information

Article

Learning Static-Adaptive Graphs for RGB-T Image Saliency Detection

Zhengmei Xu [1,†], Jin Tang [2,†], Aiwu Zhou [3,†] and Huaming Liu [1,*]

[1] School of Computer and Information Engineering, Fuyang Normal University, Fuyang 236037, China; 200107032@fynu.edu.cn
[2] Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University, Hefei 230601, China; tj@ahu.edu.cn
[3] School of Computer Science and Technology, Anhui University, Hefei 230601, China; jyang@iim.ac.cn
* Correspondence: 200806004@fynu.edu.cn
† These authors contributed equally to this work.

Abstract: Many works have been proposed on image saliency detection to handle challenging issues including low illumination, cluttered background, low contrast, and so on. Although good performance has been achieved by these algorithms, detection results are still poor based on RGB modality. Inspired by the recent progress of multi-modality fusion, we propose a novel RGB-thermal saliency detection algorithm through learning static-adaptive graphs. Specifically, we first extract superpixels from the two modalities and calculate their affinity matrix. Then, we learn the affinity matrix dynamically and construct a static-adaptive graph. Finally, the saliency maps can be obtained by a two-stage ranking algorithm. Our method is evaluated on RGBT-Saliency Dataset with eleven kinds of challenging subsets. Experimental results show that the proposed method has better generalization performance. The complementary benefits of RGB and thermal images and the more robust feature expression of learning static-adaptive graphs create an effective way to improve the detection effectiveness of image saliency in complex scenes.

Keywords: RGB-thermal; static-adaptive graph; manifold ranking; saliency detection

check for updates

Citation: Xu, Z.; Tang, J.; Zhou, A.; Liu, H. Learning Static-Adaptive Graphs for RGB-T Image Saliency Detection. *Information* **2022**, *13*, 84. https://doi.org/10.3390/info13020084

Academic Editor: Gholamreza Anbarjafari

Received: 25 December 2021
Accepted: 10 February 2022
Published: 12 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Image saliency detection aims to quickly capture the most important and useful information from a scene by using the human visual attention mechanism, which can reduce the complexity of subsequent image processing, and has been applied to numerous vision problems including image classification [1], image retrieval [2], image encryption [3,4], video summary [5], and so on. In the past few decades, researchers have proposed many saliency detection algorithms, which can be divided into bottom-up data-driven models and top-down task-driven methods. Bottom-up models [6–9] take the underlying image features and some priors into consideration, such as color, texture, orientation, and brightness. Itti et al. [10] proposed a visual attention mechanism, which opened research on saliency detection in the field of computer vision. Cheng et al. [11] introduced a regional contrast-based salient object detection algorithm, which simultaneously evaluates global contrast differences and spatial weighted coherence scores. Wang et al. [12] improved the detection effect of image saliency by optimizing seeds. Top-down models [13,14] are task driven. They use a large amount of training data with category labels and supervised learning to conduct a task-oriented analysis. Recently, most of these methods are based on deep learning, they have better performance, but their training processes are time-consuming. We focus on the bottom-up models. Many scholars have made many attempts to improve image saliency detection and have obtained good performance in simple scenes. However, the effectiveness of traditional RGB saliency detection methods decreases sharply in complex scenes, such as poor lighting or saliency objects that have the same color and texture as the background.

In recent years, multi-source sensor technology has become popular in image processing. Li et al. [15–17] simultaneously extracted RGB and thermal features for tracking, which effectively improved the effect of video target tracking at night or in rainy, hazy, and foggy weather. Zhang et al. [18] extracted the depth features of RGB images and thermal images, and then fused the two extracted features for saliency detection, which greatly improved the detection effectiveness in the case of poor illumination or similar color and texture to the background. The fusion of RGB and thermal images is proven to be effective in image saliency detection. RGB images can provide texture details with high definition in a manner consistent with the human visual system in simple scenes. By contrast, thermal images can work well in low illumination, and also have good discrimination when the target and the background have similar colors or shapes RGB-T saliency detection algorithms can obtain better results by handling challenging issues including low illumination, cluttered background, low contrast, and so on. Graph-based models [19–21] use pixels or superpixels as nodes and the similarity weight between nodes as the edge to generate the graph, which can achieve a great structure character from initial input images for RGB-T saliency detection results. However, the existing graph-based fusion models only use the static graph. The limitation of this kind of method is that it cannot explore the relationship between nodes at the target level and gain better fusion of multi-modality information. Inspired by these methods, we consider the spatial connectivity feature of graph nodes to learn a static-adaptive graph, and propose a novel RGB-thermal saliency detection algorithm to obtain more effective results, as in Figure 1.

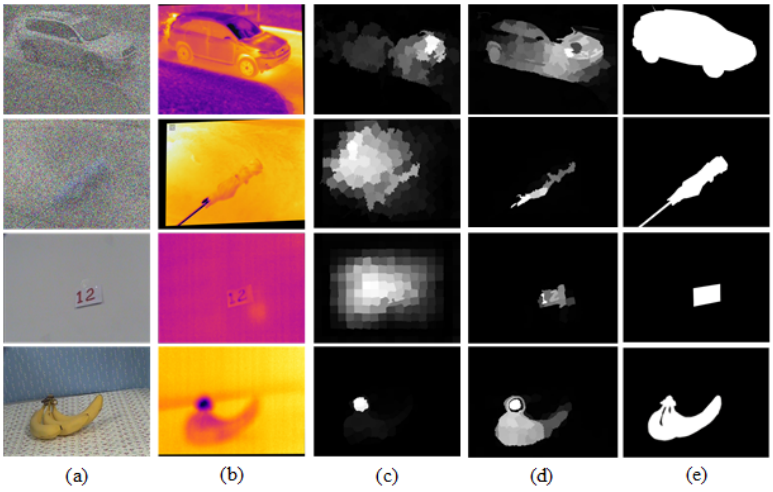

Figure 1. Comparative results of static-adaptive graph-based method with traditional static graph model. (**a**) RGB image; (**b**) thermal image; (**c**) the saliency map generated by the static graph-based model; (**d**) the saliency map generated by our model; (**e**) ground truth.

Specifically, we first extract superpixels from the two modalities and calculate their affinity matrix. Then, we learn the affinity matrix dynamically and construct a static-adaptive graph. Finally, the saliency maps can be obtained by a two-stage ranking algorithm. The contributions of this paper are summarized as follows.

- We construct an adaptive graph by sparse representation and carry out the optimization solution;
- We learn a novel static-adaptive graph model to increase the fusion capacity by considering the spatial connectivity features of graph nodes in RGB-T saliency detection;
- We compare our method with the state-of-the-art methods on an RGB-T dataset with 11 kinds of challenging subsets. The experimental results verify the effectiveness of our method.

2. Related Work

In this section, we give a brief review of methods closely related to our work. The relevant work in this paper mainly includes the graph-based method, multi-modality fusion method, and subspace-based method.

Graph-based method. In the past few decades, graph-based models have been successfully used for saliency detection problems. Harel et al. [19] proposed a graph model. The algorithm takes the pixel points as the graph nodes, constructs edges between the pixel points according to the spatial distance and feature distance, and uses Markov random field for feature fusion. Yang et al. [20] proposed a manifold ranking algorithm based on a static graph, which is a typical two-stage model to gain more accurate saliency maps. Jiang et al. [21] calculated a preliminary saliency map by Markov absorption probability on a weighted graph via partial image borders as prior background. Zhang et al. [22] used multi-scale to improve the manifold ranking algorithm. Xiao et al. [23] proposed a prior regularized multi-layer graph ranking model in which they used the prior calculating by boundary connectivity. Aytekin et al. [24] proposed a graph model that uses a convolution kernel function network to learn the connection weight coefficients.

Multi-modality fusion method. In recent years, with the development of multi-sensors, multi-modality fusion has become a new effective means to improve computer vision tasks. Li et al. [25] combined gray and thermal information to deal with target tracking in complex scenes. Li et al. [15] used multispectral (RGB and thermal) data to improve visual tracking effectiveness. Li et al. [26] established a unified RGB-T dataset and proposed a new algorithm to fuse RGB and thermal images for saliency detection, which incorporates the cross-modality consistent constraints to integrate different modalities collaboratively. RGB-D is an effective multi-modal fusion method in many aspects, such as manufacturing [27], semantic segmentation [28–30], and saliency detection [31,32]. Liu et al. [33] used three transformer encoders with shared weights to enhance multi-level features, and the algorithm they proposed effectively improves the effectiveness of saliency detection.

Subspace-based method. Subspace-based methods represent high-dimensional data in low-dimensional subspace. The purpose of subspace representation is to obtain a similarity matrix in the basic subspace of the original data. In a dataset, each data point can be reconstructed by an effective combination of other points, which are often helpful for data processing, because data can better reflect the characters of data in its low-dimensional subspace. Guo et al. [34] proposed a subspace segmentation method to jointly learn data representation and affinity matrix relationships simultaneously in a model. Li et al. [35] represented each patch with a linear combination of the remaining ones and learned the weights of the global and local features of the detection object, achieving good effectiveness in the application field of video tracking.

We learn static-adaptive graphs for saliency detection. The static graph is the traditional graph. Its structure is fixed, and it only considers the relationship between adjacent nodes. The adaptive graph is obtained by the subspace method to mine the internal relationship between superpixels. Therefore, our algorithm considers both local and global features, and has better effectiveness than the saliency detection algorithm which is only based on the static graph. In multi-modality selection, we fuse RGB image and thermal image, because RGB and thermal images have natural complementarity. Compared with the RGB-D saliency detection algorithm, the RGB-T saliency detection algorithm has much lower hardware requirements for computers, and can run well on computer with an i3 3.3G CPU and 4GB RAM.

3. Brief Review of Manifold Ranking

A manifold ranking (MR) model [20] is a typical graph-based method for saliency detection. For an image, simple linear iterative clustering (SLIC) [36] is always used to obtain n superpixels as graph nodes in most of these models. Take a graph $G = (V, E)$, where V is a node set. Some of nodes are labeled as queries and the rest need to be ranked according to their relevance to the queries. Let $\mathbf{X} = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{d \times n}$ be the character

matrix, d the dimensionality of the feature vector, and n the number of the superpixels. E is the set of undirected edges and $\mathbf{W}_{ij}$ is the edge weight between node i and node j that can be calculated by feature vectors of two nodes. Let $\mathbf{q} = [q_1, q_2, \ldots, q_n]^T$ denote an indication vector, where $q_i = 1$ if node i is a labeled query, otherwise, $q_i = 0$. The aim of MR is to gain a ranking value f_i for each graph node, which can be computed by solving Equation (1),

$$\min_f \frac{1}{2}\left(\sum_{i,j=1}^{n} \mathbf{W}_{ij}\left\|\frac{f_i}{\sqrt{\mathbf{D}_{ii}}} - \frac{f_j}{\sqrt{\mathbf{D}_{jj}}}\right\|^2 + \mu\sum_{i=1}^{n}\|f_i - q_i\|^2\right) \tag{1}$$

where $\mathbf{D}_{ii} = \sum_{j=1}^{n} \mathbf{W}_{ij}$.

To obtain more effective results, Yang et al. [20] obtained the ranking value by using the un-normalized Laplacian matrix in Equation (2),

$$\mathbf{f} = (\mathbf{D} - \lambda\mathbf{W})^{-1}\mathbf{q}, \tag{2}$$

where $\mathbf{D}$ is a degree matrix, $\mathbf{D} = diag\{\mathbf{D}_{11}, \ldots, \mathbf{D}_{nn}\}$, $\lambda = 1/(1+\mu)$.

4. Static-Adaptive Graph Learning

4.1. Static-Adaptive Graph Construction

The graph of traditional models is static; most of them only consider adjacent nodes and boundary nodes. The limitation of this kind of method is that it cannot explore the relationship between nodes at the target level. Therefore, we consider the spatial connectivity features of graph nodes to construct a static-adaptive graph, in which superpixels with similar features in the region are also connected. Take multiple graphs $G^m = (V^m, E^m), m = 1, 2, \ldots, M$, where V^m is a node set, and E^m is the set of undirected edges. Let $\mathbf{X}^m = [x_1^m, x_2^m, \ldots, x_N^m] \in \mathbb{R}^{d\times N}, m = 1, 2, \ldots, M$ be the character matrix of the m-th modality. N is the number of graph nodes. d is the dimensionality of the feature vector. As in traditional static graphs [20], when two nodes meet one of the following three conditions, they are considered to have edges.

(1) Two nodes are directly adjacent;
(2) There is a common edge between two nodes;
(3) Superpixels are on the four boundaries.

If there is an edge between two nodes, the weight of the edge is calculated by Equation (3).

$$\mathbf{W}_{i,j}^m = e^{-\gamma_0\|x_i^m - x_j^m\|}, m = 1, 2, \ldots, M, \tag{3}$$

where x_i^m denotes the mean of the i-th superpixel in the m-th modality, and γ_0 is a parameter.

We add the adaptive graph weight matrix to gain the weight matrix of the static-adaptive graph as in Figure 2, which can be calculated by Equation (4).

$$\mathbf{W} = \mathbf{W}^a + \sum_{m=1}^{M} t_m\mathbf{W}^m, \tag{4}$$

where $\mathbf{W}^a$ is the weight matrix of adaptive graph, which can be obtained by adaptive graph learning.

$\mathbf{W}^m = [\mathbf{W}_{ij}^m]_{N\times N}, m = 1, 2, \ldots, M$ is the initial weight matrix of the m-th modality. t_m can indicate the importance of different modalities of static and adaptive graphs.

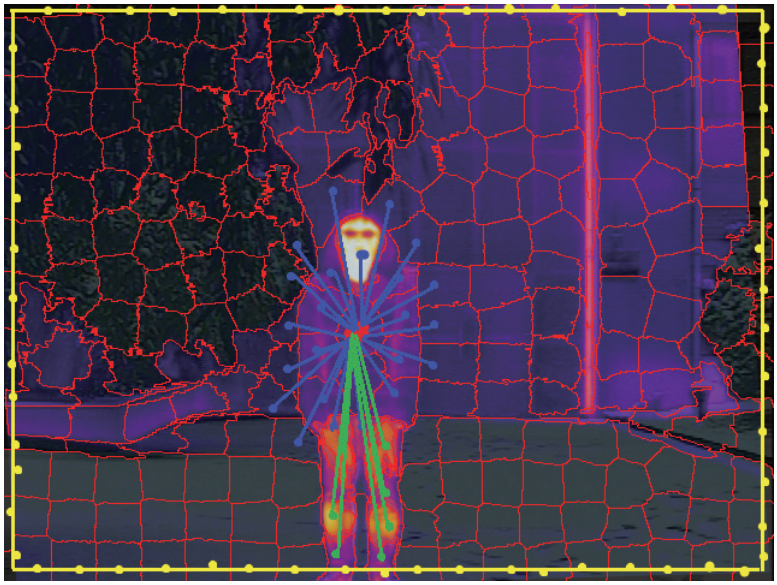

Figure 2. The general view of the static-adaptive graph on the multi-modality fusion image. The blue edges are obtained by the traditional static graph. The green edges are obtained by our adaptive graph learning model.

4.2. Adaptive Graph Learning Model Formulation

For M graphs $G^m = (V^m, E^m), m = 1, 2, \ldots, M$, we assume that all nodes in each graph belong to the same sparse subspace, in which each node can be sparsely represented by the remaining nodes. We can obtain $\mathbf{X}^m = \mathbf{X}^m\mathbf{Z}^m, m = 1, 2, \ldots, M$, where $\mathbf{Z}^m \in \mathbb{R}^{N\times N}$ is the sparse coefficient matrix. Sparse constraints can automatically select most informative neighbor nodes for each node, and make the graph more powerful. Since the nodes are often disturbed by noises, we introduce a noise matrix $\mathbf{E}^m \in \mathbb{R}^{d\times N}$ to improve the robustness. The joint sparse representation with the convex relaxation for all modalities can be written as,

$$\min_{\mathbf{Z},\mathbf{E}^m} \alpha\|\mathbf{Z}\|_1 + \beta \sum_{m=1}^{M} \|\mathbf{E}^m\|_{2,1}, \ \ s.t.\ \mathbf{X}^m = \mathbf{X}^m\mathbf{Z}^m + \mathbf{E}^m. \tag{5}$$

where α and β are balanced parameters. $\mathbf{Z} = [\mathbf{Z}^1; \cdots; \mathbf{Z}^M] \in \mathbb{R}^{N\times(M*N)}$ is the joint sparse representation coefficient matrix.

We consider the spatial connectivity feature of graph nodes and use $\mathbf{C} \in \mathbb{R}^{N\times N}$ to indicate the spatial connections of neighboring nodes.

If node i and j are 8-neighboring, $\mathbf{C}_{ij} = 1$; otherwise $\mathbf{C}_{ij} = 0$.

$$\mathbf{C}_{ij} = \begin{cases} 1, & if\ i\ and\ j\ are\ 8\text{-}neighboring, \\ 0, & else. \end{cases} \tag{6}$$

The closer the distance, the greater the relevance. Inspired by [35], to capture the global and local structure information, we employ Equation (7) to learn the adaptive graph affinity matrix.

$$\begin{aligned} &\min_{\mathbf{W}^a} \frac{\gamma}{2} \sum_{i,j=1}^{N} \|\mathbf{Z}_i - \mathbf{Z}_j\|_F^2 \mathbf{W}_{ij}^a + \frac{\delta}{2} \sum_{i,j=1}^{N} \mathbf{C}_{ij}\|\mathbf{Z}_i - \mathbf{Z}_j\|_F^2 + \lambda_1\|\mathbf{W}^a\|_F^2, \\ &s.t.\ \mathbf{W}^{aT}\mathbf{1} = \mathbf{1}, \mathbf{W}^a \geq \mathbf{0}. \end{aligned} \tag{7}$$

where γ and δ are the balancing parameters. The first item reflects the probability $\mathbf{W}_{ij}^a$ from the same cluster based on the distance between their representations $\mathbf{Z}_i$ and $\mathbf{Z}_j$. The second item indicates that two close nodes will have similar representations. $\lambda_1\|\mathbf{W}^a\|_F^2$ is

to avoid over-fitting of $\mathbf{W}^a$. $\mathbf{1}$ denotes a unit vector. ${\mathbf{W}^a}^T\mathbf{1} = \mathbf{1}, \mathbf{W}^a \geq \mathbf{0}$ are constraints to guarantee the probability property of $\mathbf{W}^a_{ij}$. We combine the Equations (5) and (7) and obtain the following optimal function,

$$\begin{aligned} &\min_{\mathbf{Z},\mathbf{E}^m,\mathbf{W}^a} \alpha\|\mathbf{Z}\|_1 + \frac{\gamma}{2}\sum_{i,j=1}^{N}\|\mathbf{Z}_i - \mathbf{Z}_j\|_F^2\mathbf{W}^a_{ij} + \frac{\delta}{2}\sum_{i,j=1}^{N}\mathbf{C}_{ij}\|\mathbf{Z}_i - \mathbf{Z}_j\|_F^2, \\ &+ \lambda_1\|\mathbf{W}^a\|_F^2 + \beta\sum_{m=1}^{M}\|\mathbf{E}^m\|_{2,1}, \\ &s.t.\ \mathbf{X}^m = \mathbf{X}^m\mathbf{Z}^m + \mathbf{E}^m, {\mathbf{W}^a}^T\mathbf{1} = \mathbf{1}, \mathbf{W}^a \geq \mathbf{0}. \end{aligned} \tag{8}$$

In order to solve the problem easily, let $\mathbf{D}^a_{ii} = \sum_{j=1}^{N}\mathbf{W}^a_{ij}$, $\mathbf{D}^c_{ii} = \sum_{j=1}^{N}\mathbf{C}_{ij}$. Equation (8) is a slightly algebraic transformation to,

$$\begin{aligned} &\min_{\mathbf{Z},\mathbf{E}^m,\mathbf{W}^a} \alpha\|\mathbf{Z}\|_1 + \gamma\, tr(\mathbf{Z}\mathbf{L}^a\mathbf{Z}^T) + \delta\, tr(\mathbf{Z}\mathbf{L}^c\mathbf{Z}^T) \\ &+ \lambda_1\|\mathbf{W}^a\|_F^2 + \beta\sum_{m=1}^{M}\|\mathbf{E}^m\|_{2,1}, \\ &s.t.\ \mathbf{X}^m = \mathbf{X}^m\mathbf{Z}^m + \mathbf{E}^m, {\mathbf{W}^a}^T\mathbf{1} = 1, \mathbf{W}^a \geq \mathbf{0}. \end{aligned} \tag{9}$$

where $\mathbf{L}^a = \mathbf{D}^a - \mathbf{W}^a$ and $\mathbf{L}^c = \mathbf{D}^c - \mathbf{C}$ are Laplacian matrices of $\mathbf{W}^a$ and $\mathbf{C}$, respectively.

4.3. Optimization

The variables in Equation (9) are not jointly convex; they are convex with respect to the subproblem of each variable when others are fixed and have a close form solution. We introduce two auxiliary variables, $\mathbf{P}^m$ and $\mathbf{Q}^m$, to make Equation (9) separable and then use the alternating direction multiplier (ADMM) algorithm [37] for optimization iteration. Then, we can obtain Equation (10).

$$\begin{aligned} &\min_{\mathbf{Z},\mathbf{E}^m,\mathbf{W}^a} \alpha\|\mathbf{Z}\|_1 + \gamma\, tr(\mathbf{Z}\mathbf{L}^a\mathbf{Z}^T) + \delta\, tr(\mathbf{Z}\mathbf{L}^c\mathbf{Z}^T) \\ &+ \lambda_1\|\mathbf{W}^a\|_F^2 + \beta\sum_{m=1}^{M}\|\mathbf{E}^m\|_{2,1}, \\ &s.t.\ \mathbf{P}^m = \mathbf{Z}^m, \mathbf{Q}^m = \mathbf{Z}^m, \mathbf{X}^m = \mathbf{X}^m\mathbf{Z}^m + \mathbf{E}^m, {\mathbf{W}^a}^T\mathbf{1} = 1, \mathbf{W}^a \geq \mathbf{0}. \end{aligned} \tag{10}$$

Thus, we obtain the Lagrange function [38] as Equation (11),

$$\begin{aligned} &\min_{\mathbf{Z},\mathbf{E}^m,\mathbf{W}^a,\mathbf{P},\mathbf{Q}} \alpha\|\mathbf{Q}\|_1 + \gamma\, tr(\mathbf{P}\mathbf{L}^a\mathbf{P}^T) + \delta\, tr(\mathbf{P}\mathbf{L}^c\mathbf{P}^T) + \lambda_1\|\mathbf{W}^a\|_F^2 \\ &+ \sum_{m=1}^{M}(\beta\|\mathbf{E}^m\|_{2,1} + \frac{\mu}{2}\|\mathbf{X}^m - \mathbf{X}^m\mathbf{Z}^m - \mathbf{E}^m + \frac{\mathbf{Y}^m_1}{\mu}\|_F^2 + \frac{\mu}{2}\|\mathbf{P}^m - \mathbf{Z}^m + \frac{\mathbf{Y}^m_2}{\mu}\|_F^2 \\ &+ \frac{\mu}{2}\|\mathbf{Q}^m - \mathbf{Z}^m + \frac{\mathbf{Y}^m_3}{\mu}\|_F^2 - \frac{1}{2\mu}(\|\mathbf{Y}^m_1\|_F^2 + \|\mathbf{Y}^m_2\|_F^2 + \|\mathbf{Y}^m_3\|_F^2)). \end{aligned} \tag{11}$$

where $\mathbf{P} = [\mathbf{P}^1; \mathbf{P}^2; \ldots; \mathbf{P}^M]$ and $\mathbf{Q} = [\mathbf{Q}^1; \mathbf{Q}^2; \ldots; \mathbf{Q}^M]$. μ is a penalty parameter; $\mathbf{Y}^m_1$, $\mathbf{Y}^m_2$, and $\mathbf{Y}^m_3$ are Lagrange multipliers.

There are five variables, $\mathbf{Z}$, $\mathbf{E}^m$, $\mathbf{W}^a$, $\mathbf{P}$, and $\mathbf{Q}$, needed to solve in Equation (11), The solver iteratively updates one variable at a time by fixing the others.

$\mathbf{Z}$-subproblem: In order to calculate $\mathbf{Z}$, we fix other variables in Equation (11); the $\mathbf{Z}$-subproblem can be written as Equation (12). Then, we divide $\mathbf{Z}$ and set it to 0 to obtain Equation (13),

$$\begin{aligned} &\min_{\mathbf{Z}} \sum_{m=1}^{M}(\frac{\mu}{2}\|\mathbf{X}^m - \mathbf{X}^m\mathbf{Z}^m - \mathbf{E}^m + \frac{\mathbf{Y}^m_1}{\mu}\|_F^2 \\ &+ \frac{\mu}{2}\|\mathbf{P}^m - \mathbf{Z}^m + \frac{\mathbf{Y}^m_2}{\mu}\|_F^2 + \frac{\mu}{2}\|\mathbf{Q}^m - \mathbf{Z}^m + \frac{\mathbf{Y}^m_3}{\mu}\|_F^2), \end{aligned} \tag{12}$$

$$\begin{aligned}\mathbf{Z}^{m,k+1} &= (\mu(\mathbf{X}^m)^T\mathbf{X}^m + 2\mu_k\mathbf{I})^{-1}(\mu_k(\mathbf{X}^m)^T\mathbf{X}^m \\ &- \mu_k(\mathbf{X}^m)^T\mathbf{E}^{m,k} + (\mathbf{X}^m)^T\mathbf{Y}_1^{m,k} + \mu_k\mathbf{P}^{m,k} + \mu_k\mathbf{Q}^{m,k} - \mathbf{Y}_2^{m,k} - \mathbf{Y}_3^{m,k}).\end{aligned} \tag{13}$$

P-subproblem: In order to calculate **P**, we fix other variables in Equation (11); the **P**-subproblem can be written as Equations (14) and (15), then dividing **P** and setting it to 0 to obtain Equation (16),

$$\min_{\mathbf{P}} \gamma\, tr(\mathbf{P}\mathbf{L}^a\mathbf{P}^T) + \delta\, tr(\mathbf{P}\mathbf{L}^c\mathbf{P}^T) + \sum_{m=1}^{M} \frac{\mu}{2}\|\mathbf{P}^m - \mathbf{Z}^m + \frac{\mathbf{Y}_2^m}{\mu}\|_F^2, \tag{14}$$

$$\min_{\mathbf{P}} \gamma\, tr(\mathbf{P}\mathbf{L}^a\mathbf{P}^T) + \delta\, tr(\mathbf{P}\mathbf{L}^c\mathbf{P}^T) + \frac{\mu}{2}\|\mathbf{P} - \mathbf{Z} + \frac{\mathbf{Y}_2}{\mu}\|_F^2, \tag{15}$$

$$\mathbf{P}^{k+1} = (\mu\mathbf{Z}^{k+1} - \mathbf{Y}_2^k)(\gamma(\mathbf{L}^a)^k + \gamma((\mathbf{L}^a)^k)^T + \delta(\mathbf{L}^a)^k + \delta((\mathbf{L}^c)^k)^T + \mu\mathbf{I})^{-1}. \tag{16}$$

Q-subproblem: In order to calculate **Q**, we fix other variables in Equation (11), then the **Q**-subproblem can be written as Equations (17) and (18). Then, divide **Q** and set it to 0, which is computed by the soft-thresholding (or shrinkage) method [39] as Equation (19),

$$\min_{\mathbf{Q}} \alpha\|\mathbf{Q}\|_1 + \sum_{m=1}^{M} \frac{\mu}{2}\|\mathbf{Q}^m - \mathbf{Z}^m + \frac{\mathbf{Y}_3^m}{\mu}\|_F^2, \tag{17}$$

$$\min_{\mathbf{Q}} \alpha\|\mathbf{Q}\|_1 + \frac{\mu}{2}\|\mathbf{Q} - \mathbf{Z} + \frac{\mathbf{Y}_3}{\mu}\|_F^2, \tag{18}$$

$$\mathbf{Q}^{k+1} = soft\_thr(\mathbf{Z}^{k+1} - \frac{\mathbf{Y}_3^k}{\mu}, \frac{\alpha}{\mu_k}). \tag{19}$$

$\mathbf{E}^m$-subproblem: In order to calculate $\mathbf{E}^m$, we fix other variables in Equation (11); then the **Q**-subproblem can be written as Equation (20). Then, by dividing **E** and setting it to 0, which is computed by the soft-thresholding (or shrinkage) method [39], we obtain Equation (21),

$$\min_{\mathbf{E}^m} \sum_{m=1}^{M} \beta\|\mathbf{E}^m\|_{2,1} + \frac{\mu}{2}\|\mathbf{X}^m - \mathbf{X}^m\mathbf{Z}^m - \mathbf{E}^m + \frac{\mathbf{Y}_1^m}{\mu}\|_F^2 \tag{20}$$

$$\mathbf{E}^{m,k+1} = S_{\frac{\beta}{\mu}}(\mathbf{X}^m\mathbf{Z}^{m,k+1} - \mathbf{X}^m - \frac{\mathbf{Y}_1^{m,k}}{\mu_k}) \tag{21}$$

$\mathbf{W}^a$-subproblem: In order to calculate $\mathbf{W}^a$, we fix other variables in Equation (11), then the $\mathbf{W}^a$-subproblem can be written as Equation (22). Then dividing $\mathbf{W}^a$ and set it to 0 obtains Equation (23),

$$\min_{\mathbf{W}^a} \gamma tr(\mathbf{P}\mathbf{L}^a\mathbf{P}^T) + \lambda_1\|\mathbf{W}^a\|_F^2 + \gamma\|\mathbf{P}_i - \mathbf{P}_j\|_F^2\mathbf{W}_{ij}^a \tag{22}$$

$$(\mathbf{W}^a)_i^{k+1} = (\frac{1 + \sum_{j=1}^{s}\hat{\mathbf{U}}_j}{s}\mathbf{1} - \mathbf{U}_{ij})_+ \tag{23}$$

where $\mathbf{U}_j \in \mathbb{R}^{N\times 1}$ is a vector whose i-th element is $\mathbf{U}_{ij} = \frac{\gamma\|\mathbf{P}_i - \mathbf{P}_j\|_F^2}{\lambda_1}$.

The Lagrange multiplier can be updated by Equation (24),

$$\begin{aligned}\mathbf{Y}_1^{m,k+1} &= \mathbf{Y}_1^{m,k} + \mu^k(\mathbf{X}^m - \mathbf{X}^m\mathbf{Z}^{m,k+1} - \mathbf{E}^{m,k+1})\\ \mathbf{Y}_2^{m,k+1} &= \mathbf{Y}_2^{m,k} + \mu^k(\mathbf{Z}^{m,k+1} - \mathbf{P}^{m,k+1})\\ \mathbf{Y}_3^{m,k+1} &= \mathbf{Y}_3^{m,k} + \mu^k(\mathbf{Z}^{m,k+1} - \mathbf{Q}^{m,k+1})\end{aligned} \tag{24}$$

5. RGB-T Salient Detection

Given a pair of RGB-T images, considering that the thermal image has stronger anti-interference ability in complex scenes, we first fuse the RGB and the thermal images at a ratio of 1:4. To generate N non-overlapping superpixels, we use a simple linear iterative clustering (SLIC) algorithm in the fused image. A two-stage ranking model is adapted to calculate the final saliency map. In the first stage, we take the boundary as a prior and select the nodes around the image as background seed queries. We use the top, bottom, left, and right sides of the image as four kinds of queries, $\mathbf{q}^t, \mathbf{q}^b, \mathbf{q}^l, \mathbf{q}^r$, which are selected separately to obtain four different detection results, $\mathbf{f}^t, \mathbf{f}^b, \mathbf{f}^l, \mathbf{f}^r$, by Equation (2). Considering that the symmetry of the image and saliency objects are often cross-left boundary and cross-bottom boundary, we select the large class nodes as queries by using the k-means method to obtain two clusters on the left and bottom boundaries separately. Then, we normalize $\mathbf{f}^k$ $(k = t, p, l, r)$ to the range between 0 and 1. The saliency value vector of N nodes $\mathbf{s}^k$ can be obtained by $\mathbf{s}^k = \mathbf{1} - \hat{\mathbf{f}}^k$ $(k = t, p, l, r)$. The saliency ranking value vector of all nodes $\mathbf{s}^1$ in the first stage can be calculated by Equation (25).

$$\mathbf{s}^1 = \mathbf{s}^t \times \mathbf{s}^b \times \mathbf{s}^l \times \mathbf{s}^r \tag{25}$$

By using the object characteristics, secondary ranking is performed to improve the first-stage saliency value. Given $\mathbf{s}^1$, we set an adaptive threshold to generate foreground as queries $\mathbf{q}_2$. Then, the Equation (2) is used to obtain the second ranking results $\mathbf{s}_2$, which are normalized to the range of 0 and 1 as $\hat{\mathbf{s}}_2$. In order to further reduce the background noise, we let $\mathbf{s} = \mathbf{s}^1 \times \mathbf{s}^2$ be the final saliency value and obtain the final salient map $\mathbf{S}$. The main steps of the two-stage RGB-T salient object detection algorithm are summarized in Algorithm 1.

Algorithm 1 The Static-Adaptive Graph based RGB-T Salient Detection Produce.

Require: The static-adaptive graph weight matrix $\mathbf{W}$, the indicator vectors of the four boundaries queries $\mathbf{q}^t$, $\mathbf{q}^b$, $\mathbf{q}^l$, $\mathbf{q}^r$.

1: Use Equation (2) to obtain $\mathbf{f}^t, \mathbf{f}^b, \mathbf{f}^l, \mathbf{f}^r$ separately;
2: $\mathbf{f}^t$, $\mathbf{f}^b$, $\mathbf{f}^l$ and $\mathbf{f}^r$ are normalized to 0 and 1;
3: Set $\mathbf{s}^t$=$\mathbf{1} - \hat{\mathbf{f}}^t$, $\mathbf{s}^b$=$\mathbf{1} - \hat{\mathbf{f}}^b$, $\mathbf{s}^l$=$\mathbf{1} - \hat{\mathbf{f}}^l$, $\mathbf{s}^r$=$\mathbf{1} - \hat{\mathbf{f}}^r$;
4: Obtain the first saliency value vector $\mathbf{s}^1 = \mathbf{s}^t \times \mathbf{s}^b \times \mathbf{s}^l \times \mathbf{s}^r$;
5: $\mathbf{s}^1$ is normalized to 0 and 1, and obtain $\hat{\mathbf{s}}^1$;
6: Use an adaptive threshold to binary $\hat{\mathbf{s}}^1$ and obtain foreground query $\mathbf{q}^2$;
7: Use Equation (2) to obtain the second saliency value vector $\mathbf{s}^2$;
8: $\mathbf{s}^2$ is normalized to 0 and 1 $\hat{\mathbf{s}}^2$;
9: Set $\mathbf{s} = \hat{\mathbf{s}}^1 \times \hat{\mathbf{s}}^2$ to suppress the background noise of image;
10: Set all superpixels value $\mathbf{s}_i$ to each pixel and obtain the final saliency map $\mathbf{S}$.

Ensure: $\mathbf{S}$ is the saliency map of the static-adaptive graph model for RGB-T saliency detection.

6. Experiment

6.1. Datasets and Experimental Settings

The RGBT-Saliency dataset [26] includes 821 pairs images with ground truth, in which the images with high diversity are recorded under different scenes and environmental conditions.

The datasets can be download from the address http://chenglongli.cn/people/lcl/journals.html (accessed on 20 December 2021).

The initial segmentation number of the superpixel N is set to 250. The edge weight coefficient θ is set to 29. Other parameters in this paper are set to $\alpha = 0.11$, $\beta = 0.15$, $\gamma = 0.04$, $\delta = 0.3$, and $\lambda_1 = 0.6$.

6.2. Measuring Standard

To verify the effectiveness of our algorithm, we compared with other methods with precision, recall, and F-measure (PRF) values, mean absolute error (MAE) values, and PR curve.

PR (*Precision*, *Recall*) curve. The PR curve is a curve with the "precision rate" as the ordinate and the "recall rate" as the abscissa. We binarize the original image S to obtain M, and then calculate the precision value and recall value by comparing M and G (ground truth) pixel by pixel in the following formula,

$$Precision = \frac{|M \cap G|}{|M|} \tag{26}$$

$$Recall = \frac{|M \cap G|}{|G|} \tag{27}$$

PRF (precision, recall, F-measure). Sometimes, the P and R indicators are contradictory, so they need to be considered comprehensively. The most common method is F-measure (also known as f-score). F-measure is the weighted average of precision and recall:

$$F_{\beta^2} = \frac{(1+\beta^2) \times P \times R}{\beta^2 \times P + R}, \tag{28}$$

where $\beta^2 = 0.3$.

MAE (mean absolute error). *MAE* is the direct calculation of the average absolute error between the salience map and the ground truth of the model output. It first binarizes them and then calculates them with the following formula:

$$MAE = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} |\overline{S}(x,y) - \overline{G}(x,y)| \tag{29}$$

where W is the width of the salient map S and the ground truth map G; H is the height of the salient map S and the ground truth map G.

6.3. Comparison Results

We compared our model with eight methods including BR [40], CA [41], MCI [42], NFI [43], SS-KDE [44], GMR [20], GR [45], and MTMR [26] on the RGBT-Saliency dataset.

We generated PR curves for 11 challenging subsets and the entire dataset, and listed their F values. The eleven subsets are eleven different challenges, which are: big salient object (BSO), bad weather (BW), center bias (CB), cross image boundary (CIB), image clutter (IC), low illumination (LI), multiple salient objects (MSO), out of focus (OF), similar appearance (SA), small salient object (SSO), and thermal crossover (TC). In Table 1, we describe in detail the division method of the eleven subsets [26].

Table 1. List of the 11 challenging subsets of RGBT-Saliency-Dataset.

Challenge	Description
BSO	The radio of ground truth salient objects over the image is more than 0.26.
BW	The image pairs are recorded in bad weather, such as snowy, rainy, hazy, or cloudy weather.
CB	The centers of salient objects are far away from the image center.
CIB	The salient objects cross the image boundaries.
IC	The image is cluttered.
LI	The environmental illumination is low.
MSO	The number of the salient objects in the image is more than one.
OF	The image is out of focus.
SA	The salient objects have similar color or shape to the background.
SSO	The radio of ground truth salient objects over the image is less the 0.05.
TC	The salient objects have similar temperature to the background.

As can be seen from Figure 3, only in the "BSO" and "CIB" subsets was our F-Measures slightly lower than the best detection result, and they were the best in the other nine subsets. Especially in the CB subset, the detection result has obvious advantages. Our detection curve has no crossover with other curves.

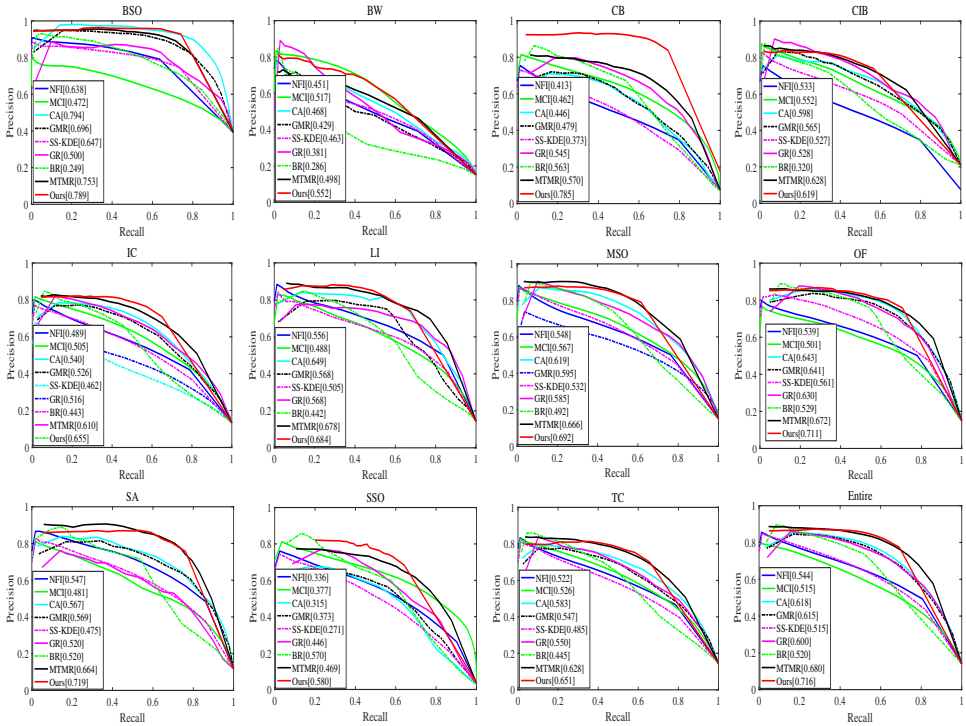

Figure 3. PR curves of the proposed approach with other baseline methods with RGB-T input on eleven subsets and the entire dataset. The $F_{0.3}$ values are shown in the legend.

The comparison results of the precision, recall, and F-measure values with other methods in different modalities as shown in Table 2. We only provide the detection results of MTMR [26] after multi-modality fusion because this model proposes to integrate multi-modal information and use multi-modal adaptive weights to detect image saliency objects. From the Table 2, we can see that the proposed algorithm is better than other methods in terms of P value and the comprehensive measure F-measure.

Table 2. Average precision (P), recall (R), F-measure (F) and mean absolute error (MAE) of our method against different kinds of methods on the RGBT-Saliency dataset. In the evaluation parameters, the larger the value of P, R, and F, the better the detection effect, while the smaller the value of MAE, the better the effect. The red font indicates the best performance. The green is second best.

Algorithm	RGB (P↑, R↑, F↑, MAE↓)	Thermal (P↑, R↑, F↑, MAE↓)	RGB-T (P↑, R↑, F↑, MAE↓)
BR [40]	0.724, 0.260, 0.411, 0.269	0.648, 0.413, 0.488, 0.323	0.804, 0.366, 0.520, 0.297
CA [41]	0.592, 0.667, 0.568, 0.163	0.623, 0.607, 0.573, 0.225	0.648, 0.697, 0.618, 0.195
MCI [42]	0.526, 0.604, 0.485, 0.211	0.445, 0.585, 0.435, 0.176	0.547, 0.652, 0.515, 0.195
NFI [43]	0.557, 0.639, 0.532, 0.126	0.581, 0.599, 0.541, 0.124	0.564, 0.665, 0.544, 0.125
SS-KDE [44]	0.581, 0.554, 0.532, 0.122	0.510, 0.635, 0.497, 0.132	0.528, 0.656, 0.515, 0.127
GMR [20]	0.644, 0.603, 0.587, 0.172	0.700, 0.574, 0.603, 0.232	0.694, 0.624, 0.615, 0.202
GR [45]	0.621, 0.582, 0.534, 0.197	0.639, 0.544, 0.545, 0.199	0.705, 0.593, 0.600, 0.199
MTMR [26]	-, -, -, -	-, -, -, -	0.716, 0.713, 0.680, 0.107
ours	0.697, 0.536, 0.603, 0.107	0.715, 0.569, 0.629, 0.112	0.804, 0.627, 0.716, 0.095

Sample Results. From the dataset, we extracted four photos with various challenges as the data source and compared the detection results of our algorithm with other algorithms for salient detection. It can be seen from the Figure 4 that our algorithm has a very robust detection effectiveness in challenging scenes such as fuzzy images, large targets, small targets, complex background, and center bias.

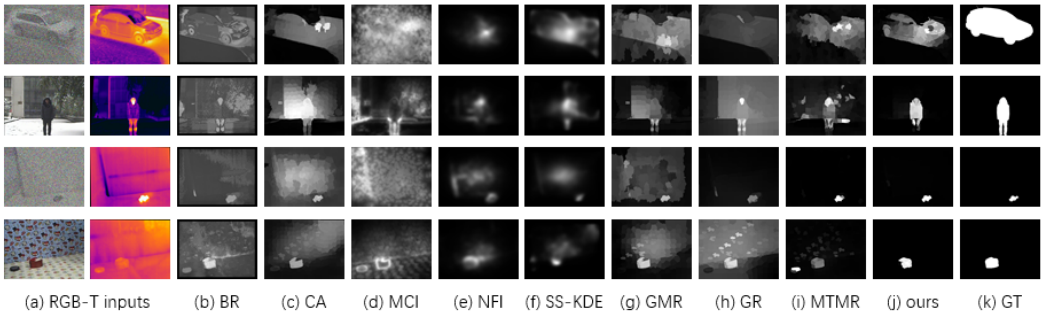

Figure 4. Sample results of the proposed approach and other baseline methods with the fusion of RGB and thermal inputs. (**a**) The first two columns are the origin RGB images and thermal images. (**b**–**i**) The results of the baseline methods with RGB and thermal inputs; (**j**) the result of our approach. (**k**) ground truth.

Runtime Results. All results were obtained on a Windows 10 64-bit operating system running Matlab 2014b with an i3 3.3G CPU and 4GB RAM. We compared the average running time with other algorithms in Table 3. Compared with the algorithm in [20], we spent more time mainly on the learning of the adaptive graph.

Table 3. Average runtime comparison on the RGBT-Saliency dataset.

Method	BR [40]	CA [41]	MCI [42]	NFI [43]	SS-KDE [44]	GMR [20]	GR [45]	MTMR [26]	Ours
Runtime(s)	21.95	3.13	58.37	33.16	2.51	2.96	6.48	3.71	5.18

6.4. Analysis of Our Approach

In our method, we compared the following four combinations of image salient detection results: (1) learning static-adaptive graphs for RGB image salient detection, called our1; (2) learning static-adaptive graph for thermal image salient detection, called our2; (3) not learning static-adaptive graphs and only fusing RGB and thermal image to detect the salient, called our3; (4) learning static-adaptive graphs for RGB-T image salient detection, called our4. It can be seen from Figure 5 that the fusion of multi-modality and the use of learning static-adaptive graphs are both effective methods to improve the salient detection.

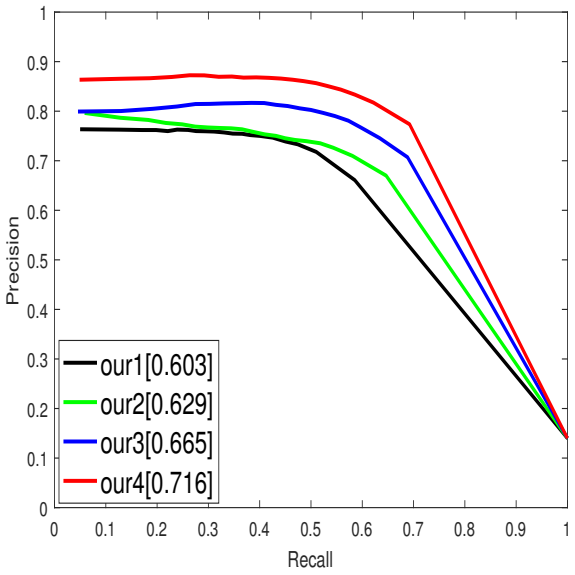

Figure 5. PR curves of our approach with its variants on the entire dataset.

Advantages. We fused thermal and RGB images for image salient detection, which can overcome the limitations of light, ambient temperature, background clutter, and color similarity in single mode. By learning the static-adaptive method, we not only retained the local features of superpixels, but also learned to mine their internal relations to obtain a better affinity matrix of superpixels and greatly improve the detection accuracy of image saliency.

Limitations. Through the experiment, we found that under complex scenes, multi-modality fusion can effectively improve the image in general. However, in some cases, the single-modality has better detection accuracy. Our future work will set the modality weight according to the image characteristics and further improve the detection effect of image saliency in complex scenes.

7. Conclusions

In this paper, we combine RGB-thermal modality information for image salient detection, which effectively improves the detection performance of single-modality RGB images under poor illumination and when the background and foreground colors are similar. At the same time, our method improves the detection accuracy of thermal images under normal lighting conditions, especially in the case of small temperature differences between the environment and the target. The image is dynamically learned, taking both global and local cues into account, and thus our method is capable of capturing the intrinsic relationship of superpixels. In the future, we will assign different weights to different modality images according to the characteristics of different modality images.

Author Contributions: Z.X. and J.T. proposed the idea, designed and performed the simulations, and wrote the paper; A.Z. and H.L. analyzed the data. All authors have read and agreed to the published version of the manuscript.

Funding: This paper is funded by the following foundations: National Natural Science Foundation of China (61906044), Natural Science Foundation of Anhui Higher Education Institution of China (KJ2019A0536, KJ2019A0529, KJ2020ZD46), Natural Science Foundation of Fuyang Normal University (2019FSKJ02ZD), Fuyang Normal University Scientific Research Project(2020KYQD0032), the Young Talents Projects of Fuyang Normal University(rcxm202001, 2021FSKJ01ZD), Fuyang City School Cooperation Project (SXHZ202103).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets can be download from the address http://chenglongli.cn/people/lcl/journals.html (accessed on 20 December 2021).

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

References

1. Wang, Q.; Lin, J.; Yuan, Y. Salient band selection for hyperspectral image classification via manifold ranking. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *27*, 1279–1289. [CrossRef]
2. Yang, X.; Qian, X.; Xue, Y. Scalable mobile image retrieval by exploring contextual saliency. *IEEE Trans. Image Process.* **2015**, *24*, 1709–1721. [CrossRef]
3. Wen, W.; Zhang, Y.; Fang, Y.; Fang, Z. A novel selective image encryption method based on saliency detection. In Proceedings of the Visual Communications and Image Processing (VCIP), Chengdu, China, 27–30 November 2016; pp. 1–4.
4. Wen, W.; Zhang, Y.; Fang, Y.; Fang, Z. Image salient regions encryption for generating visually meaningful ciphertext image. *Neural Comput. Appl.* **2018**, *29*, 653–663. [CrossRef]
5. Jacob, H.; Padua, F.L.C.; Lacerda, A.; Pereira, A.C.M. A video summarization approach based on the emulation of bottom-up mechanisms of visual attention. *J. Intell. Inf. Syst.* **2017**, *49*, 193–211. [CrossRef]
6. Zhang, L.; Ai, J.; Jiang, B.; Lu, H.; Li, X. Saliency Detection via Absorbing Markov Chain With Learnt Transition Probability. *IEEE Trans. Image Process.* **2018**, *27*, 987–998. [CrossRef] [PubMed]
7. Borji, A.; Cheng, M.M.; Jiang, H.; Li, J. Salient object detection: A benchmark. *IEEE Trans. Image Process.* **2015**, *24*, 5706–5722. [CrossRef] [PubMed]
8. Tong, N.; Lu, H.; Ruan, X.; Yang, M.H. Salient object detection via bootstrap learning. In Proceedings of the Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1884–1892.
9. Zhou, X.; Liu, Z.; Sun, G.; Wang, X. Adaptive saliency fusion based on quality assessment. *Multimed. Tools Appl.* **2017**, *76*, 23187–23211. [CrossRef]
10. Itti, L.; Koch, C.; Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1254–1259. [CrossRef]
11. Cheng, M.M.; Mitra, N.J.; Huang, X.; Torr, P.H.; Hu, S.M. Global contrast based salient region detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 569–582. [CrossRef]
12. Wang, H.; Xu, L.; Wang, X.; Luo, B. Learning Optimal Seeds for Ranking Saliency. *Cogn. Comput.* **2018**, *10*, 347–358. [CrossRef]
13. Hou, Q.; Cheng, M.M.; Hu, X.; Borji, A.; Tu, Z.; Torr, P.H. Deeply supervised salient object detection with short connections. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3203–3212.
14. Han, J.; Zhang, D.; Cheng, G.; Liu, N.; Xu, D. Advanced deep-learning techniques for salient and category-specific object detection: A survey. *IEEE Signal Process. Mag.* **2018**, *35*, 84–100. [CrossRef]
15. Li, C.; Zhao, N.; Lu, Y.; Zhu, C.; Tang, J. Weighted Sparse Representation Regularized Graph Learning for RGB-T Object Tracking. In Proceedings of the 25th ACM International Conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 1856–1864.
16. Li, C.; Zhu, C.; Huang, Y.; Tang, J.; Wang, L. Cross-modal ranking with soft consistency and noisy labels for robust rgb-t tracking. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 808–823.
17. Li, C.; Liang, X.; Lu, Y.; Zhao, N.; Tang, J. RGB-T object tracking: Benchmark and baseline. *Pattern Recognit.* **2019**, *96*, 106977. [CrossRef]
18. Zhang, Q.; Huang, N.; Yao, L.; Zhang, D.; Shan, C.; Han, J. RGB-T salient object detection via fusing multi-level CNN features. *IEEE Trans. Image Process.* **2019**, *29*, 3321–3335. [CrossRef] [PubMed]
19. Harel, J.; Koch, C.; Perona, P. Graph-Based Visual Saliency. *Adv. Neural Inf. Process. Syst.* **2006**, *19*, *545-552*.
20. 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 (CVPR), Washington, DC, USA, 23–28 June 2013; pp. 3166–3173.
21. Sun, J.; Lu, H.; Liu, X. Saliency region detection based on Markov absorption probabilities. *IEEE Trans. Image Process.* **2015**, *24*, 1639–1649. [CrossRef]
22. Zhang, L.; Yang, C.; Lu, H.; Ruan, X.; Yang, M.H. Ranking saliency. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1892–1904. [CrossRef]
23. Xiao, Y.; Wang, L.; Jiang, B.; Tu, Z.; Tang, J. A global and local consistent ranking model for image saliency computation. *J. Vis. Commun. Image Represent.* **2017**, *46*, 199–207. [CrossRef]
24. Aytekin, Ç; Iosifidis, A.; Kiranyaz, S.; Gabbouj, M. Learning graph affinities for spectral graph-based salient object detection. *Pattern Recognit. J. Pattern Recognit. Soc.* **2017**, *64*, 159–167. [CrossRef]
25. Li, C.; Cheng, H.; Hu, S.; Liu, X.; Tang, J.; Lin, L. Learning collaborative sparse representation for grayscale-thermal tracking. *IEEE Trans. Image Process.* **2016**, *25*, 5743–5756. [CrossRef]
26. Li, C.; Wang, G.; Ma, Y.; Zheng, A.; Luo, B.; Tang, J. A Unified RGB-T Saliency Detection Benchmark: Dataset, Baselines, Analysis and A Novel Approach. *arXiv* **2017**, arXiv:1701.02829.

27. Giacomo, C.; Grazia, L.S.; Christian, N.; Rafi, S.; Marcin, W. Optimizing the Organic Solar Cell Manufacturing Process by Means of AFM Measurements and Neural Networks. *Energies* **2018**, *11*, 1221.
28. Huang, Z.; Wang, X.; Huang, L.; Huang, C.; Wei, Y.; Liu, W. CCNet: Criss-Cross Attention for Semantic Segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019.
29. Hu, X.; Yang, K.; Fei, L.; Wang, K. ACNet: Attention Based Network to Exploit Complementary Features for RGBD Semantic Segmentation. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, China, 22–25 September 2019; pp. 1440–1444.
30. Zhang, J.; Yang, K.; Constantinescu, A.; Peng, K.; Müller, K.; Stiefelhagen, R. Trans4Trans: Efficient Transformer for Transparent Object Segmentation to Help Visually Impaired People Navigate in the Real World. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Nashville, TN, USA, 19–25 June 2021; pp.1760-1770.
31. Liu, N.; Han, J.; Yang, M.H. PiCANet: Learning Pixel-wise Contextual Attention for Saliency Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3089–3098.
32. Liu, Z.; Tan, Y.; He, Q.; Xiao, Y. SwinNet: Swin Transformer drives edge-aware RGB-D and RGB-T salient object detection. *IEEE Trans. Circuits Syst. Video Technol.* **2021**. [CrossRef]
33. Liu, Z.; Wang, Y.; Tu, Z.; Xiao, Y.; Tang, B. TriTransNet: RGB-D Salient Object Detection with a Triplet Transformer Embedding Network. In Proceedings of the 29th ACM International Conference on Multimedia, New York, NY, USA, 20–24 October 2021; pp. 4481-4490.
34. Guo, X. Robust Subspace Segmentation by Simultaneously Learning Data Representations and Their Affinity Matrix. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015), Buenos Aires, Argentina, 25–31 July 2015; AAAI Press: Palo Alto, CA, USA, 2015; pp. 3547–3553.
35. Li, C.; Wu, X.; Bao, Z.; Tang, J. ReGLe: Spatially Regularized Graph Learning for Visual Tracking. In Proceedings of the 25th ACM International Conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 252–260.
36. 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] [PubMed]
37. Stephen, B.; Neal, P.; Chu, E.; Borja, P.; EcKstein, J. *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*; Now Publishers Inc.: Hanover, MA, USA, 2010; Volume 3, pp. 1–122.
38. Lin, Z.; Chen, M.; Ma, Y. The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices. *arXiv* **2010**, arXiv:1009.5055.
39. Chen, M.; Ganesh, A.; Lin, Z.; Ma, Y.; Wright, J.; Wu, L. *Fast Convex Optimization Algorithms for Exact Recovery of a Corrupted Low-Rank Matrix*; Report No. UILU-ENG-09-2214; Coordinated Science Laboratory: Urbana, IL, USA, 2009.
40. Rahtu, E.; Kannala, J.; Salo, M.; Heikkilä, J. Segmenting salient objects from images and videos. In Proceedings of the European Conference on Computer Vision, Heraklion, Crete, Greece, 5–11 September 2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 366–379.
41. Qin, Y.; Lu, H.; Xu, Y.; Wang, H. Saliency detection via cellular automata. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 110–119.
42. Goferman, S.; Zelnik-Manor, L.; Tal, A. Context-aware saliency detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1915–1926. [CrossRef] [PubMed]
43. Erdem, E.; Erdem, A. Visual saliency estimation by nonlinearly integrating features using region covariances. *J. Vis.* **2013**, *13*, 11. [CrossRef]
44. Tavakoli, H.R.; Rahtu, E.; Heikkilä, J. Fast and efficient saliency detection using sparse sampling and kernel density estimation. In Proceedings of the Scandinavian Conference on Image Analysis, Ystad, Sweden, 23–25 May 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 666–675.
45. Yang, C.; Zhang, L.; Lu, H. Graph-regularized saliency detection with convex-hull-based center prior. *IEEE Signal Process. Lett.* **2013**, *20*, 637–640. [CrossRef]