F, Joachim MB (2012) Weakly supervised structured output learning for semantic segmentation. In: IEEE conference on computer vision and pattern recognition. IEEE, pp 845–852
63. Dana A (1988) Queries and concept learning. Mach Learn 2(4):319–342
64. Dana A (2004) Queries revisited. Theor Comput Sci 313(2):175–194
65. King RD, Whelan KE, Jones FM, Reiser PGK, Bryant CH, Muggleton SH, Kell DB, Oliver SG (2004) Functional genomic hypothesis generation and experimentation by a robot scientist. Nature 427(6971):247–252
66. Eric BB, Kenneth L (1992) Query learning can work poorly when a human oracle is used. In: International joint conference on neural networks. vol 8, p 8
67. Les EA, David AC, Richard EL (1990) Training connectionist networks with queries and selective sampling. In: Advances in neural information processing systems. Citeseer, pp 566–573
68. David C, Les A, Richard L (1994) Improving generalization with active learning. Mach Learn 15(2):201–221
69. Ido D, Sean PE (1995) Committee-based sampling for training probabilistic classifiers. In: Machine learning proceedings. Elsevier, pp 150–157
70. Mitchell TM (1982) Generalization as search. Artif Intell 18(2):203–226
71. Saining X, Zhuowen T (2015) Holistically-nested edge detection. In: Proceedings of the IEEE international conference on computer vision. pp 1395–1403
72. Kaiming H, Xiangyu Z, Shaoqing R, Jian S (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 770–778
73. Fisher Y, Vladlen K (2015) Multi-scale context aggregation by dilated convolutions. arXiv preprint. arXiv:1511.07122
74. Sergey I, Christian S (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning. PMLR, pp 448–456
75. Hahnloser RHR, Sebastian Seung H, Slotine J-J (2003) Permitted and forbidden sets in symmetric threshold-linear networks. Neural Comput 15(3):621–638
76. Gellért M, Wenjie L, Raquel U (2017) Deeproadmapper: extracting road topology from aerial images. In: Proceedings of the IEEE international conference on computer vision. pp 3438–3446
77. De Boer P-T, Kroese DP, Mannor S, Rubinstein RY (2005) A tutorial on the cross-entropy method. Ann Oper Res 134(1):19–67
78. Zhou W, Eero PS, Alan CB (2003) Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh asilomar conference on signals, systems & computers. IEEE, vol 2, pp 1398–1402
79. Paul J (1912) The distribution of the flora in the alpine zone. New Phytol 11(2):37–50
80. Yin L, Xiaodi H, Christof K, James MR, Alan LY (2014) The secrets of salient object segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 280–287
81. Cheng M-M, Mitra NJ, Huang X, Shi-Min H (2014) Salientshape: group saliency in image collections. Vis Comput 30(4):443–453
82. Federico P, Philipp K, Yael P, Alexander H (2012) Saliency filters: contrast based filtering for salient region detection. In: IEEE conference on computer vision and pattern recognition. IEEE, pp 733–740
83. Deng-Ping F, Cheng G, Yang C, Bo R, Ming-Ming C, Ali B (2018) Enhanced-alignment measure for binary foreground map evaluation. arXiv preprint arXiv:1805.10421
84. Deng-Ping F, Ming-Ming C, Yun L, Tao L, Ali B (2017) Structure-measure: a new way to evaluate foreground maps. In: Proceedings of the IEEE international conference on computer vision. pp 4548–4557
85. Radhakrishna A, Sheila H, Francisco E, Sabine S (2009) Frequency-tuned salient region detection. In: IEEE conference on computer vision and pattern recognition. IEEE, pp 1597–1604
86. Diederik PK, Jimmy B (2014) Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980
87. Xavier G, Yoshua B (2010) Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR workshop and conference proceedings, pp 249–256
88. Lu Z, Ju D, Huchuan L, You H, Gang W (2018) A bi-directional message passing model for salient object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 1741–1750
89. Xiaoning Z, Tiantian W, Jinqing Q, Huchuan L, Gang W (2018) Progressive attention guided recurrent network for salient object

detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 714–722
90. Tiantian W, Lihe Z, Shuo W, Huchuan L, Gang Y, Xiang R, Ali B (2018) Detect globally, refine locally: a novel approach to saliency detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 3127–3135
91. Tiantian W, Ali B, Lihe Z, Pingping Z, Huchuan L (2017) A stagewise refinement model for detecting salient objects in images. In: Proceedings of the IEEE international conference on computer vision. pp 4019–4028
92. Pingping Z, Dong W, Huchuan L, Hongyu W, Xiang R (2017) Amulet: aggregating multi-level convolutional features for salient object detection. In: Proceedings of the IEEE international conference on computer vision. pp 202–211
93. Pingping Z, Dong W, Huchuan L, Hongyu W, Baocai Y (2017) Learning uncertain convolutional features for accurate saliency detection. In: Proceedings of the IEEE international conference on computer vision. pp 212–221