




# A benchmark dataset and baseline model for co-salient object detection within RGB-D images

Ning Yang<sup>1,2</sup> · Chen Zhang<sup>1,2</sup> · Yumo Zhang<sup>1,2</sup> · Haowei Yang<sup>1,2</sup> · Ling Du<sup>3</sup> 

Received: 9 February 2021 / Revised: 12 August 2021 / Accepted: 9 September 2021 /  
Published online: 2 June 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

Within-image co-salient object detection (wCoSOD) identifies the common and salient objects within an image, which can benefit for many applications, such as reducing information redundancy, animation synthesis, and so on. Besides, the introduction of depth information that conforms to the stereo perception of human is also more conducive to accurately detecting salient objects. Thus, in this paper, we focus on a new task from the perspective of the benchmark dataset and baseline model, i.e., within-image co-salient object detection in RGB-D images. To bridge the gap the new task and algorithm verification, we first collect a new dataset containing 240 RGB-D images and the corresponding pixel-wise ground truth. Then, we propose an unsupervised method for within-image co-salient object detection in RGB-D images. Under the constraint of depth information, our model decomposes the within-image co-salient object detection task into two parts: determining the salient object proposals; combining the similarity constraint and cluster-based constraint between different proposals to locate the co-salient object and generate the final result. The experimental results on the collected dataset demonstrate that our method achieves competitive performance both qualitatively and quantitatively.

**Keywords** RGB-D images · Within-image co-salient object detection · Benchmark dataset · Unsupervised baseline method

---

✉ Ling Du  
duling@tiangong.edu.cn

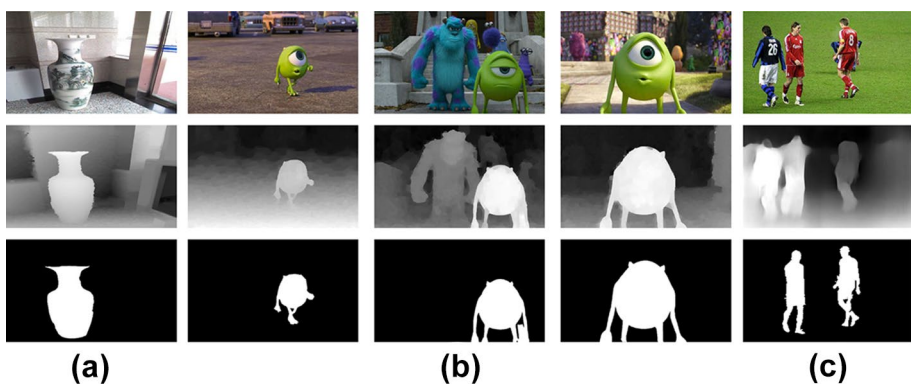
<sup>1</sup> Institute of Information Science, Beijing Jiaotong University, 100044 Beijing, China

<sup>2</sup> The Beijing Key Laboratory of Advanced Information Science and Network Technology, 100044 Beijing, China

<sup>3</sup> School of Computer Science and Technology, Tiangong University, 300387 Tianjin, China

## 1 Introduction

Salient object detection (SOD) simulates the human visual attention mechanism to locate the visually attractive object or regions from a scene [8], which has been applied in a large number of vision tasks, such as image classification [17], image retrieval [15], image compression [18], and image retargeting [13], etc. In addition, consistent with human stereo perception, RGB-D salient object detection aims to introduce the depth cue on the basis of RGB information to effectively suppress the complex backgrounds and better highlight the salient objects, as shown in Fig. 1(a). Co-salient object detection (CoSOD) in multiple images has attracted extensive attention from many successful applications. Inspired by human inductive ability and collaborative processing mechanism, co-salient object detection in RGB-D images not only pays attention to the salient object in the single image, but also determines the recurring ones in multiple related images by modeling the inter-images relationship. As shown in Fig. 1(b), RGB-D CoSOD aims to detect common salient object that exists in all three RGB-D images, i.e., the green cartoon characters. However, it is of certain practical significance for the co-salient object detection within an image, mainly detecting the salient objects with similar attributes in an image, which is called within-image co-salient object detection (wCoSOD). As shown in Fig. 1(c), the salient and recurring objects within the image are athletes in red instead of athletes in blue. This task has many potential applications in computer vision, such as reducing information redundancy [37], synthesizing realistic animations from still images [35], detecting multiple instances of an object class [19, 24], etc. The distinction between wCoSOD and CoSOD for RGB-D image is whether the processed data is multiple different images or a single image, so their corresponding modeling methods are also different. The co-salient object detection of multiple images is performed in an image group, so the correspondence between the objects between different images needs to be considered when modeling. While wCoSOD is to find a co-salient object in a single image, the corresponding relationship of different objects in the intra-image needs to be modeled. Considering the effectiveness of depth information and the significance of the wCoSOD task, in this paper, we introduce depth information into the wCoSOD task for the first time, construct a corresponding RGB-D wCoSOD benchmark dataset and provide an unsupervised baseline model.



**Fig. 1** Examples of the differences between different SOD task in RGB-D images. From the top to the bottom corresponds to the input RGB images, depth maps, and saliency ground truths. (a) RGB-D SOD. (b) RGB-D CoSOD. (c) RGB-D wCoSOD

As mentioned above, the current studies on wCoSOD only focus on RGB image without the aid of depth information. Research on RGB-D wCoSOD is expected to further increase the detection performance, while further prospering the SOD family. Thus, the first thing that needs to be solved is that there is no publicly available RGB-D wCoSOD dataset, which is not conducive to the development of this direction to a certain extent. Based on this, we first collect and annotate an RGB-D wCoSOD benchmark dataset, which laid a data foundation for subsequent algorithm research and performance evaluation. A total of 240 RGB-D images were collected, and the corresponding pixel-wise ground truths were manually annotated. In this dataset, 50% of the images have both common salient objects and non-common salient objects that need to be suppressed, which makes it very challenging. In addition, we also propose an unsupervised method to achieve RGB-D wCoSOD by considering the cluster constraint and similarity matching. The task of RGB-D wCoSOD is decomposed into two parts for construction of the model. By introducing the depth information, more accurate and compact saliency proposals are first determined. Then, the similarity that contains the depth information and cluster-based constraints between different proposals are designed to measure the correspondences and locate the co-salient object.

The main contributions can be summarized as follows:

- To the best of our knowledge, this paper focuses on the within-image co-salient object detection in RGB-D images for the first time, and constructs the first publicly available dataset for this task containing 240 RGB-D images.
- With the help of depth information, we propose an unsupervised baseline model, and model the correspondence as a consistency matching problem among different proposals under the similarity and cluster-based constraints.
- Experimental results on the collected dataset demonstrate that our method achieves competitive performance compared with the existing RGB-D SOD and RGB wCoSOD algorithms.

The rest of this paper is composed as follows. Section 2 reviews the related work. Section 3 details the proposed method. The construction of the RGB-D wCoSOD dataset is introduced in Section 4. Section 4 presents the experimental results and analysis. The conclusions are described in Section 6.

## 2 Related work

In this section, we briefly review the tasks of SOD. According to different actual needs and processing data, SOD model can be further divided into RGB SOD [6, 36], RGB-D SOD [10, 23, 29, 31], CoSOD [5, 9, 14, 20], video SOD [22, 34], and so on. Among them, CoSOD aims to detect salient and recurring objects in an image group by considering the inter-image relationship, which is similar to our task in this paper. Based on the different inter-image modeling strategies, CoSOD methods can be roughly categorized as clustering based method [14], matching based method [5], depth-induced method [9, 31], and learning based method [20].

The difference from CoSOD is that wCoSOD has only one input image, so it is necessary to model the correspondence between different objects in an image, instead of modeling the constraint relationship between the objects among different images. In view of the characteristics of wCoSOD, the researchers designed some methods to solve it. In [37,

[38], an unsupervised bottom-up wCoSOD method was proposed by finding common and saliency proposal groups through optimization and fusion. Song et al. [32] proposed a multi-scale multiple instance learning (MIL) model for wCoSOD. However, for RGB-D wCoSOD task, as yet it has still blank space. Therefore, in this work, we not only collected a dataset suitable for this task, but also proposed an unsupervised baseline model.

### 3 Proposed method

In this paper, we propose an unsupervised wCoSOD model for RGB-D images. As shown in the framework of Fig. 2, our method consists of three parts. First, we determine the saliency attribute and generate the proposal candidates. Specifically, we apply the existing RGB-D SOD models (e.g., A2dele [29]) to obtain the initial saliency map  $S^{init}$  including all the salient objects without distinguishing common attributes. At the same time, we use the Selective Search algorithm [33] to generate some proposal candidates based on the RGB-D images for common attribute judgment, and abstract the RGB image  $I$  into several superpixels  $R = \{r_i\}$ ,  $i = 1, 2, \dots, M$  through the SLIC method [2] to improve computation efficiency and structural representation. Then, the correspondence relationship between proposals at the superpixel level is calculated based on the similarity constraints  $w_{RD}$  and cluster-based constraints  $w_C$ , and then the initial co-saliency map  $S^{co-init}$  is obtained. Finally, the initial saliency map  $S^{init}$  and the initial co-saliency map  $S^{co-init}$  are weighted fused to generate the final co-saliency map  $S^{co}$ .

#### 3.1 Depth assisted proposal generation

Different from the CoSOD in an image group, there is only one image in wCoSOD task without the concept of multiple images. Therefore, how to determine the common attributes among salient objects from single image is the key to our task. Inspired by general CoSOD methods, we can model the correspondence between objects at the proposal level. For the object/region in a proposal, if it is salient and appears multiple times in the entire image, then it is the co-salient object that our task is looking for. Thus, we should generate some proposal candidates to determine the common attributes between objects.

First, the initial  $L$  proposals are extracted from the input RGB image  $I$  via the Selective Search algorithm [33], which are denoted as  $P^{init} = \{P_n^{init}\}$ ,  $n = 1, 2, \dots, L$ . Considering that

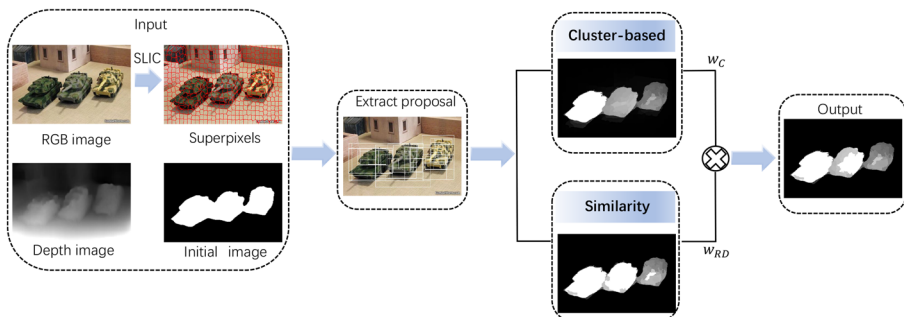


Fig. 2 The framework of the proposed RGB-D wCoSOD

only the salient proposals are useful for our final task, thus we need to ensure that there are the salient regions in the proposals. To be specific, we set three rules to filter the initial proposals: (1) The proposals should have an appropriate size to avoid covering multiple objects or partial areas inside. (2) The proposals should be salient to highlight the saliency attribute. (3) The proposals should have larger depth value to eliminate the background interference. The process can be formulated as:

$$P^{fin} = size(P^{init}) \cap sal(P^{init}) \cap dep(P^{init}), \tag{1}$$

where  $size()$  preserves proposals with the size of 1%-30%,  $sal()$  removes proposals with the average salient value of less than 0.2,  $dep()$  selects the top 80% proposals with larger average depth value. The proposals that conform to the above conditions are defined as denoted as  $P^{fin} = \{P_n\}, n = 1, 2, \dots, N$ .

### 3.2 Proposal-wise correspondence calculation

As mentioned earlier, capturing the corresponding relationship and determining the common attributes are the key points for the co-salient object detection model. In our proposed method, the correspondence is modeled as the consistency measurement of different superpixels in different proposals through similarity constraint and cluster-based constraint. In fact, the co-salient objects should have the similar attributes in terms of color appearance, depth distribution, and category. Therefore, we design the similarity constraint considering the color and depth information and the cluster-based constraint to measure the correspondence at proposal level.

Color information intuitively describes the content of an image, and depth information helps distinguish the foreground and background well. Therefore, when calculating the consistency of different superpixels in different proposals, both color and depth information are considered to express similarity constrain at the same time. We define the similarity matrix between superpixel  $r_i^p$  and  $r_j^q$ :

$$\omega_{RD}(r_i^p, r_j^q) = exp\left(-\frac{\|h_i^p, h_j^q\|_2 + \lambda |d_i^p - d_j^q|}{\sigma^2}\right), \tag{2}$$

where  $r_i^p$  represents the  $i$ -th superpixel in the  $p$ -th proposal,  $h_i^p$  denotes the mean color vector of superpixel  $r_i^p$  in the  $L^*a^*b^*$  color space,  $\|h_i^p, h_j^q\|_2$  is the euclidean distance between  $h_i^p$  and  $h_j^q$ ,  $d_i^p$  denotes the mean depth value of superpixel  $r_i^p$ ,  $\lambda$  is the depth confidence measure [10], and  $\sigma^2$  is a parameter to control strength of the similarity, which is fixed to 0.1.

In an image, the salient objects can be further divided into co-salient objects and non-co-salient objects, in which the co-salient objects belong to the same category and can be clustered together. Thus, we propose the cluster-based constraint to define the correspondence. Different from the previous work [9, 14], we consider the degree of correlation between the largest subject class of superpixels. Specifically, we firstly apply k-means method [3] to group the image into  $K$  clusters. Then, we define the class probability of superpixel  $r_i$ :

$$c_k^i = n_k^i / \sum_{k=1}^K n_k^i, \tag{3}$$

where  $n_k^i$  represents the number of pixels belonging to the  $k$ -th class in superpixel  $r_i$ .  $c_k^i$  represents the probability that superpixel  $r_i$  belongs to the cluster  $k$ . Then, the cluster

correlation between superpixel  $r_i^p$  and  $r_j^q$  is defined as the class probability of the superpixel  $r_j^q$  in the  $q$ -th proposal with consistent clustering attribute:

$$\omega_C(r_i^p, r_j^q) = c_{mk}^{j,q}, \quad (4)$$

where  $mk = \operatorname{argmax}_{k \in K} (c_k^{i,p})$  is the clustering index of superpixel  $r_i^p$  in the  $p$ -th proposal, which is defined as the cluster corresponding to the maximum class probability. In summary,  $w_{RD}$  reflects the degree of similarity between superpixels, and  $w_C$  measures the likelihood of superpixels belonging to the same class. Thus, the initial co-saliency of a superpixel is computed as the weighted sum of the initial saliency of corresponding superpixels in other proposals, which is formulated as:

$$\begin{aligned} & S^{co-init}(r_i^p) \\ &= \frac{1}{N-1} \sum_{q=1}^N \frac{1}{N_q} \sum_{j=1}^{N_q} \lambda_1 w_{RD}(r_i^p, r_j^q) \cdot \lambda_2 w_C(r_i^p, r_j^q) \cdot S^{init}(r_j^q), \end{aligned} \quad (5)$$

where  $\lambda_1$  and  $\lambda_2$  are the weighted coefficients, which can be adjusted according to different scenarios and needs. Without loss of generality, they are both set to 1 in the experiment.  $N$  is the number of proposals in the image  $I$ ,  $N_q$  is the number of superpixels in the  $q$ -th proposal  $P_q$ .

### 3.3 Within-image co-saliency map generation

Finally, the initial saliency map  $S^{init}$  and the initial co-saliency map  $S^{co-init}$  are integrated through intersection and union operation to generate the final co-saliency map  $S^{co}$ . The intersection better suppresses the interference area, and the union better ensures the consistency of the saliency map. The formula is represented as:

$$S^{co} = \frac{1}{2} \gamma_1 \cdot (S^{init} + S^{co-init}) + \gamma_2 \cdot (S^{init} \cdot S^{co-init}), \quad (6)$$

where  $\gamma_1$  and  $\gamma_2$  are the weighted coefficients, which are all set to 0.5 in experiments without loss of generality.

## 4 Dataset

Existing publicized image datasets used to evaluate RGB-D salient object detection, such as DUT-RGBD [28], NLPR [27] and NJUD [21], are mainly used to test salient object detection in RGB-D SOD. In most cases, each image contains only one salient object, which is annotated as the ground truth. In this paper, we have a different goal to detect within RGB-D images co-salient object, which is not shown in most images in the public dataset. Therefore, we collected a new benchmark dataset consisting of 240 RGB-D images for co-salient object detection within RGB-D image, which provides the corresponding pixel-wise ground truth. In the dataset, about 221 RGB-D images are selected from the current RGB-D datasets [7, 12, 21, 25–27, 39], and the remaining 19 RGB-D images are collected from the RGB within-image co-saliency dataset [37] and the depth maps are generated by

the depth estimation method [16]. For the within-image co-saliency annotation, we follow the work [9] to reproduce the ground truth. Some visual examples are shown in Fig. 3. We can see that in this dataset, some images have some interfering but salient objects (such as the Fig. 3(a)), the wCoSOD results of some images are consistent with the single-image SOD results (such as Fig. 3(b)), and there are a few images without any within-image co-salient object in it, as shown in Fig. 3(c). Moreover, our dataset contains both indoor and outdoor scenes, the object types are also relatively diverse, and the background is relatively complex, which makes our dataset very challenging. The dataset will be released after the paper is accepted.

## 5 Experiments

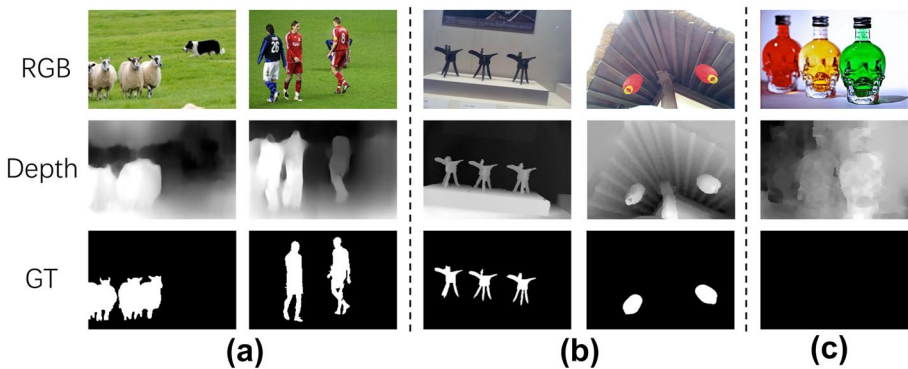
### 5.1 Experimental metrics and settings

In order to evaluate the quality of the model, we evaluate the performance of our method on the collected dataset, including the precision-recall (PR) curve, F-measure [1], S-measure [11], and MAE [8]. The precision is the percentage of correctly assigned salient pixels in the detected saliency map, and the recall represents the ratio of the detected salient pixels in the ground truth. The PR curve can be drawn using the paired precision and recall values. When the PR curve is closer to (1,1), the performance of the algorithm is better. F-measure is defined as the weighted average of precision and recall to measure both as a whole:

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

where  $\beta^2$  is set to 0.3, which is the balance parameter [4].

The mean absolute error (MAE) calculates the average absolute error between the detected saliency map  $S$  and the ground truth  $G$ .



**Fig. 3** Sample RGB images, depth images and their corresponding ground-truth images. (a) Images have some interfering but salient objects. (b) Images are consistent with the single-image SOD results. (c) Images without any within-image co-salient object

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

where  $W$  and  $H$  are the width and height of the image respectively. Obviously, the smaller the MAE value, the better the performance of the algorithm.

The S-measure value combines the regional perception ( $S_r$ ) and object perception ( $S_o$ ) to evaluate the structural similarity between the detected saliency map and ground truth.

$$S_m = \alpha * S_o + (1 - \alpha) * S_r, \quad (9)$$

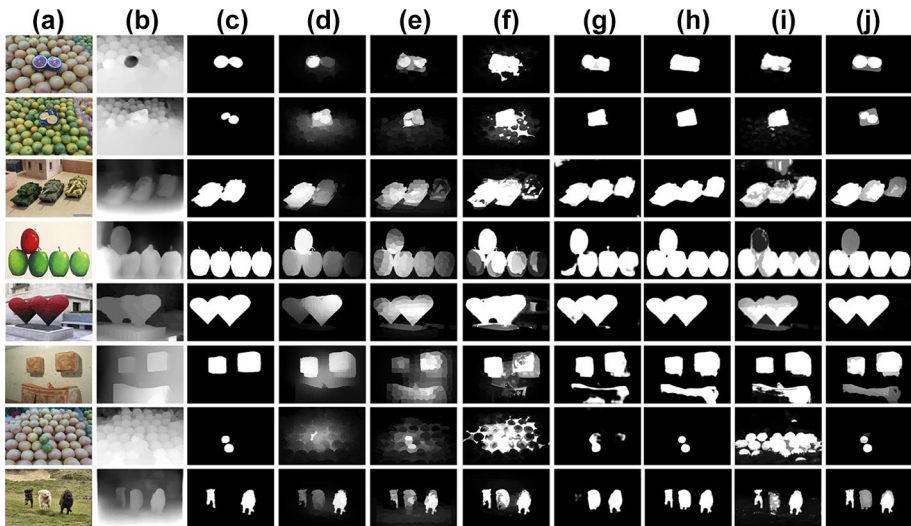
where  $\alpha$  is set to 0.5 as suggested in [11].

In the experiment, we apply the A2dele [29] model to obtain the initial saliency map. The number of superpixels  $M$  is set to 500 and  $K$  is set to 6. The experiments are conducted on 1.6 GHz frequency Intel i5 CPU and 8GB of RAM using Matlab 2016a, and the average running time is 43 seconds.

## 5.2 Performance comparison

We compare our method with six methods, i.e., DF [30], DCMC [10], CDCP [40], DMRA [28], A2dele [29], CDS [37]. CDS is an RGB wCoSOD unsupervised model. DCMC and CDCP are the RGB-D SOD unsupervised models. DF, DMRA and A2dele are the deep learning based RGB-D SOD models, in which the DF method is trained on the NLPR [27] and NJUD [21], and DMRA (2019) and A2dele (2020) are trained on the DUT-RGBD [28], NLPR [27] and NJUD [21] datasets. These training datasets are designed for RGBD SOD tasks, including the paired RGB and depth images. Because there is no RGB-D wCoSOD dataset before this work, and the core of our work in this paper is a tentative exploration of a new task. Therefore, the scale of the first dataset we constructed is relatively small, and it is mainly used for the design of unsupervised RGB-D wCoSOD model. Furthermore, the comparison algorithms in the experiment have not been retrained, and better performance should be obtained through retraining the RGB-D SOD models on a large-scale RGB-D wCoSOD dataset. This is also the direction of future efforts to collect larger-scale data to promote the development of deep learning based RGB-D wCoSOD model. Figure 4 shows the visualized results of different SOD methods. As can be seen, the RGB-D SOD models can not effectively detect the co-salient regions in the within-image due to the lack of correspondence modeling. For example, in the last image of Fig. 4, the non-common but salient object (i.e., the white dog in the middle) cannot be effectively suppressed by the A2dele method [29]. The PR curves of different methods on the proposed dataset are shown in Fig. 5. It can be seen that our model achieves higher precision on the whole PR curves. Compared with CDS, our method can better highlight the common regions by introducing of depth information and correspondence strategy. Table 1 shows the quantitative comparisons of different methods. It can be directly seen that the proposed method is better than other methods in all measurements. Specifically, for the F-measure score, our method reaches 0.7000, which improves performance by 2.5% compared to the second best algorithm (i.e., A2dele). These measurements testify to the superiority of our method.





**Fig. 4** Comparison of visual results. (a) RGB image. (b) Depth image. (c) Ground truth. (d) DF. (e) DCMC. (f) CDCP. (g) DRMA. (h) A2dele. (i) CDS. (j) Ours

**Table 1** The S-measure, MAE and F-measure of different models

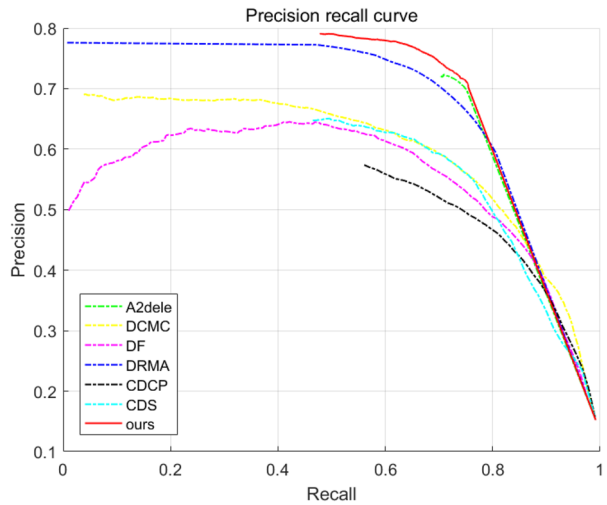
	Pub.&Year	Type	$S_m$	MAE	$F_\beta$
DF [30]	TIP'2017	RGBD Sin	0.6793	0.1538	0.5748
DCMC [10]	SPL'2016	RGBD Sin	0.6878	0.1496	0.5940
CDCP [40]	ICCV'2017	RGBD Sin	0.6509	0.1656	0.5323
DMRA [28]	ICCV'2019	RGBD Sin	0.7507	0.0904	0.6807
A2dele [29]	CVPR'2020	RGBD Sin	0.7625	0.0841	0.6830
CDS [37]	AAAI'2018	RGB wCo	0.6849	0.1520	0.5811
Ours	–	RGBD wCo	0.7728	0.0831	0.7000

“Sin” represents the single-image salient object detection model. “wCo” represents the within-image co-salient object detection

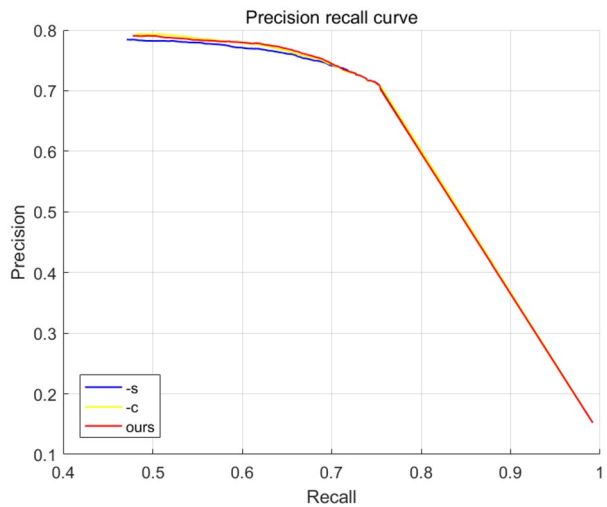
### 5.3 Ablation study

The proposed model uses two constraints, namely similarity constraint and cluster-based constraint, when calculating the co-saliency value. The PR curves of the ablation study on different mechanisms of constraints are shown in Fig. 6. It can be seen that our different constraint modules are effective. In order to verify the effectiveness of the proposed model, we conducted ablation experiments. The quantitative comparison is shown in Table 2, where “-s” represents the use of cluster-based constraint only, and “-c” represents the use of similarity constraint only. From the perspective of quantitative indicators, the performance is degraded when only one constraint is used to measure the correspondecne. Specifically, the S-measure score is 0.7692 when the cluster-based constraint is used alone. When both constraints exist, the S-measure score reaches 0.7728, with an increase of 0.5%. All these experiments demonstrate the effectiveness of the designed modules.

**Fig. 5** PR curves of different methods on the proposed dataset



**Fig. 6** PR curves of the ablation study on different mechanisms of constraints



**Table 2** Ablation study on different mechanisms of constraints

	$S_m$	MAE	$F_\beta$
-s	0.7692	0.0845	0.6950
-c	0.7722	0.0833	0.6976
Ours	0.7728	0.0831	0.7000

## 6 Conclusion

In this paper, we address the problem within-image co-salient object detection (wCoSOD) in a single RGB-D image for the first time. We collected a corresponding benchmark

dataset and proposed an unsupervised baseline method. Our method decomposes the wCoSOD task into salient object proposal generation and correspondence modelling combining the similarity and cluster-based constraints. The experimental results demonstrate the effectiveness of our method. In the future, a larger scale dataset can be constructed and some learning based methods can be designed to further improve the performance.

**Acknowledgements** This work was supported by the Beijing Nova Program under Grant Z201100006820016.

## References

1. Achanta R, Hemami S, Estrada F, Susstrunk, S (2009) Frequency-tuned salient region detection. In: Proceedings of the IEEE computer vision and pattern recognition, pp. 1597–1604
2. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süssstrunk S (2012) SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(11):2274–2282
3. Arthur D, Vassilvitskii S (2007) K-means++: The advantages of careful seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms, pp. 1027–1035
4. Borji A, Cheng MM, Jiang H, Li J (2015) Salient object detection: A benchmark. *IEEE Transactions on Image Processing* 24(12):5706–5722
5. Cao X, Tao Z, Zhang B, Fu H, Feng W (2014) Self-adaptively weighted co-saliency detection via rank constraint. *IEEE Transactions on Image Processing* 23(9):4175–4186
6. Cheng M, Mitra NJ, Huang X, Torr PHS, Hu S (2015) Global contrast based salient region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37(3):569–582
7. Cheng Y, Fu H, Wei X, Xiao J, Cao X (2014) Depth enhanced saliency detection method. In: Proceedings of international conference on internet multimedia computing and service, pp. 23–27
8. Cong R, Lei J, Fu H, Cheng MM, Lin W, Huang Q (2019) Review of visual saliency detection with comprehensive information. *IEEE Transactions on Circuits and Systems for Video Technology* 29(10):2941–2959
9. Cong R, Lei J, Fu H, Huang Q, Cao X, Hou C (2017) Co-saliency detection for RGB-D images based on multi-constraint feature matching and cross label propagation. *IEEE Transactions on Image Processing* 27(2):568–579
10. Cong R, Lei J, Zhang C, Huang Q, Cao X, Hou C (2016) Saliency detection for stereoscopic images based on depth confidence analysis and multiple cues fusion. *IEEE Signal Processing Letters* 23(6):819–823
11. Fan DP, Cheng MM, Liu Y, Li T, Borji A (2017) Structure-measure: A new way to evaluate foreground maps. In: Proceedings of the IEEE computer vision and pattern recognition, pp. 4548–4557
12. Fan DP, Lin Z, Zhang Z, Zhu M, Cheng MM (2020) Rethinking RGB-D salient object detection: models, data sets, and large-scale benchmarks. *IEEE Transactions on Neural Networks and Learning Systems* 32(5):2075–2089
13. Fang Y, Zeng K, Wang Z, Lin W, Fang Z, Lin CW (2014) Objective quality assessment for image retargeting based on structural similarity. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 4(1):95–105
14. Fu H, Cao X, Tu Z (2013) Cluster-based co-saliency detection. *IEEE Transactions on Image Processing* 22(10):3766–3778
15. Gao Y, Shi M, Tao D, Xu C (2015) Database saliency for fast image retrieval. *IEEE Transactions on Multimedia* 17(3):359–369
16. Godard C, Mac Aodha O, Firman M, Brostow GJ (2019) Digging into self-supervised monocular depth estimation. In: IEEE international conference on computer vision, pp. 3828–3838
17. Guo H, Zheng K, Fan X, Yu H, Wang S (2019) Visual attention consistency under image transforms for multi-label image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 729–739
18. Han S, Vasconcelos N (2006) Image compression using object-based regions of interest. In: IEEE international conference on image processing, pp. 3097–3100
19. He X, Gould S (2014) An exemplar-based CRF for multi-instance object segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 296–303

20. Jeong DJ, Hwang I, Cho NI (2018) Co-salient object detection based on deep saliency networks and seed propagation over an integrated graph. *IEEE Transactions on Image Processing* 27(12):5866–5879
21. Ju R, Ge L, Geng W, Ren T, Wu G (2014) Depth saliency based on anisotropic center-surround difference. In: *IEEE International conference on image processing (ICIP)*, pp. 1115–1119
22. Kim H, Kim Y, Sim JY, Kim CS (2015) Spatiotemporal saliency detection for video sequences based on random walk with restart. *IEEE Transactions on Image Processing* 24(8):2552–64
23. Li C, Cong R, Piao Y, Xu Q, Loy CC (2020) RGB-D salient object detection with cross-modality modulation and selection. In: *Proceedings of the european conference on computer vision*, pp. 225–241
24. Lin TY, Maire M, Belongie S, Hays J, Zitnick CL (2014) Microsoft coco: Common objects in context. In: *Proceedings of the european conference on computer vision*, pp. 740–755
25. Liu N, Zhang N, Shao L, Han J (2020) Learning selective mutual attention and contrast for RGB-D saliency detection. [arXiv:2010.05537](https://arxiv.org/abs/2010.05537)
26. Niu Y, Geng Y, Li X, Liu F (2012) Leveraging stereopsis for saliency analysis. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 454–461
27. Peng H, Li B, Xiong W, Hu W, Ji R (2014) RGB-D salient object detection: a benchmark and algorithms. In: *Proceedings of the IEEE european conference on computer vision*, pp. 92–109
28. Piao Y, Ji W, Li J, Zhang M, Lu H (2019) Depth-induced multi-scale recurrent attention network for saliency detection. In: *Proceedings of the IEEE international conference on computer vision*, pp. 7254–7263
29. Piao Y, Rong Z, Zhang M, Ren W, Lu H (2020) A2dele: Adaptive and attentive depth distiller for efficient RGB-D salient object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 9060–9069
30. Qu L, He S, Zhang J, Tian J, Tang Y, Yang Q (2017) RGBD salient object detection via deep fusion. *IEEE Transactions on Image Processing* 26(5):2274–2285
31. Song H, Liu Z, Xie Y, Wu L, Huang M (2016) RGBD co-saliency detection via bagging-based clustering. *IEEE Signal Processing Letters* 23(12):1722–1726
32. Song S, Yu H, Miao Z, Guo D, Ke W, Ma C, Wang S (2019) An easy-to-hard learning strategy for within-image co-saliency detection. *Neurocomputing* 358(17):166–176
33. Uijlings JR, Van De Sande KE, Gevers T, Smeulders AW (2013) Selective search for object recognition. *International Journal of Computer Vision* 104(2):154–171
34. Xi T, Zhao W, Wang H, Lin W (2017) Salient object detection with spatiotemporal background priors for video. *IEEE Transactions on Image Processing* 26(7):3425–3436
35. Xu X, Wan L, Liu X, Wong TT, Wang L, Leung CS (2008) Animating animal motion from still. In: *Proceedings of the ACM transactions on graphics*, pp. 1–8
36. Ye L, Liu Z, Li L, Shen L, Bai C, Wang Y (2017) Salient object segmentation via effective integration of saliency and objectness. *IEEE Transactions on Multimedia* 19(8):1742–1756
37. Yu H, Zheng K, Fang J, Guo H, Wang S (2018) Co-saliency detection within a single image. In: *Proceedings of the AAAI conference on artificial intelligence*, pp. 7509–7516
38. Yu H, Zheng K, Fang J, Guo H, Wang S (2020) A new method and benchmark for detecting co-saliency within a single image. *IEEE Transactions on Multimedia* 22(12):3051–3063
39. Zhu C, Li G (2017) A three-pathway psychobiological framework of salient object detection using stereoscopic technology. In: *Proceedings of the IEEE international conference on computer vision workshops*, pp. 3008–3014
40. Zhu C, Li G, Wang W, Wang R (2017) An innovative salient object detection using center-dark channel prior. In: *Proceedings of the IEEE international conference on computer vision workshops*, pp. 1509–1515

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.