SalNet: Edge Constraint Based End-to-End Model for Salient Object Detection

Le Han[1,2], Xuelong Li[1], and Yongsheng Dong[1(✉)]

[1] Center for OPTical IMagery Analysis and Learning (OPTIMAL), Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, People's Republic of China
hanle2016@opt.cn, xuelong_li@opt.ac.cn, dongyongsheng98@163.com
[2] University of Chinese Academy of Sciences, 19A Yuquanlu, Beijing 100049, People's Republic of China

Abstract. Salient object detection is a fundamental task in computer vision and pattern recognition. And it has been investigated by many researchers in many fields for a long time. Numerous salient object detection models based on deep learning have been designed in recent years. However, the saliency maps extracted by most of the existing models are blurry or have irregular edges. To alleviate these problems, we propose a novel approach named SalNet to detect the salient objects accurately in this paper. The architecture of the SalNet is an U-Net which can combine the features of the shallow and deep layers. Moreover, a new objective function based on the image convolution is further proposed to refine the edges of saliency maps by using a constraint on the L1 distance between edge information of the ground-truth and the saliency maps. Finally, we evaluate our proposed SalNet on benchmark datasets and compare it with the state-of-the-art algorithms. Experimental results demonstrate that the SalNet is effective and outperforms several representative methods in salient object detection task.

Keywords: Salient object detection · U-Net · Auto-encoder
Image convolution

1 Introduction

Human can automatically pay attention to the *region of interest* (ROI) and selectively ignore the uninterested region when they face a scene. The salient objects are just the contents of the ROI. For instance, when we are looking at a picture with a horse grazing on the hillside, we may concentrate on the horse because it is the visual salient object in this image. The task of salient object detection is to teach computers to identify and extract the salient objects of the input images as humans do. Visual saliency has been explored by numbers of

J.-H. Lai et al. (Eds.): PRCV 2018, LNCS 11259, pp. 186–198, 2018.
https://doi.org/10.1007/978-3-030-03341-5_16

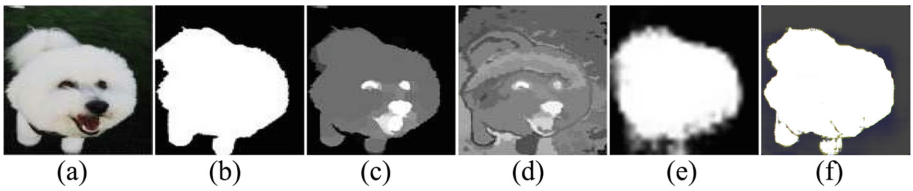

Fig. 1. Comparison of different kinds of models. (a) is the input image and the Ground-truth is (b). The saliency maps output by the background prior based model, global contrast based model and the convolution neural network based model are (c), (d) and (e), respectively. (f) is the output of the SalNet, which accurately extracted the salient objects and the edge of the object is very clear.

scholars from multiple disciplines such as neuroscience [10], cognitive psychology [7] and computer vision [4]. As a fundamental problem in computer visual, saliency detection has been successfully applied to many areas including object tracking and recognition, semantic recognition, video retrieval and scene classification [28]. Therefore, it has attracted the interest of many scholars. Numerous salient object detection models have been proposed, including the conventional models [11,17] and deep learning based models [15,22,28].

The conventional saliency detection models mainly focus on the heuristic priors or the contrast information [18]. Some previous methods utilize various visual informative knowledge as heuristic priors, i.e., the background prior, compactness prior and objectness prior. Background prior based models [6] assume that the area near the boundary of image is probably the background. It is easy to see that they usually have low accuracy if the salient objects are close to the edge of image. In compactness prior based models [27], the salient object region is considered as a connected area with perceptually homogeneous elements. However, the disadvantage of this kind of models is that they cannot detect the images with multiple salient objects very well. The objectness prior based models [27] tend to focus on the regions that are likely to contain salient objects. They are usually empirical and rely on hand-designed formulations. Due to the high-level semantic features are not considered, the objects of the saliency maps output by these prior based models are usually uneven like a heat map. The Fig. 1(c) is the output of the heuristic prior based model, in which the brightness of the different parts of the object is not the same.

The contrast based algorithms aim at investigating the difference between the image pixels or regions and the context [13], and they can be divided into the global and local context based models. Global context based models usually output a saliency map with relatively complete internal information but incomplete details. Besides, it is difficult for them to detect the salient objects with large sizes. On the contrary, local context based models can capture the detailed structures but often lose internal information. The saliency maps output by the contrast based models are usually fuzzy and the edges of the objects are not clear as presented in Fig. 1(d). The reason is that they process the images pixel

by pixel or region by region, so the relationship of adjacent pixels or regions is not considered.

The conventional saliency detection models mentioned above just leverage the low-level visual features, such as the color, contrast and various heuristic priors. The high-level information which is about the semantic knowledge has not been taken into account. So these models can hardly distinguish salient objects from the images with complex background and can not output a precise saliency map with clear edges of the objects [26].

To mine the high-level semantic information, deep learning is widely used in salient object detection task in recent years. And they delivered superior performance because the convolutional neural network can hierarchically capture the features of images. The auto-encoder is one of the best architectures of the saliency detection [18]. However, the information loss is serious when the input flows pass through the network. The reason is that the features of shallow layers and deep layers cannot be effectively combined due to the existence of a bottleneck between encoder and decoder. As a result, the saliency maps extracted by the models based on convolution neural network often have fuzzy objects which is illustrated in Fig. 1(e).

In order to alleviate these problems, we should make full use of high-level semantic information and low-level visual information. In this paper, we propose a new method named SalNet for salient object detection. The architecture of the proposed model is the U-Net which can combine multi-level features and avoid information loss [20]. Besides, in order to enrich the edge information extracted by network, we add a convolution-based edge constraint to loss function. The experimental results are presented in Fig. 1 to verify the conclusion discussed above. It can be found that our SalNet can effectively alleviate the uneven of saliency maps and blurring of objects edges, which are usually generated by existing models.

The main contributions of our work are as follows:

- We propose a new loss function to refine the objects details of saliency maps. In order to fully exploit edge information, we add a convolution based edge constraint term to loss function.
- We employ the architecture of U-Net to detect the salient objects of images accurately. Such an architecture can reduce the loss of low-level visual information, which is necessary because the saliency detection task requires the objects of saliency map and input image to be consistent on the visual information, such as shape and edge.
- The proposed SalNet is tested on four benchmark datasets (ECSSD, HKU-IS, SED1 and SED2). And the experimental results show that our saliency detection model outperforms the state-of-the-art methods on benchmark datasets.

The rest of this paper is organized as follows. In Sect. 2, some previous works are briefly introduced. Section 3 presents our model for salient object detection. To validate the proposed method, the experimental results are shown in Sect. 4. At last, Sect. 5 makes a brief conclusion for this paper.

2 Related Work

In recent years, deep learning achieved superior performance in numerous fields in computer vision, including semantic segmentation, image classification, object recognition and so on. *Convolutional Neural Network* (CNN) can fully mine the deep semantic features of images. Therefore, it is widely used in salient object detection tasks.

Zhang *et.al.* proposed a model based on the CNN and the *Maximum a Posteriori* (MAP) principle [25], which has no constraint to the number of objects in image and outputs the bounding boxes of the objects. Wang *et.al.* dealt with the saliency detection with a recurrent architecture combined with prior knowledge [22]. However, the models mentioned above only output the bounding boxes of the objects or rely on certain prior knowledge.

In order to output the saliency maps and not rely on various prior information, Liu *et al.* proposed an end-to-end salient object detection model [18]. They first utilized a convolution neural network to learn a global feature representation, and then hierarchically refined the details of saliency maps with a network named HRCNN. Zhao *et al.* utilized two convolutional neural networks to extract global context and local context information respectively, and then combined these information with a fully-connected layer to predict the saliency map of input image [28]. The *Fully Convolutional Networks* (FCN) [19], which is proposed by Long *et al.* for semantic segmentation, is widely used in the salient object detection tasks and performs well. Zhang *et al.* proposed a deep fully convolutional network based model named UCF [26], which increases the robustness and accuracy of saliency detection by learning deep uncertain convolutional features. The architecture of UCF consists of encoder and decoder. And the auto-encoder-based models are currently the best [29].

The reason why these models work well is that the convolution neural networks can capture relevant information from the convolution layers features. However, the disadvantage of these models is that their saliency maps are relatively blurred, since these salient object detection models don't aggregate the multi-level convolutional feature maps. Besides, there is a bottleneck between encoder and decoder which leads to information loss. Therefore, we adapt the U-Net as the architecture of proposed method due to the fact that it can decrease information loss by adding skip connections between corresponding layers of encoder and decoder. Experiments indicate that the SalNet performs very well in salient object detection task.

3 Proposed Method

In this section, we describe the proposed method named **SalNet**. We first present the model architecture, followed by proposed loss function. The details of the training procedure and inference are given in the last subsection.

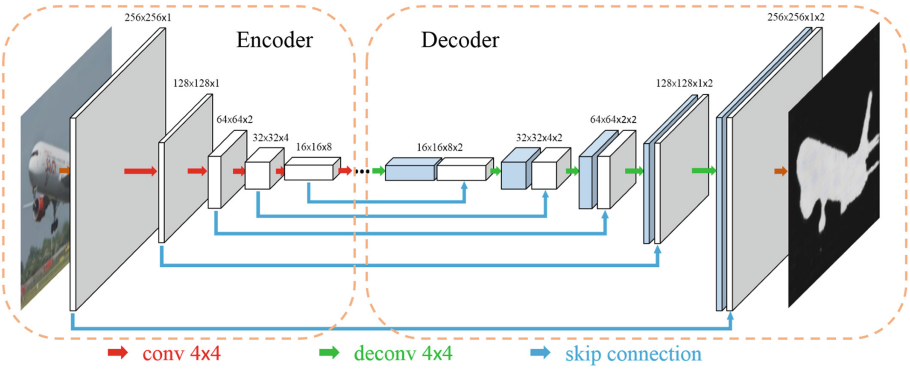

Fig. 2. The architecture we employed in the SalNet.

3.1 The Network Architecture

Many salient object detection models [22,26] use auto-encoder architecture [1,8]. The encoder is composed by a group of downsampling layers, and the size of images decreases progressively during downsampling. On the contrary, the decoder gradually restores the output of layers to input size. The disadvantage of encoder-decoder network is that the information loss is serious when image flows pass through the network. The reason is that there is a bottleneck between the encoder and decoder [9].

We leverage an U-net [20] which is an improvement of encoder-decoder network as the architecture of SalNet. The U-net adds skip connections between corresponding layers of encoder and decoder. Such a structure can reduce low-level visual information loss when image flows through the network [24], which is necessary because the salient object detection problem requires input and saliency maps to be consistent on the shape and edge of objects.

The architecture of the SalNet is illustrated in Fig. 2. The encoder is composed of 8 convolution layers of which the kernel size is 4×4 and stride is 2. The output size of encoder is $1 \times 1 \times 8$. The decoder is composed of 8 deconvolution layers with the kernel size is 4×4. In encoder, the white box of each layer is downsampled from the output of previous layer. And it is upsampled from the output of previous layer in decoder. The blue boxes of decoder are the copied feature maps of the corresponding layers in encoder.

3.2 Loss Function

In order to enrich the detail features of the salient objects extracted by network, we propose a new loss function which can be formulated as

$$Loss = \lambda_1 \mathcal{L}_{L_1}(x, y) + \lambda_2 \mathcal{L}_C(x, y) + \lambda_3 \mathcal{L}_{Conv}(x, y), \tag{1}$$

where the λ_1, λ_2 and λ_3 are the weights of L1 loss, cross entropy term and convolution loss term respectively. The x is the saliency map output by network, and y is ground-truth.

The $\mathcal{L}_{L_1}(x, y)$ is the L1 loss which is presented in Eq. 2. The L1 distance is wildly used to train neural networks. However, the network will produce blurry output if it only rely on L1 loss. So we added additional constraints to capture the high-frequency information.

$$\mathcal{L}_{L_1}(x, y) = \|y - x\|_1. \tag{2}$$

The second term is the cross entropy loss, and the definition of it is

$$\mathcal{L}_C(x, y) = -\frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} [y(i, j) \times p(i, j) + (1 - y(i, j)) ln(1 - p(i, j))], \tag{3}$$

where M and N are the length and width of image respectively, and the $p(i, j)$ is defined as follows:

$$p(i, j) = sigmoid[x(i, j)] = \frac{1}{1 + e^{-x(i,j)}}. \tag{4}$$

It performs a sigmoid calculation on the output saliency map and then calculates its cross entropy with ground-truth. It has optimized the calculation method of cross entropy so that the result will not overflow.
The last term is the convolution loss term, which can be formulated as

$$\mathcal{L}_{Conv}(x, y) = \|Conv(x) - Conv(y)\|_1. \tag{5}$$

As illustrated in Fig. 3, the value in one diagonal of the kernel is 1 and -1 in the other, and the value of the rest, including the center point, of the kernel are 0. We tested many different sizes of the convolution kernel, and some of the results are shown in Table 1. When the size is 21×21, the test results of the trained model are the best. So we set the convolution kernel size to 21×21. Such convolution kernel with a large size can fuse the object edge with information of its neighborhood, rather than just extracting the boundary of the objects as the kernel with a small size do.

3.3 Training Procedure and Inference

The optimization algorithms we used in this paper are the mini-batch *Stochastic Gradient Descent* (SGD) and the Adam solver [12], which can calculate different adaptive learning rates for different parameters. And the outputs of our SalNet can be seen from the Fig. 4.

During the inference phase, we run the network in the same way as the training time. Besides, at test phase, we apply dropout and batch normalization which uses the statistic information of the test batch. And the batch size we used in the experiments is 4.

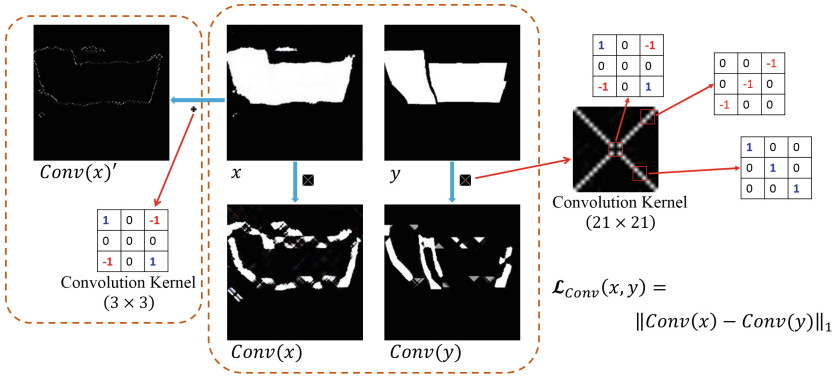

Fig. 3. The graphical diagram of the convolution kernel and comparison of results of the convolution kernels with different size. x and y represent the output of model and ground-truth, respectively. The $Conv(\cdot)$ is convolution with a 21×21 kernel, and the kernel size of $Conv(x)'$ is 3×3. The convolution operation only extracts objects edges because of too small kernel size of the convolution. Thus the relative information of objects edges and nearby regions is ignored.

4 Experiments

In this section, we describe our experimental setting and quantitative results which validate the effectiveness of our model for salient object detection.

4.1 Datasets

To show the effectiveness of the proposed method, we test it on several widely used salient object benchmark datasets, including ECSSD [23], HKU-IS [28], SED1 and SED2 [2]. ECSSD contains 1,000 semantically meaningful images with complex structure. HKU-IS has 4,447 high quality images. Many images in this dataset contain multiple salient objects, and the salient objects in many images touch the image bounding. The SED is composed of two subsets named SED1 and SED2, and both of them contain 100 images. Each image in SED1 contains one salient object, while there are two in each image in SED2.

4.2 Evaluation Metrics

In order to measure the performance of the different algorithms for salient object detection, we used three objective metrics in this paper, including the *Precision-Recall* (PR) curves, F-measure and *Mean Absolute Error* (MAE) [3].

The PR curve is based on the overlapping area between the ground-truth and estimated saliency map. We can divide the binary mask M, which is binarized by saliency map S with different thresholds, into *True Positive* (TP), *False Positive* (FP), *True Negative* (TN) and *False Negative* (FN), according

Fig. 4. Examples of our SalNet output results: the images in the first row are the inputs, the saliency masks in the second are the outputs of our method, and these in the last are the ground-truth saliency masks. It is not hard to see that the edges of the objects extracted by our model are very clear. The first column of images show that the proposed model can accurately extract the salient objects even if their colors are close to the background. The images in the second and third column indicate that the model can preserve the details and subtle structures of the object very well, and the last two columns indicate that our model can precisely detect the multiple salient objects.

to whether the $M(x, y)$ is equal to $G(x, y)$. And the precision and recall can be calculated by

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FP}. \tag{6}$$

The F-measure is a weighted harmonic mean of average precision and average recall with a non-negative weight β and can be calculated by

$$F_\beta = \frac{(1 + \beta^2) \times Precision \times Recall}{\beta^2 \times Precision + Recall}. \tag{7}$$

As suggested by existing works [3,5,18], β^2 is set to 0.3 because the precision is more important than recall. And we set the adaptive threshold T to be twice of the mean saliency value of.

The defect of the overlap-based evaluation measures is that they don't consider the true negative saliency assignments [3], so they usually give a higher score to the models which can classify the saliency areas correctly. In order to evaluate the models comprehensively, we calculate the *Mean Absolute Error* (MAE) which assess the saliency detection accuracy. The MAE can be calculated by

$$MAE = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} |S(x, y) - G(x, y)|, \tag{8}$$

in which G is the ground-truth.

Table 1. The comparison of the F-Measure and MAE results of different salient object detection models on four benchmark datasets. The models named SalNet-Kernel55 and SalNet-Kernel5151 are our models with the kernel size of convolution loss equal 5×5 and 51×51, respectively. The values in red, blue and green are the best three results of each term, respectively. The results of the proposed SalNet are almost always the best on these datasets.

	ECSSD		HKU-IS		SED1		SED2	
	F-Measure	MAE	F-Measure	MAE	F-Measure	MAE	F-Measure	MAE
DS	0.8255	0.1216	0.7851	0.0780	0.8445	0.0931	0.7541	0.1233
LEGS	0.7853	0.1180	0.7228	0.1193	0.8542	0.1034	0.7358	0.1236
UCF	0.8517	0.0689	0.8232	0.0620	0.8647	0.0631	0.8102	0.0680
ELD	0.8102	0.0796	0.7694	0.0741	0.8715	0.0670	0.7591	0.1028
DRFI	0.7331	0.1642	0.7218	0.1445	0.8068	0.1480	0.7341	0.1334
DSR	0.6621	0.1784	0.6772	0.1422	0.7909	0.1579	0.7116	0.1406
SalNet-Kernel55	0.8076	0.1013	0.8502	0.0647	0.8582	0.0827	0.8589	0.0747
SalNet-Kernel5151	0.8459	0.0819	0.8741	0.0530	0.8902	0.0665	0.8739	0.0718
SalNet	0.8468	0.0650	0.8660	0.0546	0.8918	0.0658	0.8892	0.0632

4.3 Performance Comparison with State-of-the-Art

We compare the proposed SalNet with the state-of-the-art algorithms including UCF [26], ELD [14], DS [16], LEGS [21], DRFI [11], DSR [17]. The UCF, ELD, DS and LEGS are the deep learning based algorithms, and the DRFI and DSR are conventional algorithms. The model of us is trained with 2,964 images selected randomly from the HKU-IS dataset. For fair comparison, the parameter settings of the models is the recommended by the authors, and we use some results shown in the original papers of the corresponding algorithm or the benchmark evaluation.

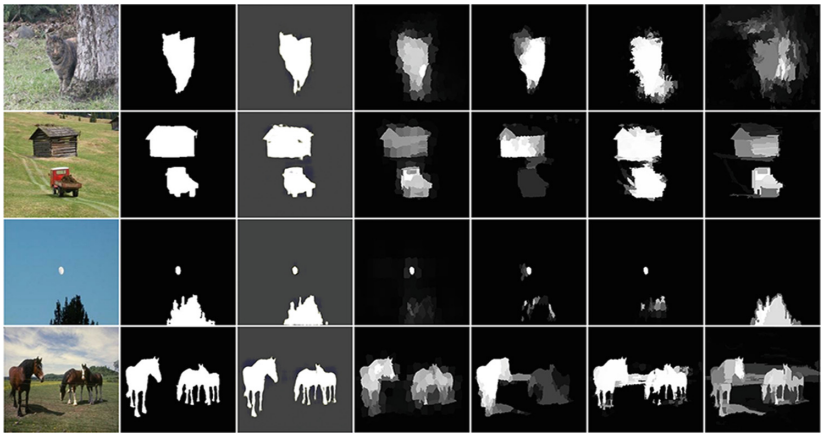

Fig. 5. Comparison with State-of-the-art models. (a) Input images; (b) Ground truth; (c) Our SalNet; (d) DS; (e) LEGS; (f) ELD; (g) DRFI.

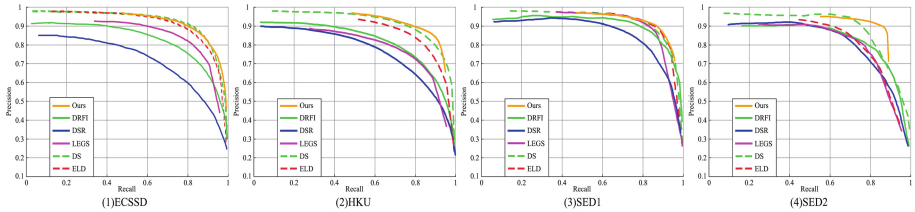

Fig. 6. The PR curves of our SalNet and the state-of-the-arts on 4 benchmark datasets

Figure 5 presents a visual comparison of the state-of-the-art models and the SalNet. As we can see from it that our saliency detection model can detect and localize the salient objects accurately. It preserves the object details which are ignored by other models. The edges of the objects detected by the SalNet is very clear and the saliency mask has a very uniform brightness, which is useful in the applications of it.

The F-measure and MAE of the proposed method and the other models are presented in Table 1, and the PR curves are illustrated in Fig. 6. We can see from them that the proposed SalNet is superior to other state-of-the-art models in all evaluation metrics across the benchmark datasets. The comparison results indicate that the SalNet can provide more accurate saliency maps with clear objects edges. In addition, we can learn from the last three rows of Table 1 that the results are not optimal when the kernel size of convolution loss term is too large or too small.

4.4 Ablation Studies

In order to verify the contributions of different components in our loss function, we evaluate three variants of the loss function and the results are illustrated in Fig. 7(b)–(d). The architecture and parameters setting of the comparison methods and the SalNet are the same. It not hard to see from (b) that the edges of the objects are quite blurred without the convolution loss.

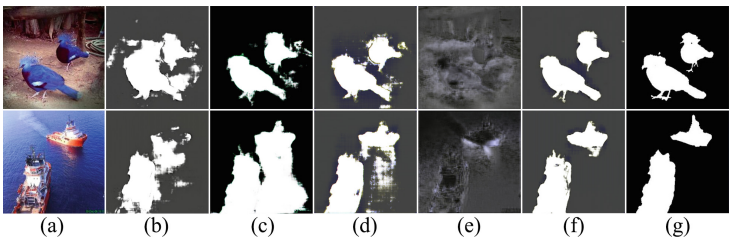

Fig. 7. The results of ablation studies. (a) Input images; (b) The output of the model with a loss function: $\mathcal{L}_{L_1}(x,y) + \mathcal{L}_C(x,y)$; (c) $\mathcal{L}_{L_1}(x,y) + \mathcal{L}_{Conv}(x,y)$; (d) $\mathcal{L}_C(x,y) + \mathcal{L}_{Conv}(x,y)$; (e) The output of the encoder-decoder model created by severing the skip connections of SalNet; (f) The output of the SalNet; (g) Ground-truth.

In order to verify whether the skip connections, which allow the low-level information combined with the features of deep layers, are useful in the SalNet, we train an encoder-decoder model which is created by deleting the skip connections of SalNet. The parameter setting and objective function are the same as the proposed method. The test results are presented in Fig. 7(e), which indicate that the skip connections can improve the detection performance of the model.

5 Conclusion

In this paper, we propose a novel end-to-end salient object detection method based on the U-Net, which can reduce the information loss by combining the features of different layers. In order to fully capture features related with the details of objects, we add a convolution based edge constraint term to the loss function. Extensive experiments demonstrate that the proposed SalNet outperforms the other state-of-the-art methods on four benchmark datasets.

Acknowledgements. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1107400, in part by the National Natural Science Foundation of China under Grants 61871470, 61761130079 and U1604153, and in part by the Program for Science and Technology Innovation Talents in Universities of Henan Province under Grant 19HASTIT026.

References

1. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)
2. Borji, A.: What is a salient object? a dataset and a baseline model for salient object detection. IEEE Trans. Image Process. **24**(2), 742–756 (2015)
3. Borji, A., Cheng, M.M., Jiang, H., Li, J.: Salient object detection: a benchmark. IEEE Trans. Image Process. **24**(12), 5706–5722 (2015)
4. Borji, A., Sihite, D.N., Itti, L.: Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study. IEEE Trans. Image Process. **22**(1), 55–69 (2013)
5. Cheng, M.M., Mitra, N.J., Huang, X., Torr, P.H., Hu, S.M.: Global contrast based salient region detection. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 569–582 (2015)
6. Han, J., Zhang, D., Hu, X., Guo, L., Ren, J., Wu, F.: Background prior-based salient object detection via deep reconstruction residual. IEEE Trans. Circuits Syst. Video Technol. **25**(8), 1309–1321 (2015)
7. Hayhoe, M., Ballard, D.: Eye movements in natural behavior. Trends Cogn. Sci. **9**(4), 188–194 (2005)
8. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. Science **313**(5786), 504–507 (2006)
9. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. arXiv preprint (2017)

10. Itti, L., Koch, C.: Computational modelling of visual attention. Nat. Rev. Neurosci. **2**(3), 194 (2001)
11. Jiang, H., Wang, J., Yuan, Z., Wu, Y., Zheng, N., Li, S.: Salient object detection: a discriminative regional feature integration approach. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2083–2090 (2013)
12. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. Computer Science (2014)
13. Klein, D.A., Frintrop, S.: Center-surround divergence of feature statistics for salient object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2214–2219. IEEE (2011)
14. Lee, G., Tai, Y.W., Kim, J.: Deep saliency with encoded low level distance map and high level features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 660–668 (2016)
15. Li, G., Yu, Y.: Deep contrast learning for salient object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 478–487 (2016)
16. Li, X., Zhao, L., Wei, L., Yang, M.H., Wu, F., Zhuang, Y., Ling, H., Wang, J.: Deepsaliency: multi-task deep neural network model for salient object detection. IEEE Trans. Image Process. **25**(8), 3919–3930 (2016)
17. Li, X., Lu, H., Zhang, L., Ruan, X., Yang, M.H.: Saliency detection via dense and sparse reconstruction. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2976–2983 (2013)
18. Liu, N., Han, J.: Dhsnet: deep hierarchical saliency network for salient object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 678–686 (2016)
19. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241 (2015)
21. Wang, L., Lu, H., Ruan, X., Yang, M.H.: Deep networks for saliency detection via local estimation and global search. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3183–3192 (2015)
22. Wang, L., Wang, L., Lu, H., Zhang, P., Ruan, X.: Saliency detection with recurrent fully convolutional networks. In: Proceedings of the European Conference on Computer Vision, pp. 825–841 (2016)
23. Yan, Q., Xu, L., Shi, J., Jia, J.: Hierarchical saliency detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1155–1162 (2013)
24. Yi, Z., Zhang, H., Tan, P., Gong, M.: Dualgan: unsupervised dual learning for image-to-image translation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2849–2857 (2017)
25. Zhang, J., Sclaroff, S., Lin, Z., Shen, X., Price, B., Mech, R.: Unconstrained salient object detection via proposal subset optimization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5733–5742 (2016)
26. Zhang, P., Wang, D., Lu, H., Wang, H., Yin, B.: Learning uncertain convolutional features for accurate saliency detection. arXiv preprint arXiv:1708.02031 (2017)
27. Zhang, Q., Lin, J., Li, W., Shi, Y., Cao, G.: Salient object detection via compactness and objectness cues. Vis. Comput. **34**(4), 473–489 (2018)

28. Zhao, R., Ouyang, W., Li, H., Wang, X.: Saliency detection by multi-context deep learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1265–1274 (2015)
29. Borji, A., Cheng, M.M., Jiang, H., Li, J.: Salient object detection: A survey, 2(4) (2014). arXiv preprint: arXiv:1411.5878