Pattern Recognition Letters 125 (2019) 124–132

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

# Finding autofocus region in low contrast surveillance images using CNN-based saliency algorithm ☆

Nan Mu [a], Xin Xu [a,b,c,*], Xiaolong Zhang [a,b]

[a] School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China
[b] Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan University of Science and Technology, Wuhan 430065, China
[c] School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

ARTICLE INFO

*Article history:*
Available online 17 April 2019

*Keywords:*
Convolutional neural network (CNN)
Autofocus
Salient object detection
Low contrast surveillance image

ABSTRACT

How to automatically locate the focus region in low contrast image is a key issue for camera-equipped surveillance devices. Due to low signal to noise ratio, the performance of autofocus will seriously decline in low contrast image, making it quite difficult to recognize the focus region. To tackle this problem, we perform autofocus by conducting a salient object detection approach. A covariance based deep learning framework is proposed to evaluate the saliency of low contrast surveillance image. Based on the mechanism of human visual system, the autofocus region can be identified by the visual salient object. In this paper, low-level features of the low contrast images are first studied and extracted. Then the mutual covariances of the segmented blocks are trained via a 7-layers convolutional neural network (CNN). Next, the initial saliency map of the testing image can be obtained by estimating the saliency score of each block via the pre-trained CNN model. Finally, the resulting saliency map is refined by introducing the local-global difference and internal similarity approaches. Experimental results demonstrate that the proposed method outperforms existing ten state-of-the-art saliency models on three public datasets and a nighttime image dataset.



## 1. Introduction

Autofocus (AF) has become one of the key technologies in surveillance device equipped with digital image sensor, which contributes the imaging equipment to automatically focus on the target objects, it influences the quality and efficiency of imaging directly. For most imaging systems, the realization of autofocus is based on the passive AF algorithm, the focusing region is determined by searching the peak of image sharpness function. However, under low contrast conditions, the sharpness function of the image becomes flat, which leads the peak position difficult to locate. For this reason, this paper selects the focus region by detecting the salient object.

Visual saliency refers to a selection mechanism, the task of which is to extract the most important information for further processing. The research of saliency detection in surveillance images has proven to be useful for computer vision applications. With the development of various saliency models, it has witnessed tremendous advances in visual saliency detection in recent years. Most of these models focus on the contrast difference between salient objects and background region. The contrast between image elements (pixel, superpixel, or region) can be analogously used to compare the saliency of these elements, thus the natural images can be converted to saliency maps.

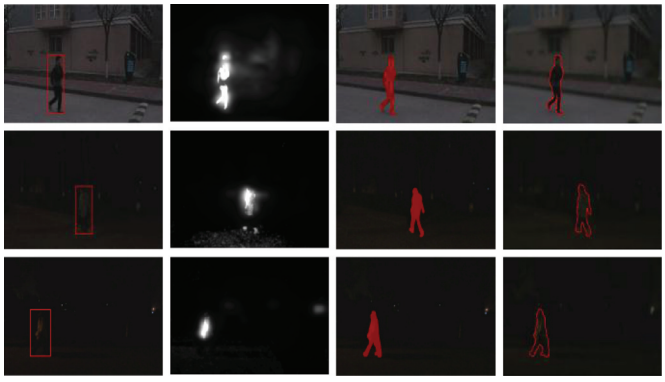
In fact, salient object detection mainly depends on the surrounding environment, it aims to find the most visually distinctive objects relative to other objects in an image. Since the process of autofocus is to set the principal subject in the scene to the clearest state to distinguish it from the background, this is completely consistent with the mechanism of salient object detection. Therefore, the objects obtained by saliency detection can be regarded as the main subjects of autofocus.

Currently, most salient models can work well in images with high contrast between foreground and background. But for detecting salient object in a relatively low contrast scene, they may face difficulties. Fig. 1 illustrates the saliency detection results and the corresponding autofocus images using the proposed model. The testing images have low lightness, low contrast and there is little difference between the visual salient objects and the background. For such images, traditional saliency approaches fail to determine the real salient object, and most existing focusing region selection methods cannot find an accurate focus position.

☆ **Conflict of interest.** The authors declare no conflict of interest.

* Corresponding author.
*E-mail address:* xuxin@wust.edu.cn (X. Xu).

https://doi.org/10.1016/j.patrec.2019.04.011
0167-8655/



(a) Input images (b) Salient objects (c) Focus regions (d) AF images

**Fig. 1.** Examples of salient object detection and autofocus results. (a) Input low contrast surveillance images with manually labeled rectangle. (b) Saliency maps obtained by the proposed saliency model. (c) Focusing region selection based on the salient object. (d) Autofocusing images.

To solve the above problems, this paper first proposed a *convolutional neural network* (CNN) based saliency model, which can extract the salient object from low contrast surveillance images efficiently. Then, the autofocusing image can be acquired according to the salient region. The overview of the proposed method is shown in Fig. 2. For the training images, 28 low-level features (4 color features, 12 steerable pyramid features, and 12 Gabor filter features) are first extracted to form a 28-dimensional feature vector. Then, the covariance matrices of random image blocks are computed, which are regarded as the training samples. Next, a 7-layers CNN model is trained to evaluate the saliency score of the testing image. For each testing image, it is first divided into non-overlapping blocks at multi-scales, then each block can be labeled with a score by the pre-trained covariance-based CNN model. At last, the local-global contrast measure and the internal similarity strategy are utilized to refine the predicting saliency values to obtain the focus regions.

The proposed saliency model has a more preferable performance than existing models for saliency detection in low contrast surveillance images, thus it has robust autofocus performance. This model has been tested on the MSRA dataset created by Liu et al. [1], the SED dataset created by Alpert et al. [2], the CSSD dataset created by Yan et al. [3], and the *nighttime image* (NI) dataset[1] created by this project to corroborate its excellent performance. After obtaining the visual salient object, which can be regarded as an optimal focus region, the focusing measure is then implemented to achieve auto focusing.

The main contributions of this paper are as follows:

(1) A covariance-based CNN model is proposed to incorporate the low-level features with high-level features to explore the properties of low contrast images.
(2) The local-global contrast and internal similarity measures are developed to adequately interpret the visual salient object and refine the structure context.
(3) By applying the information of visual attention area benefited from saliency detection, the robust autofocus region can be well identified.
(4) The performance of our model is evaluated on four datasets involving complex and/or low contrast scenes and achieves favorable results in comparison with ten state-of-the-art methods.

[1] The nighttime image dataset of this paper can be downloaded from https://drive.google.com/open?id=0BwVQK2zsuAQwX2hXbnc3ZVMzejQ.

The proposed technology has proven to be stable and robust carried out under Matlab programming environment on a standard personal computer. Experiments demonstrated that our method has a lower computational complexity and can excellently work on low contrast images, these advantages guarantee that the proposed method can be extended and implemented in electronic imaging devices and surveillance systems. In addition, the proposed method can be applied to optimize various consumer application technologies in low contrast conditions, such as person re-identification [4–7], landmark retrieval [8], cross-modal retrieval [9], multiview clustering [10,11], fine-grained visual recognition [12], etc.

The rest of this paper is organized as follows. Section 2 gives a review of related works on autofocus systems and saliency models. Section 3 describes the proposed salient object detection model. Section 4 presents the experimental comparison results of the proposed saliency model with other existing saliency models. Finally, the conclusions are drawn in Section 5.

## 2. Related works

In this section, several autofocus methods and salient object detection models are briefly summarized.

### 2.1. Focusing region selection methods

Focus region selection is the first step in autofocus system, which directly influences the accuracy of autofocus measure. Thus, a good focus region should contain the *region of interesting* (ROI) from the image, which can be obtained by searching the salient object.

In the autofocus system, the selection of focus region can not only reduce the amount of data processing and speed up the autofocus measure, but also eliminate the influence of uninteresting region on the evaluation function curve, thus it can improve the accuracy of autofocus. The common focusing region selection algorithms are shown as follows:

(1) Center region selection algorithm

In traditional optical focus camera, focusing region is the center of the image [13,14]. These methods assume that the foreground objects are often placed at the center of the picture. Although it can meet the requirement of most occasions, the performance will seriously decline when the main object is off center. For an $M \times N$ image $I(x, y)$, the center region is obtained via:

$$Rcenter = \sum_{\frac{3}{8}M \le x \le \frac{5}{8}M-1, \frac{3}{8}N \le y \le \frac{5}{8}N-1} I(x, y). \quad (1)$$

(2) Multipoint region selection algorithm

The multipoint region selection methods [15,16] select multiple windows for computing respectively, and choose an optimal focusing region, which contains the main objects. The final result is acquired by adopting a weighted fusion method, which can take advantage of these multiple windows, and can obtain a more accurate object region. However, these methods increase the overall computational cost, and introduce more noises and backgrounds.

(3) Non-uniform sampling region selection algorithm

The main idea of the non-uniform sampling is to utilize different sampling rates in different image regions, thus the center of image can keep the highest resolution, and the resolution of edge region decreases according to the distance to image center. The non-uniform sampling region selection methods [17,18] can keep the details of the central regions, and are robust to the high-resolution images, but these methods may lead to the wrong focus when the main object is not at the image center.

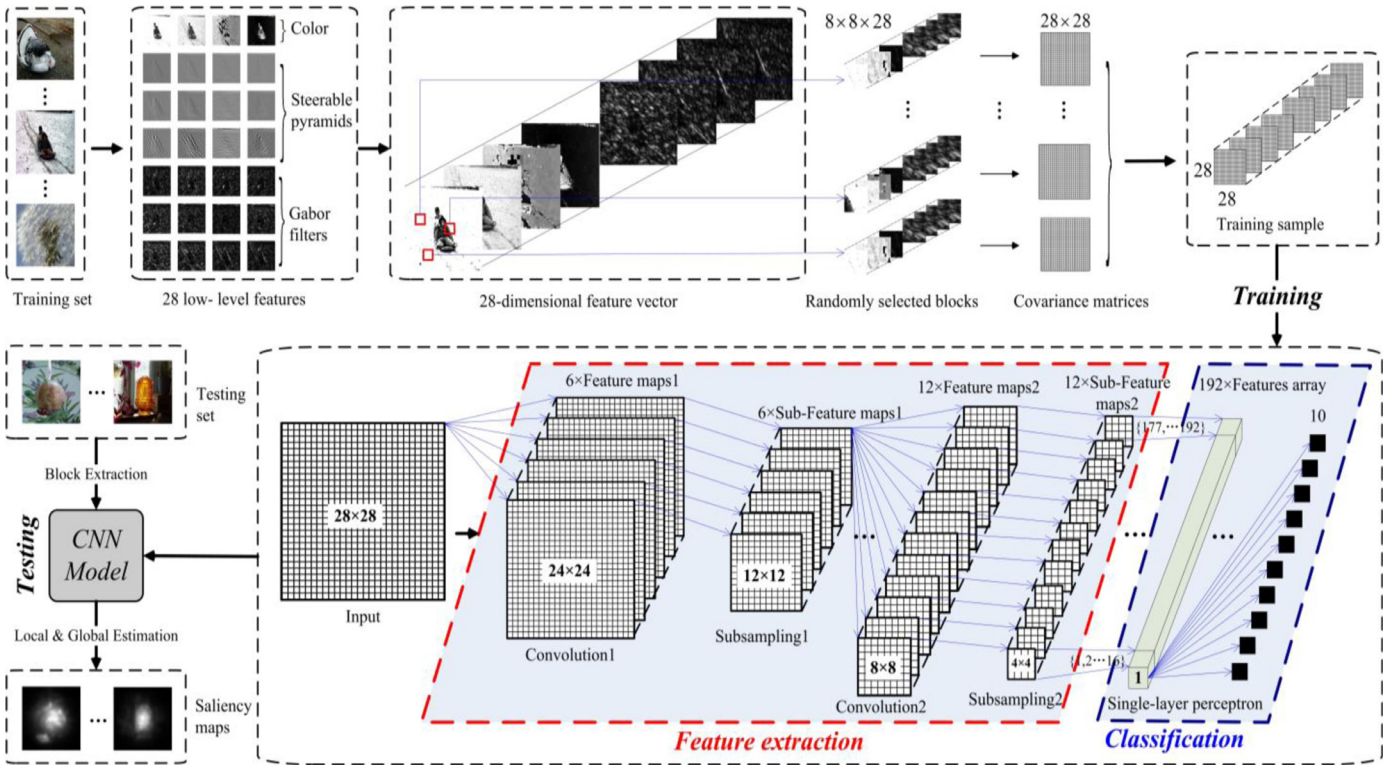



**Fig. 2.** An overview of the proposed framework.

### 2.2. Salient object detection models

Generally speaking, most existing saliency detection approaches can be broadly classified into two main categories, in which the bottom-up and top-down approaches are considered, respectively.

The bottom-up saliency models are data-driven and primarily use low-level image features. The most popular method in this category is the saliency model proposed by Itti et al. [19], which computes three feature contrasts (luminance, color and direction) in different scales. Murray et al. [20] calculated the image saliency based on the center-surround mechanism and scale-weighting function. Hou et al. [21] extracted the saliency features by utilizing the image signature, which can highlight the sparse salient region. Goferman et al. [22] proposed a context-aware saliency model, which considers the color and contrast features. Margolin et al. [23] detected the salient object by calculating inner statistics of the image patches. Yang et al. [24] introduced a graph-based mainfold ranking method to compute the superpixel saliency. Zhu et al. [25] explored boundary connectivity to measure the spatial layout and presented a principled optimization method to obtain the saliency map. Jiang et al. [26] proposed a diffusion-based saliency algorithm by building the diffusion matrix and the seed vector. Zhang et al. [27] presented a color saliency model by considering the local color-orientation interactions and spatiochromatic filtration.

The top-down salient object detection models are mainly task-driven and usually use the cognitive visual features. Tong et al. [28] introduced a bootstrap learning algorithm to train a strong saliency model. Xu et al. [29] utilized support vector machine to train distinctive features to detect the salient object. He and Lau [30] constructed a locate-by-exemplar top-down saliency model, which is based on the deep association. Li and Yu [31] presented a multiscale deep convolutional neural networks based saliency model. Yang and Yang [32] built the top-down saliency model by using the joint conditional random field and dictionary learning. Peng et al. [33] proposed a structured matrix decomposition model with tree-structured sparsity-inducing regularization and Laplacian regularization. Zhang et al. [34] learned deep uncertain convolutional features to improve the robustness and accuracy of salient object detection.

These bottom-up and top-down based methods have been successfully applied for proto-object detection. However, they perform poorly on low contrast images. It is a challenging task to acquire the effective features in low contrast images. Aiming at this problem, this paper proposed a CNN-based saliency model to detect the salient object in low contrast images, which has been verified efficient to find the visual saliency region under low contrast conditions. The saliency model can be applied to autofocus system, which obtains optimum autofocusing results in low contrast surveillance images.

## 3. Proposed saliency model

Since the focus objects are more likely existed in salient regions, the key point to achieve autofocus in low contrast surveillance image is to select an ideal salient region. The details of the proposed CNN-based local-global contrast and internal similarity driven salient object detection algorithms are presented in this section.

### 3.1. Low-level features extraction

For an input image, 28 low-level visual features (denoted as $\{fa(x,y)\}a = 1,\cdots,28$ are extracted to form a multi-dimensional feature vector $F(x,y) = [f1(x,y),\cdots,fa(x,y),\cdots,f28(x,y)]^T$. The extracted 28 features include: (1) four color features, which contain the strength feature obtained by averaging the three channels of RGB color space, the lightness feature obtained from LAB color space, the hue and saturation features obtained from HSV color space, (2) 12 steerable pyramid features obtained by filtering the

grayscale image at three scales on four directions, (3) 12 Gabor filter features obtained by performing the Gabor filter at 12 orientations. These 28 features are less affected by the image noise and image contrast, thus the effective information which can represent the properties of low contrast image can be well preserved.

### 3.2. Region covariance based CNN model

In the training phase, the covariance matrices of the randomly segmented blocks are computed to constitute the training sample. The training image is first divided into $8 \times 8$non-overlapping blocks $B(i)$, $i = 1, \cdots, n$. Each block can be represented as a $28 \times 28$ covariance matrix via [35]:

$$Ci = \frac{1}{num - 1} \sum_{b=1}^{num} (Fb(x, y) - \mu^*)(Fb(x, y) - \mu^*)^T, \tag{2}$$

where$\{Fb(x, y)\}b = 1, \cdots, num$ denote the 28-dimensional feature points inside $B(i)$ and $\mu^*$ is the mean value of these points. Because the region covariance has strong adaptability to rotation, scaling and brightness changes, and the extracted features can be nonlinearly fused by computing their covariance matrices, the optimal visual information of the low contrast image is largely obtained, which greatly increases the validity of the training sample and contributes a lot to the accuracy of the final saliency results.

Then, CNN is introduced to train the covariance matrix and test the saliency values of image blocks. The training samples are $28 \times 28$ covariance matrices of 10,000 blocks randomly selected from 100 training images. The corresponding 10 labels $\{0, 0.1, \cdots, 0.9\}$ of each matrix are computed as the proportion of salient pixels of each block in the ground-truth image. The proposed CNN framework is built based on the Palm's Deep Learning Toolbox [36]. Since the input layer of the proposed network structure is the covariance matrix of image block, the size of input layer only depends on the dimension of the extracted features, and it is irrelevant to the size of the training image and the block. This operation can greatly reduce the algorithm complexity of training.

For each $28 \times 28$ covariance matrix, which is the input layer, the first convolutional layer consists of 6 feature maps connected to it through 6 $5 \times 5$ kernels. The second layer has 6 subsampling layers by $2 \times 2$ average pooling operation. The third convolutional layer consists of 12 feature maps connected to the 6 average pooling layers through 72 $5 \times 5$ kernels. The fourth layer has 12 subsampling layers by the $2 \times 2$ average pooling operation. Then, 192 features of the fourth layer are concatenated into a feature vector which feeds into the final classification form by fully connection. The final layer consists of 10 output neurons which corresponding to 10 labels. By introducing the learning strategy, the low-level features are trained to become high-level features, which contain more semantic information of the salient objects. Since the trained deep model has powerful expressive ability, the proposed method has excellent performance in locating the salient object in low contrast images.

### 3.3. Local-global contrast approach

In the testing phase, the input image is first rescaled into the size of $256 \times 256$, then the image is segmented into multi-scale non-overlapping blocks, the sizes of these blocks are $8 \times 8$, $16 \times 16$, and $32 \times 32$, respectively. The multi-scale strategy can suppress the interference of background noise.

By extracting the mentioned 28 visual features, the original image can be abstracted into a 28-dimensional feature vector, each block can be represented by a $28 \times 28$ covariance matrix, which is the element of testing sample. Using the pre-trained deep CNN model, each covariance matrix can be marked with a saliency label, which is the saliency descriptor of the corresponding block $B(i)$.

The saliency value of $B(i)$ is measured by computing the weighted average saliency label differences between $B(i)$ and its surrounding neighborhoods $B(j)$ in a local-global form via:

$$SV(i) = \frac{1}{n} \sum_{j=1, j \neq i}^{n} \frac{|B(i) - B(j)|}{1 + |c(i) - c(j)|}, \tag{3}$$

where $|\cdot|$ denotes the Euclidean distance, $c(i)$ and $c(j)$ denote the spatial centers of $B(i)$ and $B(j)$, respectively. The local-global contrast information helps to find the autofocus areas which are most noticeable to the human eye.

### 3.4. Internal similarity measure

The inter-pixel similarity measure is also introduced to refine the resulting saliency map. Each block $B(i)$ is assigned to a pixel-level histogram $Hk(i)$, which is calculated based on the color quantization table with $m$ entries. The histogram is normalized to have $\sum_{k=1}^{m} Hk(i) = 1$. The internal similarity between two blocks $B(i)$ and $B(j)$ is obtained by:

$$S(i, j) = \frac{S_{color}(i, j)}{|c(i) - c(j)|}, \tag{4}$$

where the color similarity $Scolor(i, j)$ is computed as the sum of intersection between each histogram as:

$$Scolor(i, j) = \sum_{k=1}^{m} \min\{Hk(i), Hk(j)\}. \tag{5}$$

The final saliency value for each block is recalculated by exploiting the inter-pixel similarity measure, thus the blocks with higher similarity can possess more similar values.

$$SV'(i) = \frac{\sum_{j=1}^{n} S(i, j) \cdot SV(j)}{\sum_{j=1}^{n} S(i, j)}. \tag{6}$$

Since the saliency values of all the blocks are computed, the resulting saliency map $Smap$ can be obtained, then it is normalized in the interval [0, 1]. To further enhance the performance, the generated saliency map is smoothed by a median filter, which can better highlight the edges of the salient objects. The internal similarity strategy is beneficial to acquire the structure information to locate the autofocus regions.

After detecting the salient object, the focusing region can be selected accurately based on it. The performance evaluation of the saliency maps obtained by the proposed model is described in the next section.

## 4. Experimental results

A number of experiments were conducted to validate the performance of the proposed saliency model on four datasets: (1) the MSRA dataset [1], in which the principle salient objects are labeled by different human subjects, (2) the SED dataset [2], which provides the ground truth segmented by three human subjects, (3) the CSSD dataset [3], which is more challenging, including complex scenes, and (4) the NI dataset created by the proposed research, which contains plenty of low contrast images in the evening, the resolution of these various images is $640 \times 480$.

The proposed saliency model is compared with ten existing state-of-the-art saliency models including *non-parametric* (NP) model [20], *image signature* (IS) model [21], *context-aware* (CA) model [22], *patch distinction* (PD) model [23], *graph-based manifold ranking* (GBMR) model [24], *saliency optimization* (SO) model [25], *bootstrap learning* (BL) model [28], *generic promotion* (GP) model [26], *spatiochromatic context* (SC) model [27], and *structured matrix decomposition* (SMD) model [33].

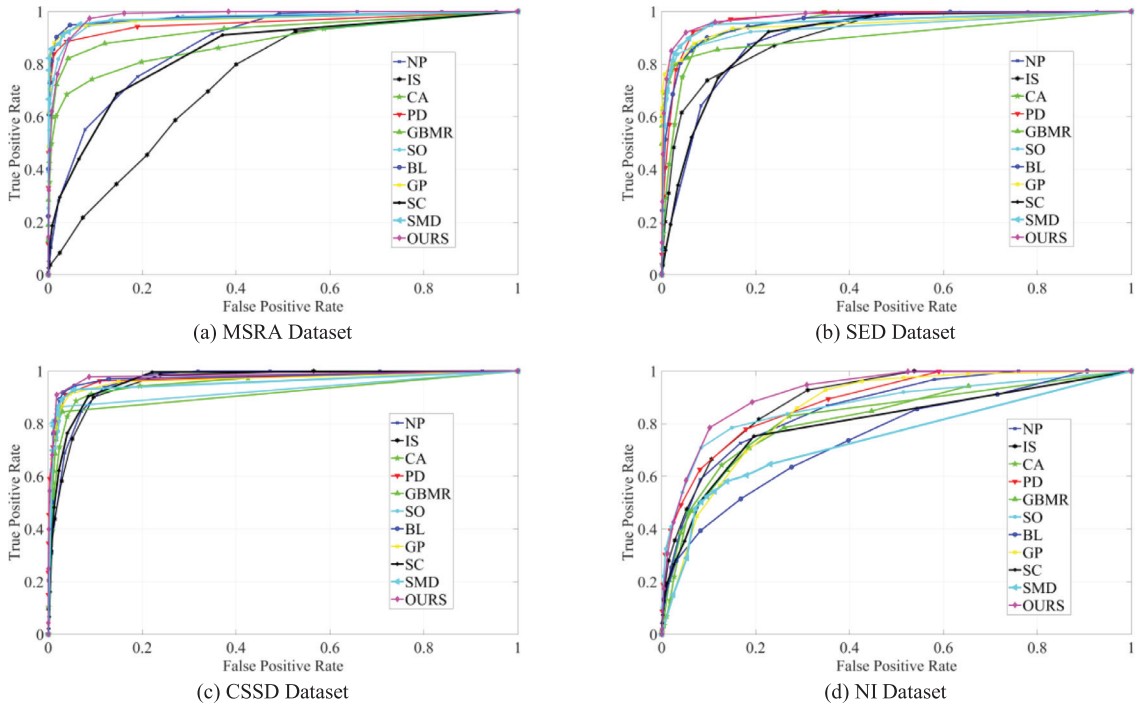



**Fig. 3.** The ROC performance comparisons of various saliency models on four datasets.

**Table 1**
The AUC performance of saliency maps from various saliency models on four datasets.

| | NP | IS | CA | PD | GBMR | SO | BL | GP | SC | SMD | OURS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| MSRA | 0.8760 | 0.7401 | 0.8805 | 0.9562 | 0.9311 | 0.9731 | 0.9772 | 0.9726 | 0.8458 | 0.9755 | **0.9843** |
| SED | 0.9022 | 0.9096 | 0.9543 | 0.9718 | 0.9130 | 0.9402 | 0.9604 | 0.9516 | 0.9043 | 0.9613 | **0.9801** |
| CSSD | 0.9630 | 0.9571 | 0.9526 | 0.9697 | 0.9126 | 0.9249 | 0.9717 | 0.9650 | 0.9648 | 0.9552 | **0.9827** |
| NI | 0.8597 | 0.8937 | 0.8242 | 0.8853 | 0.8202 | 0.8699 | 0.7514 | 0.8548 | 0.8022 | 0.7401 | **0.9316** |

In order to evaluate the performance of the proposed saliency model, the *receiver operating characteristic* (ROC) curve is introduced to test the accuracy of the generated saliency maps. The ROC curve is a two-dimensional graph which contains the *true positive rate* (TPR) and the *false positive rate* (FPR). Given the ground-truth binary mask (denoted as $GT(x, y)$) and the obtained saliency map $SMap(x, y)$ $(0 \le SMap(x, y) \le 1)$, a threshold $t$, $(0 \le t \le 1)$) is used to get the binary masks $Bt(x, y)$, in which 0 denotes the background and 1 denotes the salient objects. The TPR and FPR can be computed via:

$$\mathrm{TPR} = E\left(\prod_t Bt(x,y)\cdot GT(x,y)\right), \tag{7}$$

$$\mathrm{FPR} = E\left(\prod_t (1 - Bt(x,y))\cdot GT(x,y)\right). \tag{8}$$

By plotting the obtained TPRs and FPRs, the ROC curve is generated. The ROC performance comparisons of the ten models and the proposed model are shown in Fig. 3, which are tested on the MSRA, SED, CSSD and the NI datasets, respectively.

It can be seen from Fig. 3 that the proposed model has a better performance than other ten state-of-the-art saliency models on MSRA, SED, CSSD, and NI datasets, the overall performance will decline in the nighttime images which have a relatively low contrast.

The *area under the curve* (AUC) is calculated to give an intuitive comparison. The AUC can indicate how well the generated saliency map predicts the human interesting area. Table 1 shows the AUC value of the various saliency models on the four datasets. It can be observed that the proposed model has state-of-the-art performance on the mentioned four datasets.

For an objective comparison to quantitatively evaluate the performance of the detected salient object, the precision, recall criteria are then introduced, which compare the difference between binarized saliency map and the ground-truth mask. Let $R(\cdot)$ represents the salient region, given the ground-truth binary mask $GT$ and binary mask (denoted as $Bmask$) of the saliency map, the precision and recall can be calculated via:

$$precision = \frac{R(B_{mask} \cap GT)}{R(B_{mask})}, \tag{9}$$

$$recall = \frac{R(B_{mask} \cap GT)}{R(GT)}. \tag{10}$$

To further evaluate the accuracy of obtained binary mask of the saliency map, the F-measure is given by:

$$Fmeasure = \frac{(1+\beta^2)\, precision \times recall}{\beta^2 \times precision + recall}. \tag{11}$$

The proposed method uses $\beta^2 = 0.3$ to weigh the precision and recall. The comparisons of precision, recall, and $F$-measure of these various models are shown in Fig. 4.

As shown in Fig. 4, the $F$-measure value of the proposed method is relatively higher than the other ten models, which indicates an excellent performance to predict the human eye gaze. The recall rate of the various saliency models is not high on the nighttime image dataset, the possible reason is that the salient objects in NI dataset are too small, which results in a low $F$-measure performance.

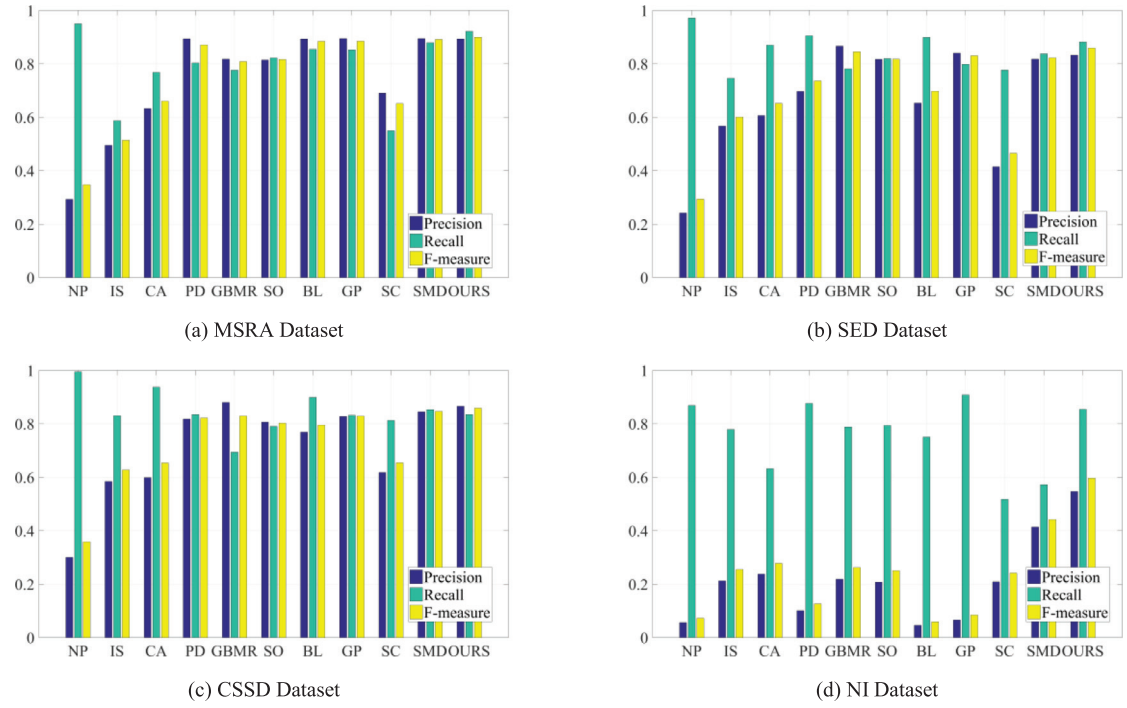



**Fig. 4.** The precision, recall, and F-measure performance comparisons of various saliency models on four datasets.

**Table 2**
The computational run-time (in second) of various saliency models on four datasets.

| | NP | IS | CA | PD | GBMR | SO | BL | GP | SC | SMD | OURS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| MSRA | 2.437 | 0.303 | 80.44 | 11.332 | 1.541 | 0.337 | 16.403 | 1.448 | 30.670 | 2.520 | 5.643 |
| SED | 1.457 | 0.201 | 72.654 | 5.450 | 0.938 | 0.438 | 43.236 | 1.230 | 30.330 | 2.708 | 3.093 |
| CSSD | 1.638 | 0.294 | 78.134 | 6.444 | 0.833 | 0.578 | 39.202 | 1.445 | 29.144 | 2.405 | 4.992 |
| NI | 5.340 | 0.903 | 111.35 | 19.024 | 4.832 | 2.124 | 103.13 | 8.031 | 38.751 | 12.748 | 7.546 |

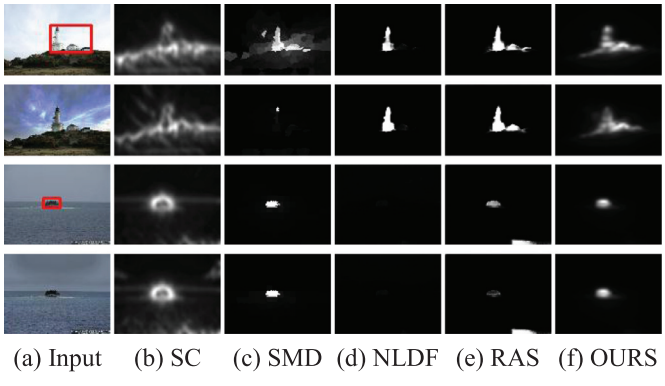



**Fig. 5.** Saliency performance comparison of various models by reducing the image contrast. (a) Input image of different contrasts, the first row and the third row are the original normal light images, the principle salient objects are circled by a red rectangle; the second row and the fourth row are images that are used for comparison, and their contrasts are reduced compared to the original images. (b–c) Traditional low-level feature based models. (d–e) High-level feature based deep learning models. (f) The proposed models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

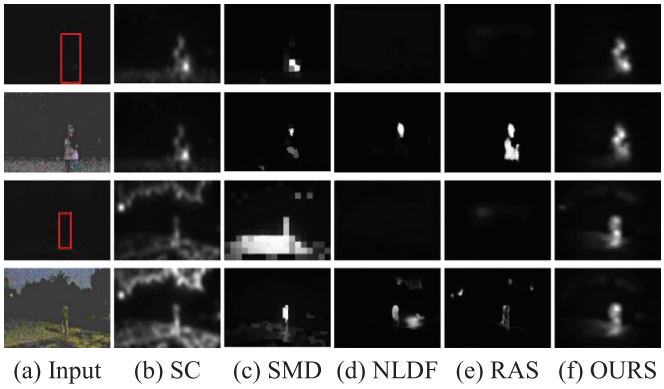



**Fig. 6.** Saliency performance comparison of various models by increasing the image contrast. (a) Input image of different contrasts, the first row and the third row are the original nighttime images, the principle salient objects are circled by a red rectangle; the second row and the fourth row are images that are used for comparison, and their contrasts are enhanced compared to the original images. (b–c) Traditional low-level feature based models. (d–e) High-level feature based deep learning models. (f) The proposed models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The run-time performance is also considered to evaluate the efficiency of various algorithms. The experiment is measured on a standard personal computer with 1.6GHZ CPU and 8GB RAM. All approaches use Matlab implementations. It can be observed from Table 2 that the run-time of IS method is time-saving, but can only generate the low-resolution saliency maps. The computational complexity of the proposed method is slightly higher than the superpixel-based method GBMR, whereas our method can get more accurate estimations.

To illustrate the impact of the image contrast on the proposed and other methods, two comparative experiments are presented in Figs. 5 and 6 to demonstrate the effectiveness of various

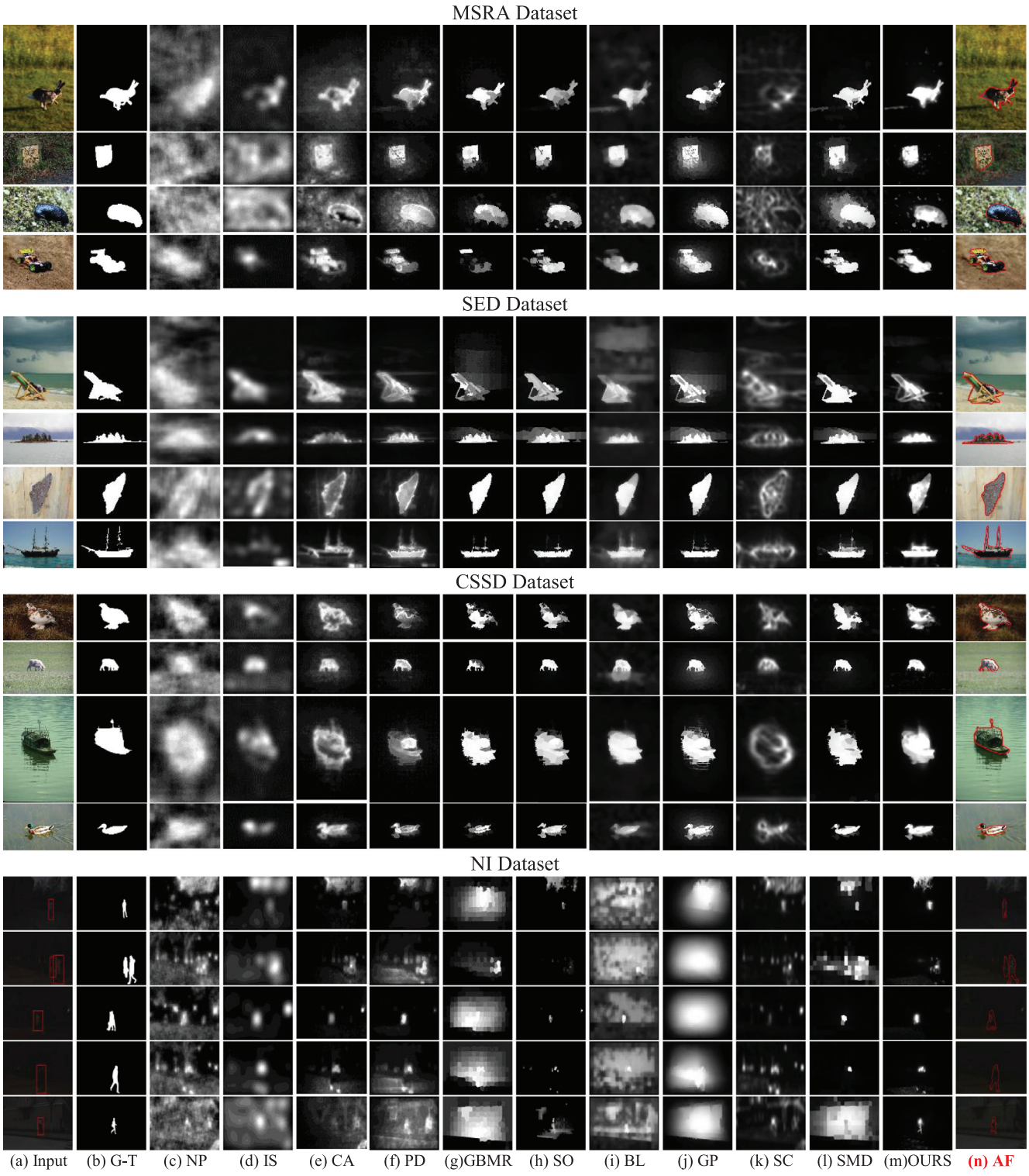



**Fig. 7.** Qualitative comparisons on MSRA, SED, CSSD, and NI datasets. (a) testing low contrast images; (b) ground-truth binary masks; (c-l) saliency maps obtained by various state-of-the-art saliency models; (m) saliency maps obtained by the proposed model; (n) autofocus images based on the salient objects.

models under different contrast conditions. We configure the experiments from two questions. (1) Whether reducing the contrast of normal light image affects the performance of various models, shown in Fig. 5. (2) Whether increasing the contrast of night images can improve the performance of various models, shown in Fig. 6. The contrast of the input image is changed by subtracting light and adding light respectively. In addition, several neural network models are also considered for comparison. The results of the proposed model are mainly compared with four state-of-the-art saliency models, including two traditional low-level feature based models (SC [27] and SMD [33]) and two high-level feature based deep learning models (NLDF [37] and RAS [38]).

(1) The effect of decreasing image contrast on the model performance is shown in Fig. 5.

It can be seen from Fig. 5 that the traditional low-level feature based models (SC and SMD) are susceptible to the interference of complex backgrounds. Relatively, the high-level feature based deep learning models (NLDF and RAS) are more robust to identify the real salient objects. When the contrast of the normal light image is reduced, the difference between the foreground and the background becomes smaller. As a result, the traditional models detect more background regions, and the deep models have difficult to determine the right salient objects. Whereas, the proposed model, which incorporates the low-level feature with the high-level feature, has consistent good performance on these images of different contrasts.

(2) The effect of increasing image contrast on the model performance is shown in Fig. 6.

It can be seen from Fig. 6 that both the traditional low-level feature based models (SC and SMD) and the high-level feature based deep learning models (NLDF and RAS) have poor performance on the low contrast nighttime images. When the contrast of the nighttime image is enhanced, the visibility of the foreground is increased as well as the background. As a result, these models can partially detect the correct salient objects. However, the contrast of the background region is also improved, which misleads these models to find the true salient objects, thus affecting the accuracy of the models. Unlike these models, the proposed model utilizes the covariance based convolutional neural network to guide the low-level features to train the high-level features, which is robust under different contrast conditions. Meanwhile, the local-global contrast approach and the inter-pixel similarity measure make our model less susceptible to background interference. Therefore, the change of image contrast has little effect on the proposed model.

The subjective comparisons are shown in Fig. 7. From Fig. 7, the saliency maps obtained by the GBMR, SO, BL, GP, SMD and the proposed models have uniform salient regions, the saliency objects of our model are more similar with the ground-truth binary masks. The saliency maps of NP and IS models can't clearly distinguish the salient region from their surroundings. The CA, PD and SC models have good detection effects, but the detected salient objects are not uniform, and their time consumption are very high. It is also evident that our model can better detect the salient objects in low contrast images, and is more effective than the other saliency models, these models cannot correctly detect the real salient objects under the conditions of low contrast background.

The autofocus images are shown in Fig. 7(n), in which the focusing regions are circled by red line. Since the salient objects can be detected accurately by the proposed model, it can help to select a more optimal focus region in low contrast images.

## 5. Conclusions

In this paper, an effective CNN-based saliency model is proposed to select the focusing region in low contrast surveillance image, which is based on covariance matrix to train the deep learning model. The autofocus region is refined by the local-global contrast and inter-pixel similarity. Experiments have been carried out on the public available MSRA, SED, CSSD datasets and our nighttime image dataset for salient object detection. Results show that the proposed method outperforms the ten state-of-the-art saliency models. Most of the existing saliency computational methods fail to perform well on low contrast images, while the proposed approach has excellent performance on this task. Based on the results of salient object detection, the proposed algorithm can more quickly and accurately detect the most important autofocusing areas than conventional auto-focusing methods especially in low contrast images, which has great application value in surveillance system.

## Acknowledgment

This work was supported by the N atural Science Foundation of China (61602349, U1803262, 61440016 and 61273225).

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.patrec.2019.04.011.

## References

[1] T. Liu, J. Sun, N.-N. Zheng, X. Tang, H.-Y. Shum, Learning to detect a salient object, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[2] S. Alpert, M. Galun, R. Basri, A. Brandt, Image segmentation by probabilistic bottom-up aggregation and cue integration, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[3] Q. Yan, L. Xu, J. Shi, J. Jia, Hierarchical saliency detection, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 1155–1162.
[4] L. Wu, Y. Wang, J. Gao, X. Li, Deep adaptive feature embedding with local sample distributions for person re-identification, Pattern Recognit. 73 (2018) 275–288.
[5] L. Wu, Y. Wang, X. Li, J. Gao, What-and-where to match: deep spatially multiplicative integration networks for person re-identification, Pattern Recognit. 76 (2018) 727–738.
[6] L. Wu, Y. Wang, J. Gao, X. Li, Where-and-when to look: deep siamese attention networks for video-based person re-identification, IEEE Trans. Multimedia (2018) 1–13.
[7] L. Wu, Y. Wang, L. Shao, M. Wang, 3-D PersonVLAD: learning deep global representations for video-based person reidentification, IEEE Trans. Neural Netw. Learn. Syst. (2019) 1–13.
[8] Y. Wang, X. Lin, L. Wu, W. Zhang, Effective multi-query expansions: collaborative deep networks for robust landmark retrieval, IEEE Trans. Image Process. 26 (3) (2017) 1393–1404.
[9] L. Wu, Y. Wang, L. Shao, Cycle-consistent deep generative hashing for cross–modal retrieval, IEEE Trans. Image Process. 28 (4) (2019) 1602–1612.
[10] Y. Wang, X. Lin, Lin Wu, W. Zhang, Q. Zhang, X. Huang, Robust subspace clustering for multi-view data by exploiting correlation consensus, IEEE Trans. Image Process. 24 (11) (2015) 3939–3949.
[11] Y. Wang, L. Wu, X. Lin, J. Gao, Multiview spectral clustering via structured low-rank matrix factorization, IEEE Trans. Neural Netw. Learn. Syst. 29 (10) (2018) 4833–4843.
[12] L. Wu, Y. Wang, X. Li, J. Gao, Deep attention-based spatially recursive networks for fine-grained visual recognition, IEEE Trans. Cybernetics (2018) 1–12.
[13] J.-S. Lee, Y.-Y. Jung, B.-S. Kim, S.-J. Ko, An advanced video camera system with robust AF, AE, and AWB control, IEEE Trans. Consumer Electron. 47 (3) (2002) 694–699.
[14] Y. Sun, S. Duthaler, B.J. Nelson, Auto focusing algorithm selection in computer microscopy, in: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005, pp. 70–76.
[15] J. He, R. Zhou, Z. Hong, Modified fast climbing search auto-focus algorithm with adaptive step size searching technique for digital camera, IEEE Trans. Consum. Electron. 49 (2) (2003) 257–262.
[16] C.H. Chin, Y. Zhang, C. Wen, Visual servo auto-focusing using recursive weighted least-squares for machine vision inspection, in: Proceedings of IEEE Conference on Electronics Packaging Technology, 2003, pp. 777–780.
[17] K. Zhu, W. Jiang, D. Wang, X. Zhou, J. Zhang, An effective focusing algorithm based on non-uniform sampling, in: Proceedings of IEEE International Workshop on VLSI Design and Video Technology, 2005, pp. 276–279.
[18] F. Meglio, G. Panariello, G. Schirinzi, Three dimensional SAR image focusing from non-uniform samples, in: Proceedings of IEEE International Geoscience and Remote Sensing Symposium, 2007, pp. 528–531.
[19] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1254–1259.
[20] N. Murray, M. Vanrell, X. Otazu, C.A. Parraga, Saliency estimation using a non–parametric low-level vision model, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 433–440.
[21] X. Hou, J. Harel, C. Koch, Image Signature, Highlighting sparse salient regions, IEEE Trans. Pattern Anal. Mach. Intell. 34 (1) (2012) 194–201.
[22] S. Goferman, L. Zelnik-Manor, A. Tal, Context-aware saliency detection, IEEE Trans. Pattern Anal. Mach. Intell. 34 (10) (2012) 1915–1926.
[23] R. Margolin, A. Tal, L. Zelnik-Manor, What makes a patch distinct? in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 1139–1146.
[24] C. Yang, L. Zhang, H. Lu, X. Ruan, M.-H. Yang, Saliency detection via graph-based manifold ranking, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2013 3166–3137.
[25] W. Zhu, S. Liang, Y. Wei, J. Sun, Saliency optimization from robust background detection, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2814–2821.

[26] P. Jiang, N. Vasconcelos, J. Peng, Generic promotion of diffusion-based salient object detection, in: Proceedings of IEEE International Conference on Computer Vision, 2015, pp. 217–225.
[27] J. Zhang, M. Wang, S. Zhang, X. Li, X. Wu, Spatiochromatic context modeling for color saliency analysis, IEEE Trans. Neural Netw. Learn. Syst. 27 (6) (2016) 1177–1189.
[28] N. Tong, H. Lu, M. Yang, Salient object detection via bootstrap learning, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1884–1892.
[29] X. Xu, N. Mu, H. Zhang, X. Fu, Salient object detection from distinctive features in low contrast images, in: Proceedings of IEEE International Conference on Image Processing, 2015, pp. 3126–3130.
[30] S. He, R.W.H. Lau, Exemplar-driven top-down saliency detection via deep association, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 5723–5732.
[31] G. Li, Y. Yu, Visual saliency detection based on multiscale deep CNN features, IEEE Trans. Image Process. 25 (11) (2016) 5012–5024.
[32] J. Yang, M.-H. Yang, Top-down visual saliency via joint CRF and dictionary learning, IEEE Trans. Pattern Anal. Mach. Intell. 39 (3) (2017) 576–588.
[33] H. Peng, B. Li, H. Ling, W. Hu, W. Xiong, S.J. Maybank, Salient object detection via structured matrix decomposition, IEEE Trans. Pattern Anal. Mach. Intell. 39 (4) (2017) 818–832.
[34] P. Zhang, D. Wang, H. Lu, H. Wang, B. Yin, Learning uncertain convolutional features for accurate saliency detection, in: Proceedings of IEEE International Conference on Computer Vision, 2017, pp. 212–221.
[35] O. Tuzel, F. Porikli, P. Meer, Region covariance: a fast descriptor for detection and classification, in: Proceedings of European Conference on Computer Vision, 2006, pp. 589–600.
[36] R.B. Palm, Prediction as a candidate for learning deep hierarchical models of data *M.S. thesis*, DTU Informatics, Technical University of Denmark, Lyngby, Denmark, 2012.
[37] Z. Luo, A. Mishra, A. Achkar, J. Eichel, S. Li, P.-M. Jodoin, Non-local deep features for salient object detection., in: Proceedings of the Conference on Computer Vision and Pattern Recognition, 2017, pp. 6593–6601.
[38] S. Chen, X. Tan, B. Wang, X. Hu, Reverse attention for salient object detection, in: Proceedings of the European Conference on Computer Vision, 2018, pp. 236–252.