entropy

MDPI

Article

Exploring Focus and Depth-Induced Saliency Detection for Light Field

Yani Zhang [1], Fen Chen [1,2,*], Zongju Peng [1,2], Wenhui Zou [2] and Changhe Zhang [1]

[1] School of Electrical and Electronic Engineering, Chongqing University of Technology, Chongqing 400054, China; yani_zhang@stu.cqut.edu.cn (Y.Z.); pengzongju@126.com (Z.P.); 13614503112@163.com (C.Z.)

[2] Faculty of Information Science and Engineering, Ningbo University, No. 818, Ningbo 315211, China; zouwench@163.com

* Correspondence: chenfen1@cqut.edu.cn

Abstract: An abundance of features in the light field has been demonstrated to be useful for saliency detection in complex scenes. However, bottom-up saliency detection models are limited in their ability to explore light field features. In this paper, we propose a light field saliency detection method that focuses on depth-induced saliency, which can more deeply explore the interactions between different cues. First, we localize a rough saliency region based on the compactness of color and depth. Then, the relationships among depth, focus, and salient objects are carefully investigated, and the focus cue of the focal stack is used to highlight the foreground objects. Meanwhile, the depth cue is utilized to refine the coarse salient objects. Furthermore, considering the consistency of color smoothing and depth space, an optimization model referred to as color and depth-induced cellular automata is improved to increase the accuracy of saliency maps. Finally, to avoid interference of redundant information, the mean absolute error is chosen as the indicator of the filter to obtain the best results. The experimental results on three public light field datasets show that the proposed method performs favorably against the state-of-the-art conventional light field saliency detection approaches and even light field saliency detection approaches based on deep learning.

Keywords: light field; saliency detection; focus cue; foreground; color and depth-induced cellular automata

Citation: Zhang, Y.; Chen, F.; Peng, Z.; Zou, W.; Zhang, C. Exploring Focus and Depth-Induced Saliency Detection for Light Field. *Entropy* **2023**, *25*, 1336. https://doi.org/10.3390/e25091336

Academic Editor: Wei Li

Received: 31 July 2023
Revised: 30 August 2023
Accepted: 5 September 2023
Published: 15 September 2023

1. Introduction

The light field is a densely sampled image array, which has brought humans closer to recording the real world. Compared with traditional images, the light field can record the intensity and direction of light rays and has a stronger expressive ability. With the development of refocusing and rendering techniques, many light field applications have been generated, such as depth estimation [1,2] and light field super-resolution [3,4]. As an important image preprocessing, light field saliency detection is crucial to promote the research of light field applications.

Saliency detection aims to identify important regions that are interesting or obvious to human eyes and improve the understanding of computer vision applications, such as semantic segmentation [5], object detection [6], image recognition [7], etc. Saliency detection methods based on RGB images make it difficult to accurately detect salient objects in cluttered scenarios by color, compactness, or contrast cues. Moreover, saliency detection methods based on RGB-D images are easily misled by depth maps, as shown in Figure 1.

Existing methods [8–14] detected salient regions via hand-crafted features such as color, depth, and focus, while having a limited exploration of the light field and a few related studies. The deep learning methods, leveraging powerful extraction and expression capabilities, promote the development of light field saliency detection. Piao et al. [15] used a single sub-aperture image to synthesize the complex light field. In [16], a micro-lens

image is used to predict salient regions through a convolutional network, which did not consider the correlation between sub-aperture maps. Wang et al. [17] designed a cross-modal feature fusion module to fuse the aggregated features from various modalities in the three networks of all-focus, depth, and focal stack images. Liang et al. [18] adopted a weakly supervised network and exploited the features of focal stack and depth maps to generate pixel-level pseudo-saliency maps. Jiang et al. [19] leveraged attention mechanisms to explore cross-modal complementarity and overcome information loss in the focal stack. Yuan et al. [20] refined the focal stack with depth modality, which enhances the structure and location information of salient objects in the focal stack.

However, there is no traditional method for exploring the relationship between the focal stack and the depth map, and how to use the interaction between these to improve the performance of saliency detection in the light field is still a problem worth thinking about.

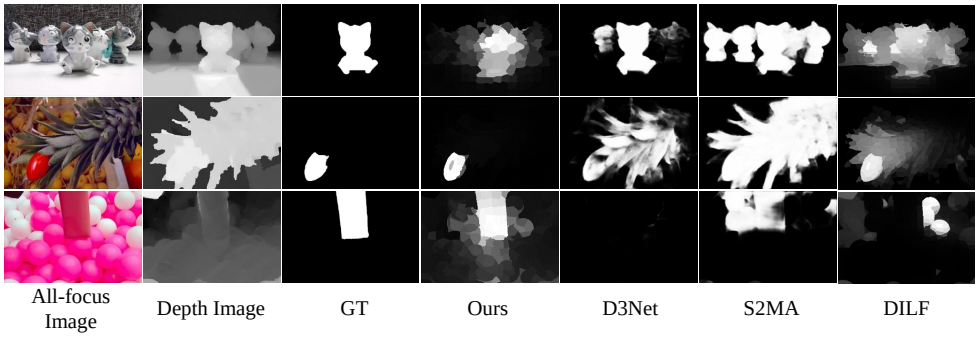


Figure 1. Comparison of the detection results due to depth image misleading by different methods (D3Net [21], S2MA [22], DILF [10]) on the DUT-LF [15] dataset.

The key to successfully identifying salient objects is via exploring the potential features and discovering the interactions between different cues in the light field. In this paper, we propose a light field saliency detection algorithm, in which the cues among the focal stack, depth, and all-focus images are fully explored and utilized to improve saliency detection via complementation and fusion. Specifically, the coarse region of the salient object is localized by combining the color and depth compactness. For the focal stack, we compute the background and foreground probabilities to highlight the foreground. Among these, the foreground object is enhanced by the depth contrast and the foreground probability. In addition, we introduce the local geodesic saliency cues [23] and combine them with the depth compactness to refine the saliency map. Furthermore, taking into account the consistency and smoothness of the saliency object, we design a saliency optimization model, the color and depth-induced cellular automata (CDCA), by exploiting the complementarity of the color and depth cues to optimize the salient detection map with higher accuracy. Finally, to avoid the influence of low-quality saliency maps, we use a filter to obtain excellent saliency detection results by comparing the values of the mean absolute error (MAE).

To summarize, the main contributions of this paper are:

1. We introduce a light field saliency detection method taking into account interactions among the depth, focus, color, and compactness cues.
2. We separate foreground and background by taking advantage of the focal stack and depth image. Exploring the depth feature, we extract the depth compactness and local saliency cues to emphasize local regions for refinement. At the same time, we integrate foreground probability with a depth contrast map to highlight the foreground.
3. We develop the CDCA optimization model, which integrates color and depth cues to improve the spatial consistency of saliency detection results.

The rest of the paper is organized as follows: Section 2 overviews the related works on RGB, RGB-D, and light field saliency detection. In Section 3, the proposed method and formulation are described in detail. The experimental results are presented and analyzed in Section 4. Finally, we summarize the proposed method in Section 5.

2. Related Works

This section will briefly review the related work on RGB images, RGB-D images, and light field saliency detection methods.

2.1. RGB Image Saliency Detection

RGB image saliency detection models can be divided into two categories, top-down and bottom-up models. Top-down methods are mainly task-oriented for saliency detection. Bottom-up methods mainly use the low-level features of an image, such as color, texture, and contrast information for saliency detection. This is especially true after the appearance of the simple linear iterative cluster (SLIC) [24] algorithm, which improves computational efficiency for image segmentation. Initially, saliency detection methods [25] used global and local contrast information as salient features for detection. Yang et al. [26] considered both foreground and background cues in a different way and ranked the similarity of the image regions via graph-based manifold ranking. Zhu et al. [27] analyzed the spatial distribution of the region and regarded the region with a large boundary border as the background. Zhou et al. [28] recovered the falsely suppressed salient regions by combining image internal compactness and local contrast information. Inspired by the automatic cellular machine, Yao et al. [29] used the propagation mechanism of the single-layer cellular automaton (SCA) to find the intrinsic correlation of similar regions and dynamically update the saliency map. At the same time, a multi-layer automatic cellular (MCA) optimization algorithm is proposed to integrate the advantages of the salient features.

It is difficult to achieve predictions that are highly consistent with human perception only by relying on low-level features, thus motivating many deep learning-based salient object detection models. Lou et al. [30] introduced a U-Net network to fuse multi-level context and perform salient object detection in both local and global manners. Wang et al. [31] added a pyramid attention structure and edge detection module to the network to obtain accurate salient area edges while expanding the receptive field. To make full use of the global context information, Chen et al. [32] utilized several context-aware modules to incorporate multiple levels of features with global contextual information. Since humans assign more attention to moving objects, Zhou et al. [33] proposed a motion-attention transfer network for zero-shot video object segmentation within the encoder, which not only inherits the advantages of multi-modal learning but also utilizes motion-attention to facilitate appearance learning. Fully supervised networks rely on a large amount of annotated data labels; Lai et al. [34] adopted weak supervision to improve salient objects in complex scenes by exploring the nature of visual attention patterns.

However, it is difficult to detect more accurate saliency results only by limiting features in challenging scenarios.

2.2. RGB-D Image Saliency Detection

According to the perceptual characteristics of human eyes, salient objects are often located in the foreground. Nowadays, many RGB-D saliency detection methods reduce misjudgments in challenging scenes by adding depth information. Niu et al. [35] demonstrated that stereoscopic information can provide a useful complement to existing RGB image saliency detection. Peng et al. [36] proposed a multi-stage RGB-D saliency detection model and demonstrated that depth information can improve the robustness of saliency results. Ren et al. [37] performed saliency detection by fusing global priors such as regional contrast, depth information, and background. Cong et al. [38] proposed a saliency detection method based on depth confidence and multi-information fusion to reduce the impact of depth images. Zhu et al. [39] found that the center-dark channel prior can distinguish the foreground and background. Cong et al. [23] refined the generated saliency detection map by extracting the shape and contour of salient objects in the depth image and achieved excellent results.

In addition, the method of RGB-D saliency detection using deep learning [21,22,40,41] can achieve better results. Zhu et al. [40] designed an independent sub-network to extract

depth cues and guide the main network to improve the saliency detection performance. Liu et al. [22] proposed a selective attention mechanism to weigh the mutual attention to filter unreliable information. Fan et al. [21] proposed a new RGB-D dataset and designed a deep cleaning unit to filter low-quality salient results. Zhao et al. [41] designed a depth-awareness framework by excavating depth information and exploiting the low-level boundary cues to achieve accurate salient detection results. However, RGB-D image unavoidably suffer from the influence of poor-quality depth images on the detection results, especially when the salient and non-salient objects are at the same depth.

2.3. Light Field Saliency Detection

The rich features of the light field are beginning to be used to supplement the insufficiency of RGB and RGB-D images. Li et al. [8] calculated the likelihood score for the first time to distinguish the foreground and background of light field images for saliency detection. Li et al. [9] constructed salient and non-salient dictionaries from feature vectors, but the salient results could not provide a better visual experience for human eyes. Zhang et al. [10] combined depth, color contrast, and background probability for saliency detection and obtained the complete salient region. Wang et al. [11] fused color contrast, background prior, depth, and focus information through a Bayesian framework, but did not fully explore the depth image. Zhang et al. [12] used multiple cues to improve the performance of saliency detection by complementing the differences between different information. The single-layer cellular automaton can enhance the saliency consistency between similar regions. Inspired by [29], Piao et al. [13] proposed a depth-guided automatic cellular machine model (DCA) that can automatically optimize saliency maps based on depth, focus, and color information. Wang et al. [14] calculated the degree of focus to generate depth information to reduce the dependence on the depth map.

With the continuous improvement of light field datasets, many models have begun to use convolutional networks to extract the salient features of the light field. Piao et al. [15] proposed a multi-view object detection network to synthesize multi-view images for saliency detection. Zhang et al. [16] proposed an end-to-end convolutional network to extract the salient features of micro-lens images. Wang et al. [42] and Zhang et al. [43] have committed to exploring the correlation and fusion between focal stack and all-focus images to improve saliency detection performance, but they need to rely on high-quality focal stacks. To aggregate cross-level features, Wang et al. [17] propose a cross-modal feature fusion module to fuse features from various modalities from three sub-networks. Liang et al. [18] designed a weakly supervised learning framework based on a pseudo-ground truth to solve the problem of unclear edges of salient objects in complex structures. Jiang et al. [19] utilized the attention mechanism to explore cross-modal complementarities and generated object edges and edge features to progressively refine regional features to achieve edge-aware detection. Zhang et al. [44] designed a multi-task collaborative network for light field salient object detection to explore the complementary coherence among multiple cues, including spatial, edge, and depth information. Feng et al. [45] exploited the relationship between light field cues to identify clean labels from pixel-level noisy labels for saliency detection. Yuan et al. [20] used the multi-modal feature guidance method to refine the focal stack, enhance the structure and position information about the salient objects in the focal stack, and improve accuracy.

Compared with conventional methods, deep learning methods can obtain high-quality and remarkable results but require greater computing power, which increases the cost of experiments to a certain extent. We focus extensively on designing a low-cost detection model and adopting the interaction and complementarity between light field cues to improve the saliency detection performance in challenging scenarios.

3. Methodology

In this paper, we make full use of the advantages of focus, depth, and color to improve the accuracy of the saliency detection model. Figure 2 shows the framework of the proposed

method and the framework mainly has the following stages: (1) We obtain the global compactness saliency to locate the compactly distributed areas. (2) The foreground and background probabilities are calculated and respectively combined with the depth contrast cue and the global compactness map to highlight the salient objects, as shown in Figure 3. (3) The global compactness saliency is refined by exploring high-quality depth compactness and local geodesic cues. (4) The saliency maps are optimized by the CDCA model to obtain a more perfect saliency map. (5) We design an output model to obtain excellent saliency detection results by judging the MAE value. Figure 4 shows the visual processes of each step. In the next section, this paper presents the details of the method.

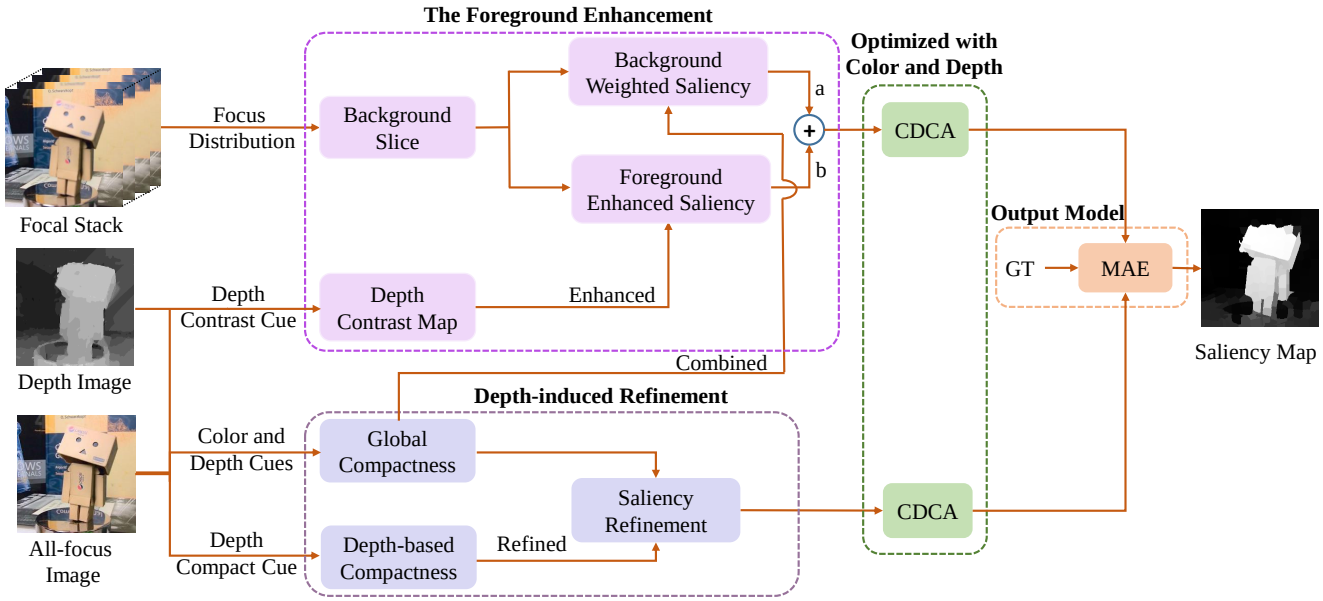


Figure 2. The framework of the proposed method where "a" and "b" are the different weighted coefficient to balance the background saliency map and foreground saliency map.

3.1. Compactness Based on Color and Depth Cues

In the spatial domain, the salient region usually has a compact spread. In the depth domain, the region closest to the camera often contains a concentrated distribution in the depth image. Motivated by this, we integrate color and depth to define global compactness, which can be used to distinguish salient objects from the background.

We divide an image into compact and homogenous superpixels by the SLIC algorithm [24] and construct a graph $G = (V, E)$, where V represents the generated superpixel node set, and E represents the distance between adjacent nodes' connection set. Therefore, the similarity between superpixels v_i and v_j in the Lab color space and depth space is defined as:

$$a_{ij}^{c} = exp(-||c_i - c_j||_2/\sigma^2) \tag{1}$$

$$a_{ij}^{d} = exp(-\lambda_d \cdot |d_i - d_j|/\sigma^2) \tag{2}$$

where c_i is the mean color value of v_i superpixels in the Lab color space, the mean depth value of v_i superpixels in d_i depth space, $\lambda_d = exp((1 - m_d) \cdot CV \cdot H - 1)$ is the depth confidence [38], m_d is the mean value of the depth map, CV is the coefficient of variation, and H denotes the depth frequency entropy. λ_d is used to judge the quality of a depth image. The higher the value of λ_d, the better the quality of the depth image. $\sigma^2 = 0.1$ [26] is a constant that controls the strength of the similarity. Here, the global compactness based on color and depth is defined as:

$$S_{CS}(i) = 1 - norm(cc(i) + cd(i)) \tag{3}$$

where $norm(x) = (x - x_{min})/(x_{max} - x_{min})$ is a function that normalizes x to the range of [0, 1]. $cc(i)$ and $cd(i)$ are the color compactness and depth compactness of the superpixel v_i, respectively. They are expressed as follows:

$$cc(i) = \frac{\sum_{j=1}^{N} n_j \cdot a_{ij}^c \cdot \|b_j - \mu_i\|}{\sum_{j=1}^{N} n_j \cdot (a_{ij}^c + a_{ij}^d)} \tag{4}$$

$$cd(i) = \frac{\sum_{j=1}^{N} n_j \cdot a_{ij}^d \cdot \|b_j - p_0\|}{\sum_{j=1}^{N} n_j \cdot (a_{ij}^c + a_{ij}^d)} \tag{5}$$

where N represents the number of superpixels, n_j represents the number of pixels in the superpixel v_i, $b_j = [b_j^x, b_j^y]$ represents the centroid coordinates of the superpixel v_j, $\mu_i = \left[\frac{\sum_{j=1}^{N} a_{ij}{}^c \cdot n_j \cdot b_j^x}{\sum_{j=1}^{N} a_{ij}{}^c \cdot n_j}, \frac{\sum_{j=1}^{N} a_{ij}{}^c \cdot n_j \cdot b_j^y}{\sum_{j=1}^{N} a_{ij}{}^c \cdot n_j}\right]$ represents the spatial average value, and p_0 is the coordinate of the center.

3.2. Exploring Focus for Foreground Enhancement

Effectively distinguishing the foreground from the background is a key step in salient detection. Considering the problem that salient objects are mostly located in the foreground and the foreground is not easy to obtain, we analyze the focus distribution in the focal stack to select the background slice, and determine the foreground by finding the background. At the same time, global compact saliency maps based on color and depth information can comprehensively detect salient objects in images, but the detection results are rough. To further suppress the interference caused by the background, we fuse the background and foreground probabilities with the global compact map and the depth-contrast saliency map, respectively, to achieve the purpose of separating the foreground and background.

The focal stack is a set of focused slices focused on the foreground and background, and the difference of the focus point will lead to the sharpness difference of different regions. Considering the advantages of the center prior and background prior, we detect background regions and compute background and foreground probabilities to highlight salient objects.

In order to highlight the foreground and suppress the background, we select the background slice and compute the background probability [10] to refine the global compactness saliency map, and the result $S_{bg}(i)$ is as follows:

$$S_{bg}(i) = \sum_{i=1}^{N} S_{cs}(i) \cdot Pb_{bg}(i) \tag{6}$$

$$Pb_{bg}(i) = 1 - exp(-\frac{U_{bg}(i)^2}{2\sigma_{bg}^2} \cdot \left\|p_0 - U_{pos}^*(i)\right\|^2) \tag{7}$$

where we set σ_B= 1 to ensure that the background probability is maximized, $U_{foc}(i)$ is the mean value of superpixel v_i of the slice, $||C - U_{pos}^*(i)||$ is used to measure the superpixel spatial information related to the superpixel and the image center, and $U_{pos}^*(i)$ is the normalized average coordinate of the superpixel v_i.

At the same time, to fully extract the depth cues, we introduce the foreground probability $P_{fg}(i)$ [13] to highlight the foreground objects. Figure 3 shows that the foreground probability enhances the depth cues. The foreground saliency map $S_{fg}(i)$, induced by the depth cues, is defined as:

$$S_{fg}(i) = \sum_{j=1}^{N} S_D(i) \cdot P_{fg}(j) \tag{8}$$

$$P_{fg}(i) = exp(-\frac{U_{foc}(i)^2}{2\sigma_F^2} \cdot \left\|p_0 - U_{pos}^*(i)\right\|^2 \cdot \|1 - d(i)\|^2) \tag{9}$$

where σ_F= 0.2, and the depth-induced contrast saliency S_D and spatial weight factors are:

$$S_D(i) = \sum_{j=1}^{N} W_{pos}(i,j) \cdot \|d_i - d_j\| \tag{10}$$

$$W_{pos}(i,j) = exp(-\frac{||U^*_{pos}(i) - U^*_{pos}(j)||^2}{2\sigma_w^2}) \tag{11}$$

where σ_w= 0.67.

A depth prior has a great help in distinguishing a salient object from the background. To further emphasize the salient object, we distribute more weight to the foreground salient maps, and the saliency detection result map S_{FF} is obtained by weighted fusion as:

$$S_{FF} = \alpha \cdot S_{fg} + (1 - \alpha) \cdot S_{bg} \tag{12}$$

where α is set to 0.7.

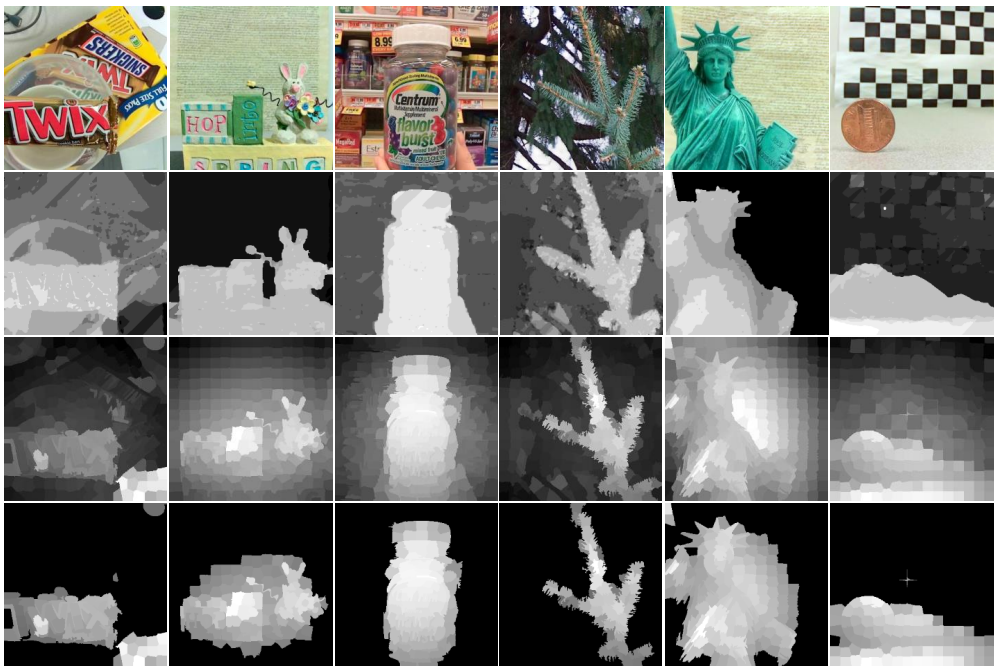


Figure 3. Foreground probability enhances depth contrast saliency map (from top to bottom are the all-focus image, depth image, depth contrast saliency map, and enhancement result).

3.3. Depth-Induced Refinement

In order to avoid the influence of a poor-quality depth image, the depth confidence [38] is introduced to calculate the high-quality depth compactness and refine the foreground. The improved depth compactness $S_{dc}(i)$ is defined as:

$$S_{dc}(i) = 1 - N(\frac{\sum_{j=1}^{N} n_j \cdot a_{ij}^c \cdot a_{ij}^d \cdot \|b_j - p_0\| \cdot exp(-\frac{\lambda_d \cdot d_i}{\sigma^2})}{\sum_{j=1}^{N} n_j \cdot a_{ij}^c \cdot a_{ij}^d}) \tag{13}$$

Observing the depth map contained in existing light field datasets, we find that the quality of the depth map correlates with the detection of high-quality salient objects. Considering the quality of depth maps of different datasets, we adopt the average depth confidence value as a benchmark to extract depth information. When the quality of the depth map is reliable (i.e., $\lambda_d \geq mean$), the salient object has an obvious depth contrast with the background, and the depth compactness can refine the saliency map. When the quality of the depth map is poor (i.e., $\lambda_d < mean$), we utilize the global compactness to

strengthen the salient features. For refinement, the saliency map combined with the depth feature is defined as:

$$S_{DF}(i) = \begin{cases} 0.5 \cdot S_{CS}(i) + 0.5 \cdot N(S_{dc}(i) \cdot S_g(i) + S_{dc}(i)), \lambda_d \geq mean \\ 0.5 \cdot S_{CS}(i) + 0.5 \cdot N(S_{dc}(i) \cdot S_{CS}(i) + S_{dc}(i)), \lambda_d < mean \end{cases} \quad (14)$$

where the local geodesic distance saliency $S_g(i)$ [23] is introduced to accumulate edge weights along the shortest path from a superpixel to a background node, *mean* is the average depth confidence of the depth image, and the specific value is placed in the experiment section.

3.4. Color and Depth-Induced Optimization

We observe that the edges of the generated saliency maps are not clear, the salient objects are incomplete, and the background still has subtle disturbances. Considering the consistency of color smoothing and depth space, the proposed algorithm adds depth information on the basis of the existing optimization model, and the improved optimization model can obtain salient detection results with clearer edges and improved quality.

To obtain a meticulous saliency map, we improve upon a model called the color and depth-induced cellular automata (CDCA) to optimize saliency maps. The CDCA model is mainly based on two considerations: (1) Depth information helps highlight foreground objects. We integrate the color and depth cues to define cell neighbors can improve the optimization accuracy. (2) To avoid the influence of poor-quality depth maps, superpixels on the image boundaries are considered background seeds. When the quality of the depth map is reliable, the CDCA model can effectively optimize the saliency map, and its synchronous update rule can greatly improve the detection of incomplete prior saliency maps in challenging scenes. Therefore, we construct the impact factor matrix f_{ij}^{cd} as follows:

$$f_{ij}^{cd} = \begin{cases} exp(-(||c_i - c_j||_2 + |d_i - d_j|)/\sigma^2), j \epsilon NB(i) \\ 0, i = j \quad or \quad otherwise \end{cases} \quad (15)$$

where σ^2 is a parameter controlling the strength of similarity, and set σ^2= 0.1. $NB(i)$ is the set of the neighbors of cell i. To normalize the influence factor matrix, we generate the degree matrix $D = diag\{d_1, d_2, \ldots, d_{N2}\}$, $d_i = \sum_j f_{ij}^{cd}$. Then, the normalized influence factor matrix is:

$$F^* = D^{-1} \cdot F \quad (16)$$

In order to balance the importance of the cell's current state and the cell's neighbors' state, a coherence matrix $C = diag\{c_1, c_2, \ldots, c_N\}$ is constructed to promote the evolution of the cell. Then, the consistency calculation of each cell to its current state is:

$$c_i = \frac{1}{max(f_{ij})} \quad (17)$$

In order to control c_i in the range of $[b, a + b]$, the constants a and b are set to 0.6 and 0.2, respectively. Then, the coherence matrix is $C^* = diag\{c_1^*, c_2^*, \ldots, c_N^*\}$ as:

$$c_i^* = a \cdot \frac{c_i - min(c_j)}{max(c_i) - max(c_j)} + b \quad (18)$$

Here, the synchronous update rule is defined as:

$$S^{t+1} = C^* \cdot S^t + (I - C^*) \cdot F^* \cdot S^* \quad (19)$$

where S^t is the refined saliency map when t= 0, and the ultimate saliency map after N1 time steps is denoted as S^{t+1}.

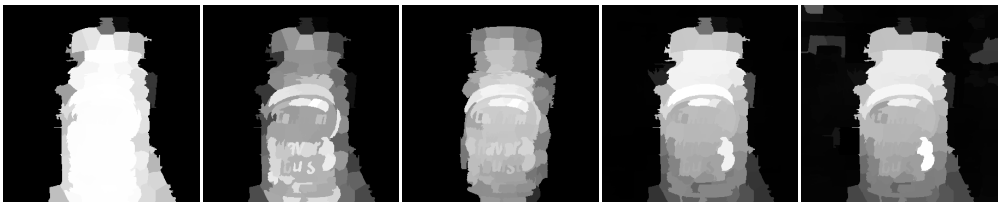
3.5. Output Model Using MAE

To reduce the redundant information brought by saliency map fusion, the output result mainly considers the following aspects: (1) When the quality of the depth image is poor, the noise introduced by the low-quality depth map should be avoided. (2) When the quality of the depth image is reliable, salient objects can be identified by depth contrast cues. (3) When the focal stack and the depth image are both reliable, we filter to obtain the excellent saliency results. Figure 4 shows the visual process of the proposed method.

We design a simple screening filter, judging the mean absolute error (MAE) value, to obtain the optimal prediction result. The higher the saliency map accuracy is, the smaller the value of the MAE is. The final saliency detection result is denoted as:

$$S_{LF} = \begin{cases} S_{DF}, MAE_{S_{DF}} \leq MAE_{S_{FF}} \\ S_{FF}, MAE_{S_{DF}} > MAE_{S_{FF}} \end{cases} \tag{20}$$

(a) GT (b) S_{CS} (c) S_{fg} (d) S_{bg} (e) S_g



(f) S_{dc} (g) S_{DF} (h) S_{FF} (i) S^*_{FF} (j) S^*_{DF}

Figure 4. The visual process of the proposed method.

4. Experiment and Verification

4.1. Dataset and Parameter Setup

In this paper, we select the existing public light field datasets, LFSD [8], HFUT [12], and DUT-LF [15] to verify the effectiveness and robustness of the proposed method. The LFSD dataset captures 60 indoor and 40 outdoor scenes, most of which have a single salient object and reliable depth image quality. The HFUT dataset contains 255 pictures, which not only contains a large number of challenging scenes, such as small objects, multiple targets, or image blur, but also the depth images are poor quality. The DUTLF-FS dataset consists of 1000 training images and 462 test images. The salient object has the characteristics of small size, low contrast with the background, and multiple salient objects without connection. At the same time, some images are affected by light intensity. In the experiments, we use test images of the DUT-LF dataset for verification and comparison, and the experiments run on Matlab 2018b.

We set the number of superpixels to 200 in all experiments. In the CDCA model, the number of time steps $N1$=20. In the depth refinement stage, the average depth confidence of LFSD [8] and DUT-LF [15] is 0.22, and that of HFUT [12] is 0.03.

4.2. Performance Evaluation Measures

To conduct a quantitative performance evaluation, we compute the precision-recall (PR) curve, F-measure, WF-measure [46], E-measure [47], S-measure [48], and mean absolute error (MAE) to evaluate the state-of-the-art detection models used for comparison.

The PR curve reflects the relationship between precision and recall. By binarizing the saliency map and the ground-truth map with a threshold, the value of precision and

recall can be calculated. The F-measure is the result calculated by the weighted sum of the precision and recall. The higher the F-measure value is, the more effective the model is:

$$F_\beta = \frac{(1+\beta^2) \cdot Precision \cdot Recall}{\beta^2 \cdot Precision + Recall} \tag{21}$$

where $\beta^2 = 0.3$.

In [46], considering the correlation between the pixels of the saliency map and the position of the wrong pixels, a weighting function ω is added in precision and recall to represent the importance of the pixels and the degree of dependence between different pixels. The weighted F-measure is defined as follows:

$$F_\beta^\omega = \frac{((1+\beta^2) Precision^\omega \cdot Recall^\omega)}{(\beta^2 \cdot Precision^\omega + Recall^\omega)} \tag{22}$$

The E-measure is used to evaluate the structural similarity between the saliency detection map and the ground-truth map [47], and its specific formula is:

$$E_S = \frac{1}{W \cdot H} \sum_{x=1}^{W} \sum_{y=1}^{H} \phi(x,y) \tag{23}$$

where $\phi(x,y)$ is the enhanced alignment matrix.

The S-measure is used to obtain the two characteristics of pixel-level matching and image-level statistics [48], and its specific formula is:

$$S_\lambda = \lambda \cdot S_o + (1-\lambda) \cdot S_\gamma \tag{24}$$

where S_o and S_γ represent object-aware and region-aware structural similarity, respectively, and λ is a balance parameter and is set to 0.5.

The MAE expresses the similarity between the saliency map and the true value map, which is used to measure the average error between each pixel of the binarized saliency map and the ground truth. The MAE is expressed as follows:

$$MAE = \frac{1}{W \cdot H} \sum_{x=1}^{W} \sum_{y=1}^{H} \|S(x,y) - GT(x,y)\| \tag{25}$$

where W and H, respectively, represent the width and height of the image, $S(x,y)$ is the continuous saliency map, and $GT(x,y)$ is the the binary ground truth.

4.3. Comparison with State-of-the-Art Methods

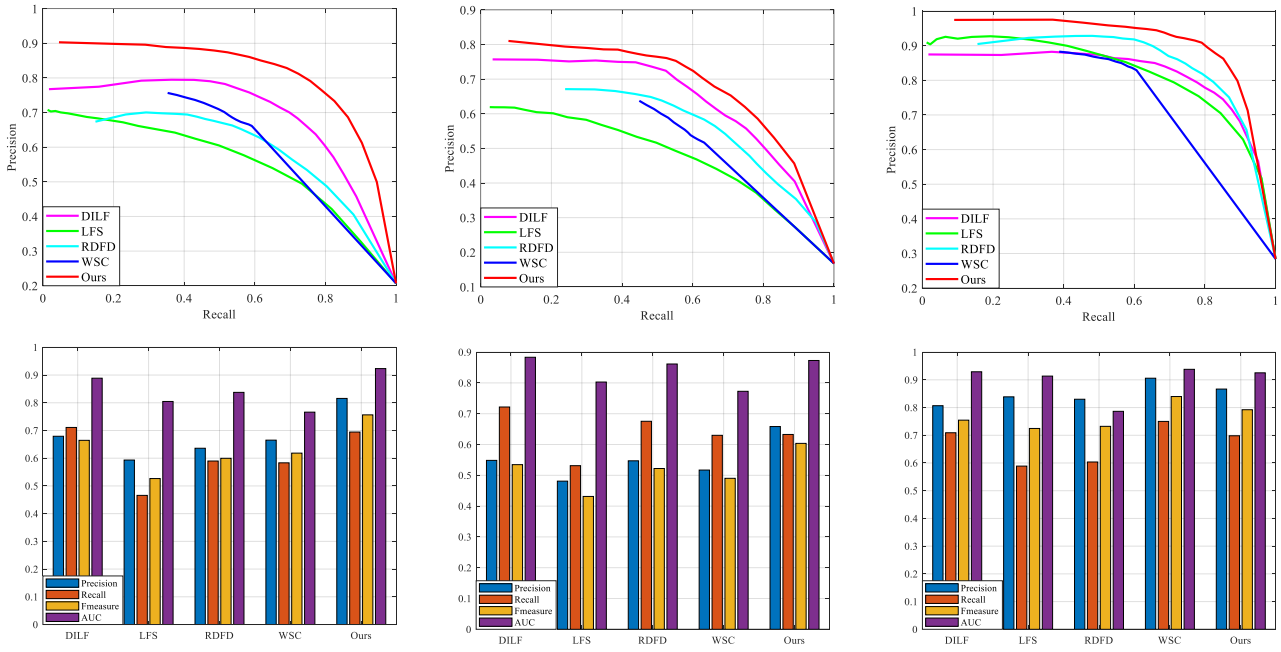
In this paper, we focus on bottom-up saliency detection models and qualitatively compare them with the state-of-the-art conventional light field methods (LFS [8], WSC [9], DILF [10], and RDFD [14]). All comparative saliency maps were provided by the authors or run on publicly available code. Figure 5 shows the visual comparison of the proposed method with the others on the LFSD [8], HFUT [12], and DUT-LF [15] datasets, and the proposed method achieves the highest PR curve. On the HFUT and DUT-LF datasets shown in Figure 5a,b, the proposed method can improve the detection performance in challenging scenes by exploring the interaction and complementarity among different light field features when the quality of the depth map and focal stack is poor. When the quality of the image is reliable, the proposed method can achieve the superior saliency detection performance shown in Figure 5c.

Table 1 shows the quantitative performance evaluation of the proposed method and the others (LFS [8], WSC [9], DILF [10], RDFD [14]). The proposed method achieves the best score, and the saliency detection results obtained are superior to the latest conventional methods. It is demonstrated that salient object detection performance in challenging scenarios can be improved by exploiting the interaction and complementarity among the salient features of the light field.

Table 1. Quantitative comparison between the proposed method and the state-of-the-art saliency detection methods on different datasets (E-measure (E_β), S-measure (S_α), WF-measure (F_β^ω), F-measure (F_β), and MAE(M)) (bold: best).

Method	LFSD [8]					HFUT [12]					DUT-LF [15]				
	E_β	S_α	F_β^ω	F_β	M	E_β	S_α	F_β^ω	F_β	M	E_β	S_α	F_β^ω	F_β	M
LFS [8]	0.749	0.660	0.470	0.725	0.219	0.666	0.565	0.260	0.426	0.222	0.742	0.585	0.309	0.525	0.228
WSC [9]	0.778	0.693	0.637	0.735	0.163	0.679	0.613	0.428	0.485	0.154	0.787	0.656	0.527	0.617	0.151
DILF [10]	0.828	0.790	0.654	0.787	0.149	0.693	0.672	0.430	0.530	0.151	0.813	0.725	0.517	0.663	0.157
RDFD [14]	0.813	0.760	0.647	0.792	0.152	0.691	0.619	0.355	0.518	0.215	0.782	0.658	0.443	0.599	0.192
Ours	**0.847**	**0.812**	**0.720**	**0.840**	**0.124**	**0.746**	**0.687**	**0.455**	**0.600**	**0.148**	**0.841**	**0.759**	**0.557**	**0.756**	**0.144**

To prove that the proposed method is better than other state-of-the-art saliency detection methods, we compare them with related RGB/RGB-D and light field methods, including RGB conventional methods (DCLC [28], BSCA [29]), RGB-D methods (CDCP [39], DCMC [38], D3Net [21], S2MA [22], PDNet [40]), and light field methods (DLSD [15], MAC [16], DCA [13], NoiseLF [45]). The experiments show that the proposed method not only outperforms conventional saliency detection methods but also achieves a more accurate salient object detection at a lower computational cost than deep learning methods, as shown in Table 2 and Figure 6.



(a) Compare on HFUT dataset (b) Compare on DUT-LF dataset (c) Compare on LFSD dataset

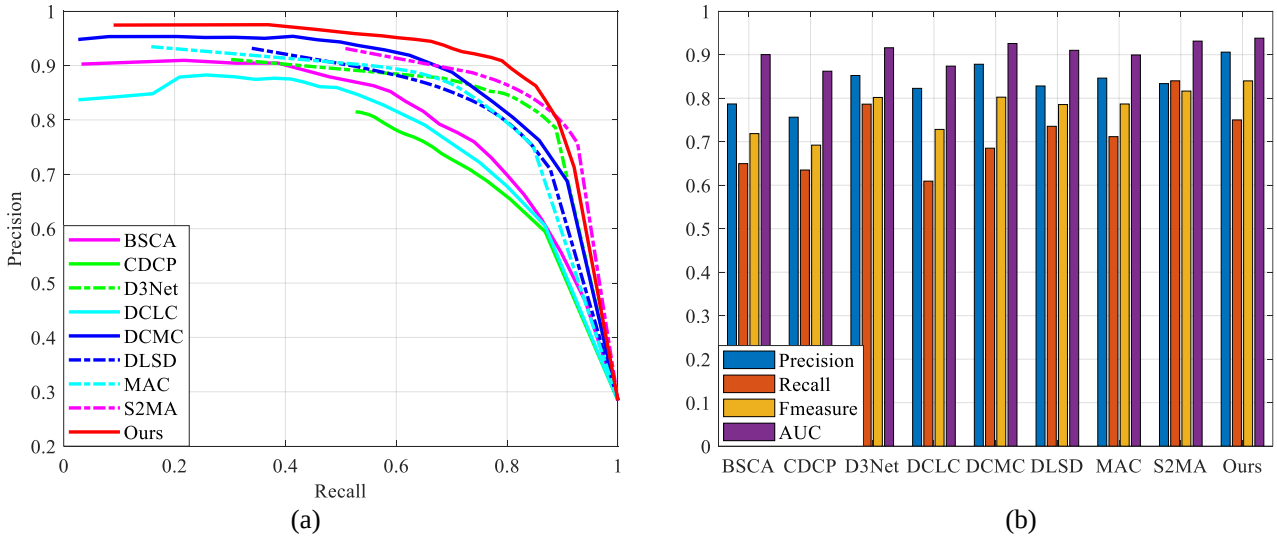
Figure 5. Performance comparison with the proposed method and the state-of-the-art conventional light field saliency detection methods.

Table 2. Quantitative comparisons between the proposed method and the state-of-the-art deep learning and traditional methods on the LFSD dataset ((E-measure (E_β), S-measure (S_α), WF-measure (F_β^ω), F-measure (F_β), and MAE(M)). The top three models are highlighted in red, blue, and green, respectively.

Category	Method	E_β	S_α	F_β^ω	F_β	M
RGB	DCLC [28]	0.765	0.668	0.511	0.728	0.200
	BSCA [29]	0.780	0.723	0.549	0.719	0.198

Table 2. *Cont.*

Category	Method	E_β	S_α	F_β^ω	F_β	M
RGB-D	CDCP [39]	0.763	0.699	0.595	0.692	0.181
	DCMC [38]	0.817	0.722	0.584	0.802	0.172
RGB-D^{DL}	D3Net [21]	0.840	0.808	0.751	0.802	0.107
	S2MA [22]	0.851	0.820	0.764	0.816	0.105
	PDNet [40]	-	-	-	0.822	0.075
Light Field^{DL}	DLSD [15]	0.840	0.778	0.703	0.785	0.125
	MAC [16]	0.819	0.768	0.681	0.787	0.133
	NoiseLF [45]	-	-	-	0.804	0.111
Light Field	DCA [13]	-	-	-	0.831	0.133
	Ours	0.847	0.812	0.720	0.840	0.124



(a) (b)

Figure 6. Performance comparison of the proposed method with other state-of-the-art saliency detection methods (RGB/RGB-D/light field) on LFSD dataset (the dotted line is the deep learning method). (**a**) Precision–recall curves of different methods. (**b**) Average precision, recall, F -measure, and AUC of different methods.

4.4. Ablation Study

In this section, we analyze the effectiveness of using depth and color cues in the CDCA model and the contributions of different components of the proposed method.

4.4.1. The Effectiveness of the CDCA Model

To prove the effectiveness of the CDCA model in the proposed method, we compare the results of utilizing the CDCA model and non-optimization, respectively. The experimental results show that the saliency map generated by the CDCA model is closer to the ground truth with a clearer edge. Table 3 demonstrates that the CDCA model helps improve the performance of saliency detection. The weighted F-measure considers the correlation between the pixels of the saliency map, while the saliency map that has not been optimized by the CDCA model contains more redundant information, which increases the correlation between pixels.

Figure 7 shows that the depth feature can correct the saliency map when there are complex colors in the scenes (rows 1 and 2). When the depth image cannot highlight the salient objects (rows 3 and 4), the color contrast cue can be complemented with the depth, resulting in a clearer salient result. Through quantitative and qualitative analyses, the CDCA model

can improve the accuracy of saliency detection by utilizing the complementarity of color and depth.

Table 3. The effectiveness of the CDCA model in the proposed method on the LFSD dataset (E_β), S-measure (S_α), WF-measure (F_β^ω), F-measure (F_β) ,and MAE(M)) (bold: best).

Method	E_β	S_α	F_β^ω	F_β	M
Compactness	0.801	0.700	0.577	0.743	0.178
w/o CDCA	0.844	0.797	**0.721**	0.831	0.125
+ CDCA	**0.847**	**0.812**	0.720	**0.840**	**0.124**

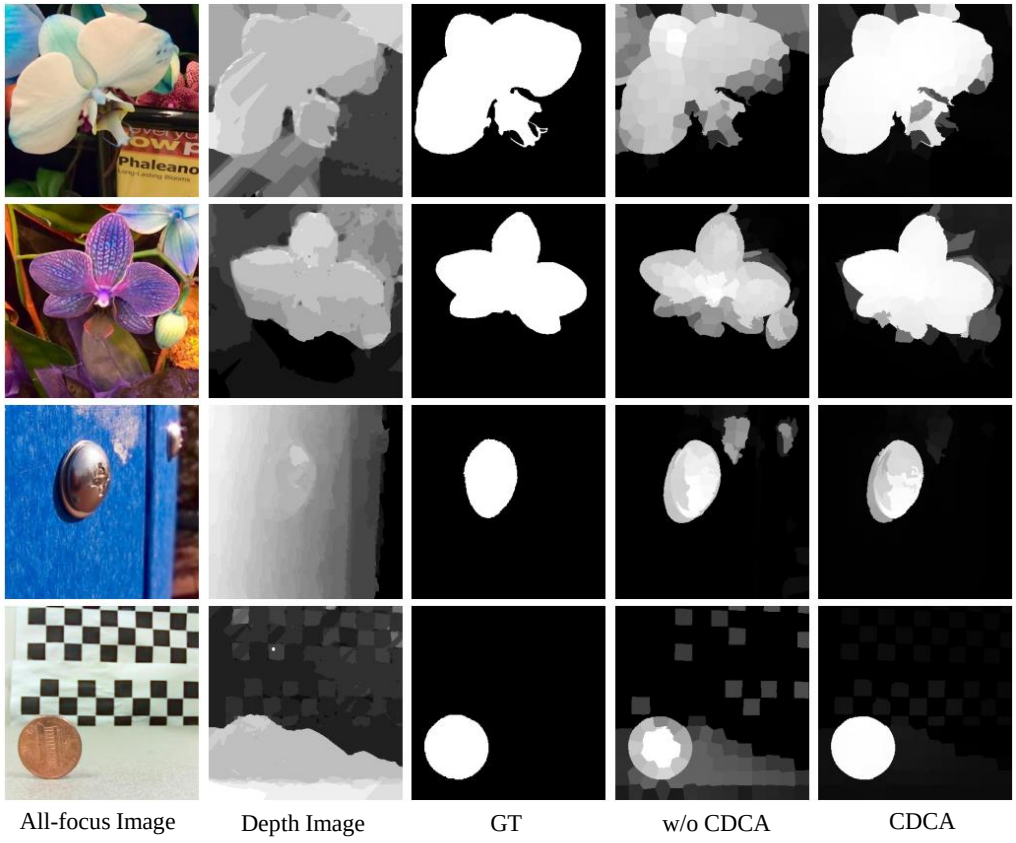


Figure 7. The comparison of the saliency maps after optimization by the proposed CDCA model and without the CDCA model on the LFSD dataset.

To verify the superiority of the CDCA model and explain the difference between the proposed CDCA model and the SCA [29] model, we compare the corresponding saliency map (generated by LFS [8], DILF [10], WSC [9], and RDFD [14]) optimized by the CDCA model and the SCA model on the LFSD [8] dataset. The suffix with CDCA in Table 4 denotes the saliency map optimized by the CDCA model. As shown in Table 4, the LFS method has the most obvious effect after optimization; the F-measure increased by 10.6%, and the MAE decreased by 29.68%. The CDCA model, which adds color and depth cues to increase spatial consistency, optimizes the saliency maps with a higher precision. Therefore, the CDCA model is suitable for light field images and has a strong generalization.

Table 4. Performance comparison between the proposed method and the light field saliency detection methods after the CDCA optimization (F-measure (F_β) (bold: best).

Method	Ours	LFS	LFS-CDCA	DILF	DILF-CDCA	WSC	WSC-CDCA	RDFD	RDFD-CDCA
F_β	**0.840**	0.725	0.802	0.787	0.814	0.735	0.800	0.792	0.820
MAE	**0.124**	0.219	0.154	0.149	0.149	0.163	0.154	0.152	0.147

Figure 8 shows the performance comparison of the existing light field saliency detection methods after optimization of the SCA [29] and CDCA models on the LFSD [8] dataset. The PR curve shows that the CDCA model achieves a higher PR curve compared to the SCA model. Figure 9 shows the visual comparison of the optimized saliency map between the CDCA and SCA models on the LFSD [8] dataset. It is observed that when the background color is similar to the salient object, the CDCA model can optimize the more accurate salient detection results because of the depth feature. In the case of a reliable depth image, the CDCA model can obtain more accurate salient edges. It is undeniable that the CDCA model is susceptible to the impact of poor-quality depth images. However, combining color and depth information compensates for the negative impact of the depth image quality to a certain extent.

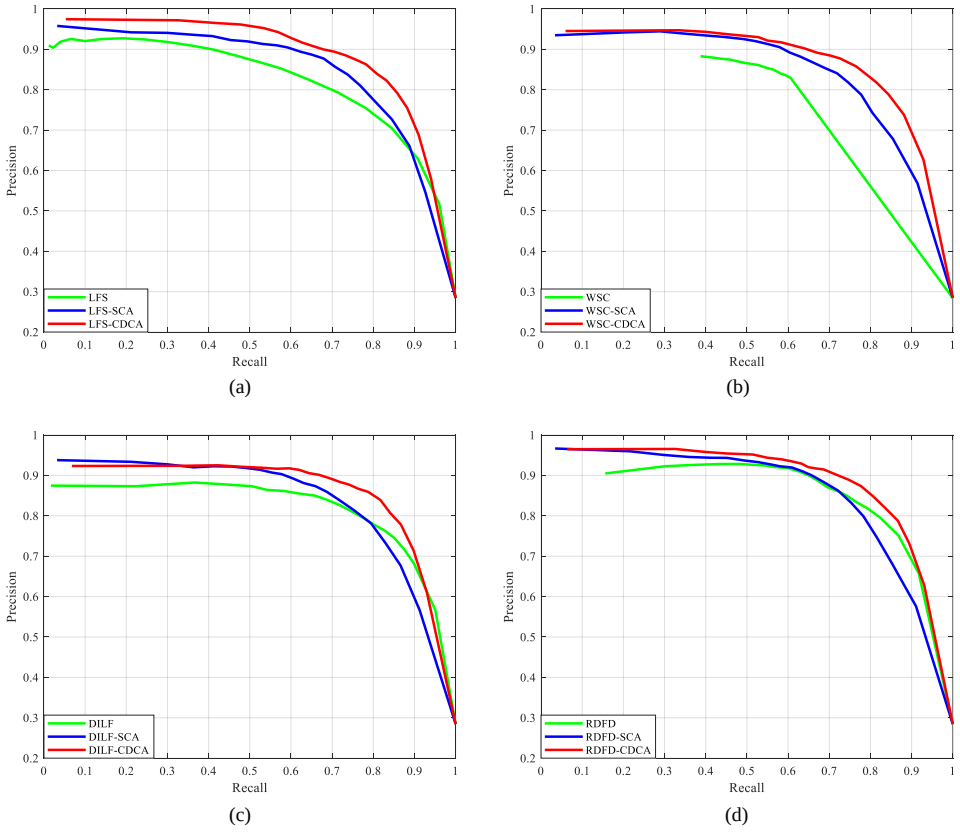


Figure 8. The comparison of the PR curves after optimization by the proposed CDCA model and the SCA model on the LFSD dataset. (**a**) LFS method. (**b**) WSC method. (**c**) DILF method. (**d**) RDFD method.

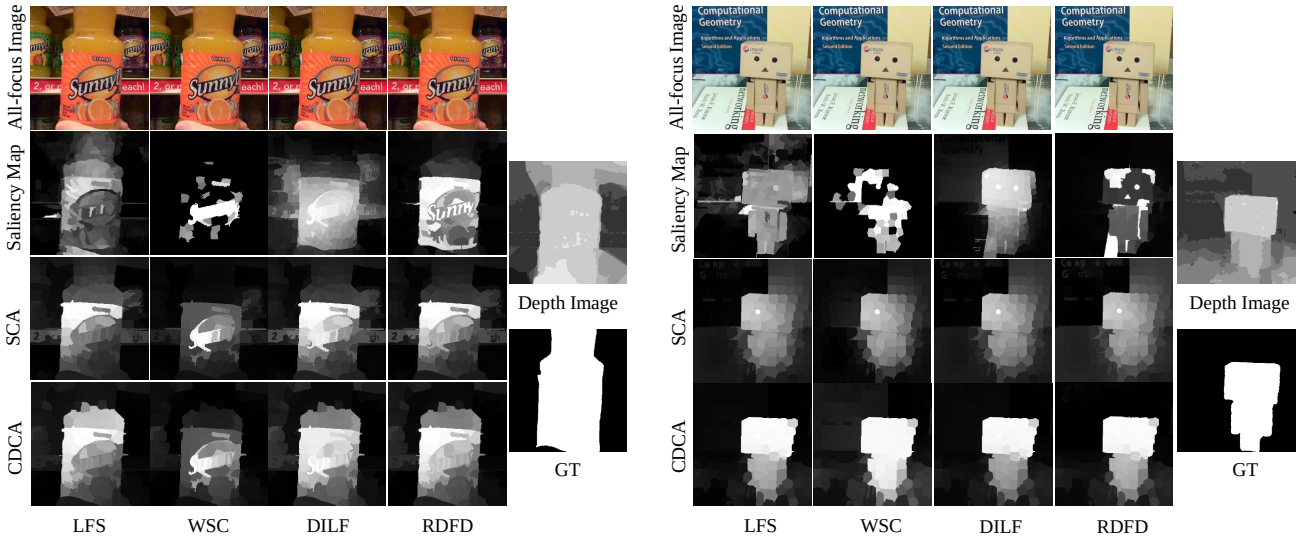


Figure 9. The visual comparison of the saliency maps after optimization by the proposed CDCA model and the SCA model on the LFSD dataset.

4.4.2. The Effectiveness of the Focal Stack and Depth

To verify the effectiveness of interaction and complementarity between different light field cues, we evaluate a variant of the proposed method by sequentially joining the focal stack and the depth map on the LFSD [8] dataset, as shown in Table 5. The focal stack and depth image have improved all indicators, and it is also proved that the CDCA model can obtain higher-precision saliency detection results.

Table 5. Performance comparison of each component in the whole algorithm where FocalStack+ represents the contribution of salient features in the focal stack and Depth+ represents the contribution of depth cue to the model. ((E-measure (E_β), S-measure (S_α), WF-measure (F_β^ω), F-measure (F_β), and MAE(M)) (bold: best).

Settings	E_β	S_α	F_β^ω	F_β	M
Compactness	0.801	0.700	0.577	0.743	0.178
FocalStack+	0.829	0.764	0.678	0.807	0.142
Depth+	0.791	0.746	0.624	0.747	0.162
FocalStack *+	0.828	0.791	0.697	0.811	0.134
Depth *+	0.834	0.794	0.674	0.828	0.141
MAE Filter	**0.847**	**0.812**	**0.720**	**0.840**	**0.124**

* Optimized by CDCA model.

Figure 10 shows the visual comparison between the proposed method and others on the LFSD [8] dataset. The foreground and background in the all-focus images are relatively similar (rows 2, 3, 6, 8), and the salient objects can be effectively identified with the supplement of depth information. The depth images in the fourth and fifth rows can easily mislead detection, while focus and color cues can play a corrective role.

As illustrated in Figure 11, we also exhibit some failures brought by the proposed method. The performance of the method is partially dependent on the accuracy of the depth map and the focus region of the focal stack. If the depth map is seriously blurred or amorphous, it yields incorrect results. If the foreground and background are similar in color and disorderly, it is necessary to rely on depth compactness and focus to extract and detect salient objects. Therefore, obtaining an accurate depth image as well as a perfect focal region of the focal stack remains a challenging problem.

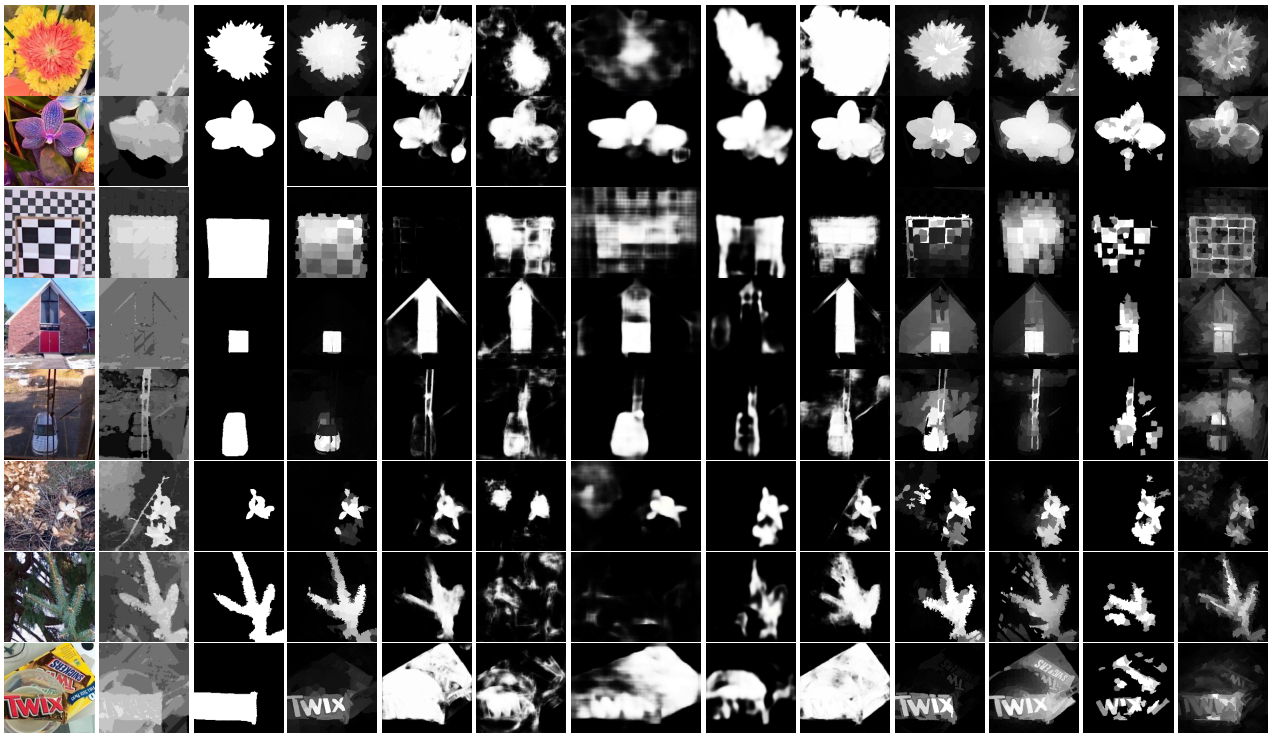


Figure 10. The visual comparison of our saliency maps with other state-of-the-art methods on LFSD dataset.

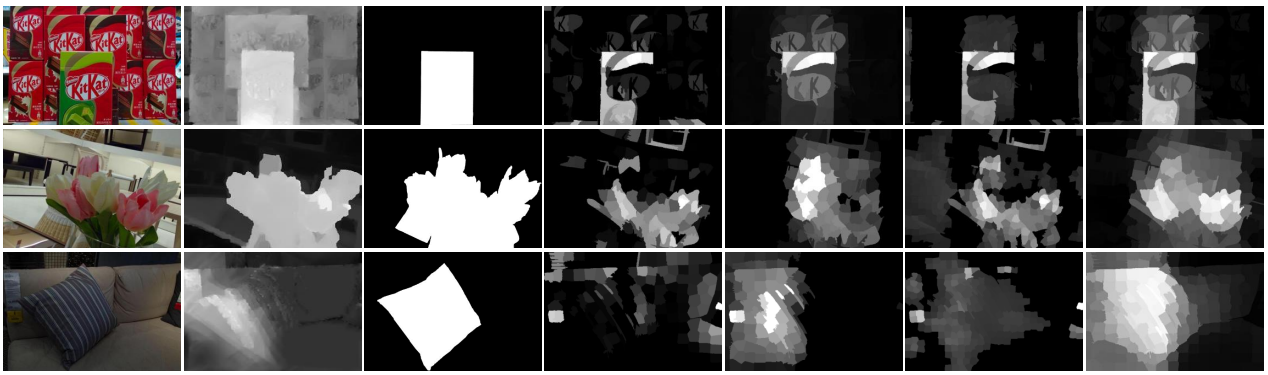


Figure 11. Some failure results by the proposed method.

5. Conclusions

In this paper, we propose a light field saliency detection method based on focus and depth, which explores the interaction and complementarity among focus, color, and depth cues of the light field to improve the saliency detection performance. Firstly, coarse salient regions are localized by combining the compactness of color and depth. Then, the interplay of depth and focus information is used to highlight the foreground and suppress the background. At the same time, the local depth cue is used to enhance the global features to refine the salient map. Secondly, inspired by spatial consistency, we utilize the complementarity of color and depth information to improve previous optimization models, resulting in remarkable results with higher accuracy. Finally, to avoid the influence of image quality, we design an output model with the MAE as the screening index.

The proposed method can obtain high-quality saliency detection results with lower computational cost by deeply exploring different cues of the light field. According to

the comprehensive comparison of public datasets and ablation experiments, it is proved that the proposed method is far superior to the conventional light field saliency detection methods and even better than some state-of-the-art methods based on deep learning.

Author Contributions: Y.Z. designed and completed the algorithm and drafted the manuscript. F.C. polished the manuscript. Z.P. gave professional guidance and reviewed the manuscript. W.Z. and C.Z. proofread the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant No. 623710812, the Natural Science Foundation of Chongqing under Nos. cstc2021jcyj-msxmX0411 and CSTB2022NSCQ-MSX0873, the Science and Technology Research Program of the Chongqing Municipal Education Commission under No. KJZD-K202001105, and the Scientific Research Foundation of the Chongqing University of Technology under Nos. 2020zdz029 and 2020zdz030.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jeon, H.G.; Park, J.; Choe, G.; Park, J.; Bok, Y.; Tai, Y.W.; So Kweon, I. Accurate depth map estimation from a lenslet light field camera. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1547–1555.
2. Shin, C.; Jeon, H.G.; Yoon, Y.; Kweon, I.S.; Kim, S.J. Epinet: A fully-convolutional neural network using epipolar geometry for depth from light field images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4748–4757.
3. Jin, J.; Hou, J.; Yuan, H.; Kwong, S. Learning light field angular super-resolution via a geometry-aware network. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 11141–11148.
4. Wang, S.; Zhou, T.; Lu, Y.; Di, H. Detail-preserving transformer for light field image super-resolution. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual Event, 22 February–1 March 2022; Volume 36, pp. 2522–2530.
5. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [PubMed]
6. Fan, D.P.; Ji, G.P.; Sun, G.; Cheng, M.M.; Shen, J.; Shao, L. Camouflaged object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 2777–2787.
7. Wei, Y.; Xia, W.; Huang, J.; Ni, B.; Dong, J.; Zhao, Y.; Yan, S. CNN: Single-label to multi-label. *arXiv* **2014**, arXiv:1406.5726.
8. Li, N.; Ye, J.; Ji, Y.; Ling, H.; Yu, J. Saliency detection on light field. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2806–2813.
9. Li, N.; Sun, B.; Yu, J. A weighted sparse coding framework for saliency detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 5216–5223.
10. Zhang, J.; Wang, M.; Gao, J.; Wang, Y.; Zhang, X.; Wu, X. Saliency Detection with a Deeper Investigation of Light Field. In Proceedings of the IJCAI, Buenos Aires, Argentina, 25–31 July 2015; pp. 2212–2218.
11. Wang, A.; Wang, M.; Li, X.; Mi, Z.; Zhou, H. A two-stage Bayesian integration framework for salient object detection on light field. *Neural Process. Lett.* **2017**, *46*, 1083–1094.
12. Zhang, J.; Wang, M.; Lin, L.; Yang, X.; Gao, J.; Rui, Y. Saliency detection on light field: A multi-cue approach. *ACM Trans. Multimed. Comput. Commun. Appl.* (*TOMM*) **2017**, *13*, 1–22.
13. Piao, Y.; Li, X.; Zhang, M.; Yu, J.; Lu, H. Saliency detection via depth-induced cellular automata on light field. *IEEE Trans. Image Process.* **2019**, *29*, 1879–1889. [CrossRef] [PubMed]
14. Wang, X.; Dong, Y.; Zhang, Q.; Wang, Q. Region-based depth feature descriptor for saliency detection on light field. *Multimed. Tools Appl.* **2021**, *80*, 16329–16346.
15. Piao, Y.; Rong, Z.; Zhang, M.; Li, X.; Lu, H. Deep Light-field-driven Saliency Detection from a Single View. In Proceedings of the IJCAI, Macao, China, 10–16 August 2019; pp. 904–911.
16. Zhang, J.; Liu, Y.; Zhang, S.; Poppe, R.; Wang, M. Light field saliency detection with deep convolutional networks. *IEEE Trans. Image Process.* **2020**, *29*, 4421–4434.
17. Wang, A. Three-stream cross-modal feature aggregation network for light field salient object detection. *IEEE Signal Process. Lett.* **2020**, *28*, 46–50. [CrossRef]
18. Liang, Z.; Wang, P.; Xu, K.; Zhang, P.; Lau, R.W. Weakly-supervised salient object detection on light fields. *IEEE Trans. Image Process.* **2022**, *31*, 6295–6305. [CrossRef] [PubMed]
19. Jiang, Y.; Zhang, W.; Fu, K.; Zhao, Q. MEANet: Multi-modal edge-aware network for light field salient object detection. *Neurocomputing* **2022**, *491*, 78–90. [CrossRef]

20. Yuan, B.; Jiang, Y.; Fu, K.; Zhao, Q. Guided Focal Stack Refinement Network for Light Field Salient Object Detection. *arXiv* **2023**, arXiv:2305.05260.
21. Fan, D.P.; Lin, Z.; Zhang, Z.; Zhu, M.; Cheng, M.M. Rethinking RGB-D salient object detection: Models, data sets, and large-scale benchmarks. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 2075–2089. [CrossRef]
22. Liu, N.; Zhang, N.; Han, J. Learning selective self-mutual attention for RGB-D saliency detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 13756–13765.
23. Cong, R.; Lei, J.; Fu, H.; Hou, J.; Huang, Q.; Kwong, S. Going from RGB to RGBD saliency: A depth-guided transformation model. *IEEE Trans. Cybern.* **2019**, *50*, 3627–3639. [CrossRef]
24. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [CrossRef]
25. Cheng, M.M.; Mitra, N.; Huang, X.; Torr, P.; Hu, S.M. Salient Object Detection and Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *37*, 569–582. [CrossRef]
26. Yang, C.; Zhang, L.; Lu, H.; Ruan, X.; Yang, M.H. Saliency detection via graph-based manifold ranking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 3166–3173.
27. Zhu, W.; Liang, S.; Wei, Y.; Sun, J. Saliency optimization from robust background detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2814–2821.
28. Zhou, L.; Yang, Z.; Yuan, Q.; Zhou, Z.; Hu, D. Salient region detection via integrating diffusion-based compactness and local contrast. *IEEE Trans. Image Process.* **2015**, *24*, 3308–3320. [CrossRef]
29. Qin, Y.; Lu, H.; Xu, Y.; Wang, H. Saliency detection via cellular automata. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 110–119.
30. Luo, Z.; Mishra, A.; Achkar, A.; Eichel, J.; Li, S.; Jodoin, P.M. Non-local deep features for salient object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6609–6617.
31. Wang, W.; Zhao, S.; Shen, J.; Hoi, S.C.; Borji, A. Salient object detection with pyramid attention and salient edges. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 1448–1457.
32. Chen, Z.; Xu, Q.; Cong, R.; Huang, Q. Global context-aware progressive aggregation network for salient object detection. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 10599–10606.
33. Zhou, T.; Wang, S.; Zhou, Y.; Yao, Y.; Li, J.; Shao, L. Motion-attentive transition for zero-shot video object segmentation. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 13066–13073.
34. Lai, Q.; Zhou, T.; Khan, S.; Sun, H.; Shen, J.; Shao, L. Weakly supervised visual saliency prediction. *IEEE Trans. Image Process.* **2022**, *31*, 3111–3124.
35. Niu, Y.; Geng, Y.; Li, X.; Liu, F. Leveraging stereopsis for saliency analysis. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 454–461.
36. Peng, H.; Li, B.; Xiong, W.; Hu, W.; Ji, R. RGBD salient object detection: A benchmark and algorithms. In Proceedings of the Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; Proceedings, Part III 13; Springer: Cham, Switzerland, 2014; pp. 92–109.
37. Ren, J.; Gong, X.; Yu, L.; Zhou, W.; Ying Yang, M. Exploiting global priors for RGB-D saliency detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, USA, 7–12 June 2015; pp. 25–32.
38. Cong, R.; Lei, J.; Zhang, C.; Huang, Q.; Cao, X.; Hou, C. Saliency detection for stereoscopic images based on depth confidence analysis and multiple cues fusion. *IEEE Signal Process. Lett.* **2016**, *23*, 819–823. [CrossRef]
39. Zhu, C.; Li, G.; Wang, W.; Wang, R. An innovative salient object detection using center-dark channel prior. In Proceedings of the IEEE international Conference on Computer Vision Workshops, Venice, Italy, 22–29 October 2017; pp. 1509–1515.
40. Zhu, C.; Cai, X.; Huang, K.; Li, T.H.; Li, G. PDNet: Prior-model guided depth-enhanced network for salient object detection. In Proceedings of the 2019 IEEE International Conference on Multimedia and Expo (ICME), Shanghai, China, 8–12 July 2019; pp. 199–204.
41. Zhao, Y.; Zhao, J.; Li, J.; Chen, X. RGB-D salient object detection with ubiquitous target awareness. *IEEE Trans. Image Process.* **2021**, *30*, 7717–7731. [CrossRef]
42. Wang, T.; Piao, Y.; Li, X.; Zhang, L.; Lu, H. Deep learning for light field saliency detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea (South), 27 October–2 November 2019; pp. 8838–8848.
43. Zhang, M.; Ji, W.; Piao, Y.; Li, J.; Zhang, Y.; Xu, S.; Lu, H. LFNet: Light field fusion network for salient object detection. *IEEE Trans. Image Process.* **2020**, *29*, 6276–6287.
44. Zhang, Q.; Wang, S.; Wang, X.; Sun, Z.; Kwong, S.; Jiang, J. A multi-task collaborative network for light field salient object detection. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *31*, 1849–1861. [CrossRef]
45. Feng, M.; Liu, K.; Zhang, L.; Yu, H.; Wang, Y.; Mian, A. Learning from pixel-level noisy label: A new perspective for light field saliency detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 1756–1766.
46. Margolin, R.; Zelnik-Manor, L.; Tal, A. How to evaluate foreground maps? In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 248–255.

47. Fan, D.P.; Gong, C.; Cao, Y.; Ren, B.; Cheng, M.M.; Borji, A. Enhanced-alignment measure for binary foreground map evaluation. *arXiv* **2018**, arXiv:1805.10421.
48. Fan, D.P.; Cheng, M.M.; Liu, Y.; Li, T.; Borji, A. Structure-measure: A new way to evaluate foreground maps. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4548–4557.