



A Genetic Programming Approach for Constructing Foreground and Background Saliency Features for Salient Object Detection

Shima Afzali^(✉), Harith Al-Sahaf, Bing Xue, Christopher Hollitt,
and Mengjie Zhang

School of Engineering and Computer Science, Victoria University of Wellington,
Wellington, New Zealand
{shima.afzali,harith.al-sahaf,bing.xue,
christopher.hollitt,mengjie.zhang}@ecs.vuw.ac.nz

Abstract. Salient Object Detection (SOD) methods have been widely investigated in order to mimic human visual system in selecting regions of interest from complex scenes. The majority of existing SOD methods have focused on designing and combining handcrafted features. This process relies on domain knowledge and expertise and becomes increasingly difficult as the complexity of candidate models increases. In this paper, we develop an automatic feature combination method for saliency features to relieve human intervention and domain knowledge. The proposed method contains three phases, two Genetic Programming (GP) phases to construct foreground and background features and a spatial blending phase to combine those features. The foreground and background features are constructed to complement each other, therefore one can improve other's shortcomings. This method is compared with the state-of-the-art methods on four different benchmark datasets. The results indicate the new automatic method is comparable with the state-of-the-art methods and even improves SOD performance on some datasets.

Keywords: Salient object detection · Foreground · Background
Genetic programming

1 Introduction

Visual saliency detection is a fundamental research and real life problem in neuroscience, psychology, and computer vision [7]. Salient Object Detection (SOD) is a process of identifying and localizing regions including objects that attract more attention than other parts of an image when examined by a human viewer [7].

In the past two decades, various types of saliency features have been designed for the SOD task by domain experts. Using the existing collection of features

saves us from designing similar or redundant features. However, manually selecting features from the existing features and combining them is not an efficient way and not guarantee the optimal combination. Liu et al. [7] developed some well-known SOD features including local, regional, and global features. However, their proposed method loses its performance in some challenging images due to lack of more informative features and a suitable combination method. Lin et al. [6] proposed a method to detect salient object by extracting multiple features such as local contrast, global contrast, and background prior. They refined local and global contrasts by object center priors and then combined the refined features to salient region detection, and the feature combination part has been manually designed by the authors.

In order to have a more precise saliency map, saliency features are required to complement each other. Some features can complement each other, while some others may corrupt others' efficacy. A good feature combination method explores complementary characteristics of features and finds an optimal way to combine these features. However, in the literature, authors often have not paid attention to the complementary characteristic of features.

The aforementioned issues motivates us to develop a method which can automatically explore a set of the different features, select informative ones, consider their complementary characteristic and combine them suitably. Genetic Programming (GP) [5] is a search strategy to automatically evolve solutions (programs) by automatically exploring different possible combinations of features. GP has a flexible tree-based representation which also allows searching the space of various integration operations to combine different features. Thus, the aforementioned capabilities of GP make it suitable choice to develop a GP-based automatic feature combination method to address the aforementioned issues.

The overall goal of this study is to develop an automatic method to combine features to construct two new informative features. We propose a new method which focuses on two important parts of the image, foreground objects and background. In the proposed method, two GP-based foreground and background feature construction phases are developed. The GP-based foreground feature mainly targets the foreground object, while the GP-based background feature focuses on suppressing background. Specifically, this paper aims to fulfill the following objectives:

- Develop new automatic feature combination method to construct two new informative features; and

- Design two new fitness functions to evaluate the evolved solutions (individuals) by GP method.

2 The Proposed Method

In this paper, the overall process contains three phases, two GP-based feature construction phases to build foreground (FG) and background (BG) features, respectively, and a spatial blending phase to combine the constructed features. GP is utilized to find a good combination of the input features to construct

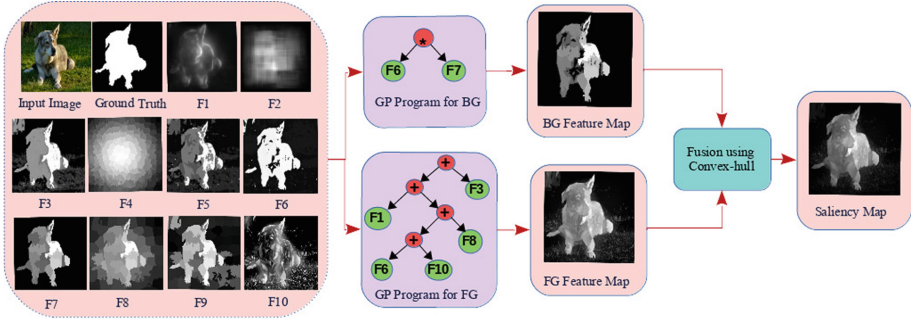


Fig. 1. Scheme of the proposed method.

FG and BG features. The process of the complete method is depicted in Fig. 1. For the first GP phase, GP-based foreground feature construction (GPFPG), we focus on constructing the FG feature in order to effectively highlight foreground object(s). In this phase, GP takes a set of saliency feature maps as input and constructs FG feature as output that is a combination of those features. For the second GP phase, GP-based background feature construction (GPBG), GP is used to construct the BG feature to suppress background. GPBG takes saliency features and the function set as input to combine features, and returns a constructed feature as output. In contrast to GPFPG, GPBG utilizes a different fitness function in constructing the BG feature (see details in Sect. 3.3). Fitness function for GPFPG: saliency detection is a type of classification model that classifies pixels into, salient or non-salient groups. Since saliency detection is a Bernoulli distribution problem, binary entropy is chosen as the fitness measure. Here, binary entropy is employed to enhance precision of salient regions by decreasing the difference between the constructed feature and the ground truth.

$$H(p, q) = -p \log q - (1 - p) \log(1 - q) \quad (1)$$

where p is the ground truth value, q is the saliency value which is calculated by the GP program, and $H(p, q)$ is the entropy value between the ground truth and the saliency map. The fitness function is the average entropy of all the training images. The lower entropy shows the better fitness value for the GP program.

Table 1. GP parameters.

Population size	Generations	Mutation rate	Crossover rate
100	50	0.19	0.8
Elitism rate	Tree depth	Selection type	Tournament size
0.01	2–4	Tournament	7

Fitness function for GPBG: *recall* is employed as the fitness function for GP because recall operates as a pessimistic measure of saliency, so attempts to suppress background regions. For the final GP phase, an object center prior map and spatial blending is employed to combine the constructed FG and BG features [9].

3 Experiment Design

In this work, the performance of the proposed method is evaluated using three widely used SOD datasets including SED1 [4], MSRA10K [7], and ECSSD [4]. Each dataset is split into a training set (60%), a validation set (20%) and a test set (20%). Each of the GP methods were run 30 times on each dataset.

Similar parameter values are used for both GP methods, GPFBC and GPBG. Table 1 summarizes the GP parameters. The parameter settings mostly follow the suggested values from the literature [3]. The initial population is created by the ramped half-and-half method. In this study, the population size is set to 100 to reduce the computational time. The tree depth was limited to 2–4, since it prevents individuals to growing inefficiently and becoming more complex. For the function set, both GP methods use a simple set of the commonly used arithmetic operations including addition, subtraction, and multiplication. Each function in the set $\{+, -, \times\}$, takes two saliency feature maps as input in 2D-array and returns another 2D-array saliency feature map as output. For the terminal set, different types of features is collected based on different characteristics of the saliency features from the literature. Here, nine saliency features are taken from the previous work [2], and the SUSAN edge detector is also added to the feature set [8]. The performance of the proposed method is evaluated using precision-recall (PR) curve, receiver operating characteristic (ROC) curve, and F-measure [4]. GPFBC is compared to seven other methods, five methods are selected from [4] including DRFI, GS, GMR, SF, RBD, and two other methods MSSS [1] and wPSO [2].

4 Results and Discussions

4.1 Quantitative Comparison

Based on the precision-recall curves in Fig. 2(a) and (b), GPFBC outperforms most other methods, but is slightly worse than RBD and DRFI. On the ECSSD dataset in Fig. 2(c), GPFBC performs better than RBD and also has a comparable result with wPSO. Based on the ROC curves in Figs. 3(a)–(c), GPFBC has the second best Area Under Curve (AUC) on all three data sets, where DRFI has the best AUC. GPFBC has a higher true positive rate in relation to false positive rate comparing to all the other methods apart from DRFI. Figure 4(a) shows that GPFBC has slightly lower average precision, recall, and F-measure to DRFI, RBD, and GS, but it has better performance than the other methods on

the SED dataset. In Fig. 4(b), GPFBC has better results than most of the methods on the ASD dataset, while DRFI and RBD have slightly better results than GPFBC. On the ECSSD dataset, GPFBC has a slightly lower average precision than wPSO and DRFI, but a higher average recall than wPSO (Fig. 4(c)). The ECSSD dataset contain more complex images than the SED and ASD datasets. Although GPFBC performs well on the ASD and SED datasets, it has better performance on ECSSD regarding average precision, recall, and F-measure. Generally, GPFBC shows a comparable or even better performance compared to the other methods except for DRFI. Although the performance of the GPFBC method is not as good as the DRFI method, GPFBC uses only 10 features and DRFI employs a 93 dimensional feature vector.

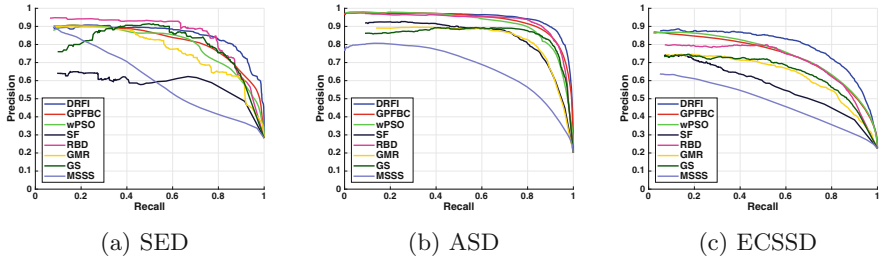


Fig. 2. Precision-recall curves of GPFBC compared to seven other methods.

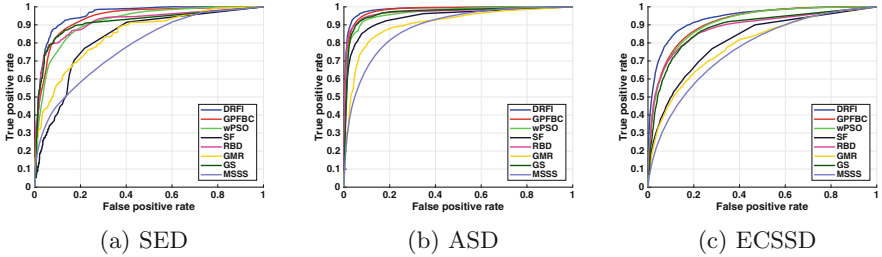


Fig. 3. ROC curves of GPFBC compared to seven other methods.

4.2 Qualitative Comparison

Some sample saliency maps are shown in Figs. 5 and 6 to illustrate the qualitative performance of GPFBC and the seven other methods. It can be seen that the performance of GPFBC is mostly good on the challenging and complex images, e.g., images having non-homogeneous foreground object (e.g., 4th row),

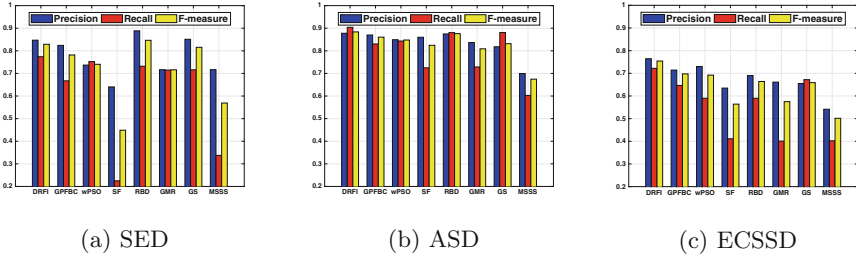


Fig. 4. Average precision, recall, and F-measure of GPFBC compared to seven other methods.

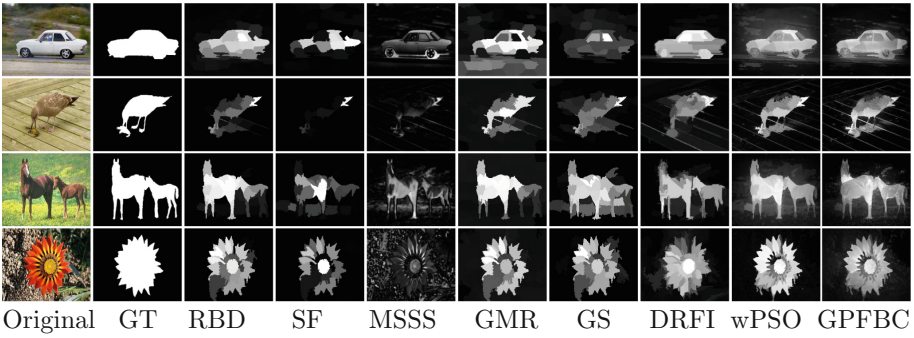


Fig. 5. Some visual examples of the new method and seven other SOD methods.

cluttered/complex background regions (e.g., 1st and 3rd rows), having more than one salient object (e.g., 3rd row), having similar color with the background (e.g., 2nd row). Generally, GPFBC shows the highest quality on suppressing background and completely detecting foreground object(s). However, it may fail in some challenging images (Fig. 6), since it has the lack of enough informative features such as shape information, texture features, and high-level features.

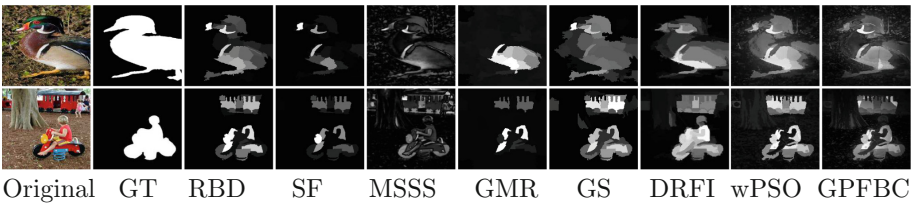


Fig. 6. Some visual examples of the new method and seven other SOD methods.

5 Conclusions

In this study, an automatic feature combination method is developed to construct two new informative features using GP to focus on the foreground object and the background, respectively. The first GP method takes input saliency features and generates a foreground feature, which is mainly good at highlighting foreground objects. The second GP method focuses on generating background feature, that mostly suppresses background for SOD. The results show that GP has a promising capability for exploring a large search space and finding a good way to combine different input saliency features. The findings motivate us to further explore GP for developing a fully automatic feature combination method in our future work that does not rely on the spatial blending approach in the third phase of the proposed method.

References

1. Achanta, R., Süsstrunk, S.: Saliency detection using maximum symmetric surround. In: Proceedings of the 17th IEEE International Conference on Image Processing, pp. 2653–2656. IEEE (2010)
2. Afzali, S., Xue, B., Al-Sahaf, H., Zhang, M.: A supervised feature weighting method for salient object detection using particle swarm optimization. In: Proceedings of the IEEE Symposium Series on Computational Intelligence, pp. 1–8 (2017)
3. Al-Sahaf, H., Al-Sahaf, A., Xue, B., Johnston, M., Zhang, M.: Automatically evolving rotation-invariant texture image descriptors by genetic programming. *IEEE Trans. Evol. Comput.* **21**(1), 83–101 (2017)
4. Borji, A., Cheng, M.M., Jiang, H., Li, J.: Salient object detection: a survey. arXiv preprint [arXiv:1411.5878](https://arxiv.org/abs/1411.5878), pp. 1–26 (2014)
5. Koza, J.R.: Genetic Programming (1997)
6. Lin, M., Zhang, C., Chen, Z.: Predicting salient object via multi-level features. *Neurocomputing* **205**, 301–310 (2016)
7. Liu, T., et al.: Learning to detect a salient object. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(2), 353–367 (2011)
8. Smith, S.M., Brady, J.M.: SUSANA new approach to low level image processing. *Int. J. Comput. Vis.* **23**(1), 45–78 (1997)
9. Yang, C., Zhang, L., Lu, H.: Graph-regularized saliency detection with convex-hull-based center prior. *IEEE Signal Process. Lett.* **20**(7), 637–640 (2013)