



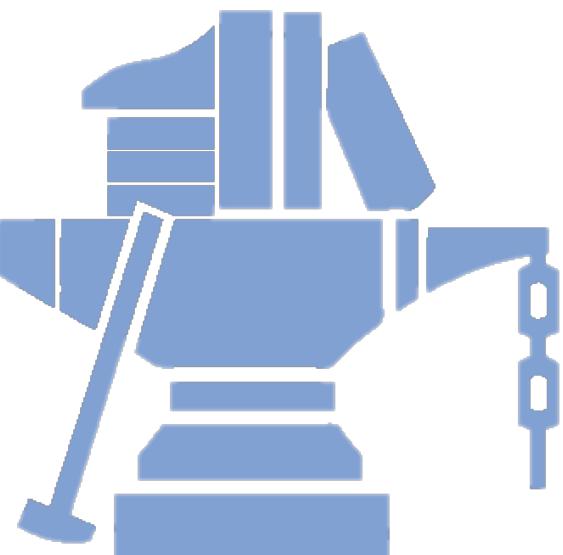
数字媒体信息处理研究中心
Center of Digital Media Information Processing



When Saliency Detection Meets Different Data Sources: Theory and Model

Runmin Cong
Beijing Jiaotong University



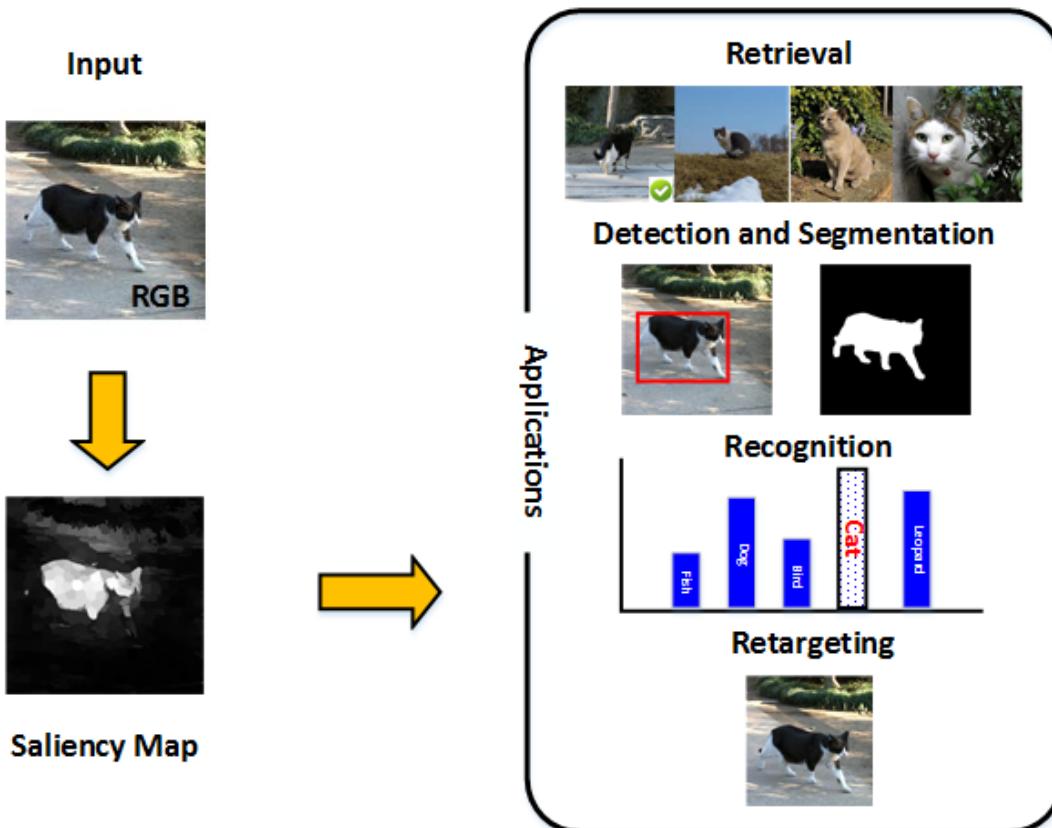


Outline

- Introduction
- RGBD Saliency Detection
 - Going from RGB to RGBD saliency: A depth-guided transformation model
- RGBD Co-saliency Detection
 - Co-saliency detection for RGBD images based on multi-constraint feature matching and cross label propagation
- Video Saliency Detection
 - Video saliency detection via sparsity-based reconstruction and propagation
- Conclusion

Introduction

- What is saliency detection?

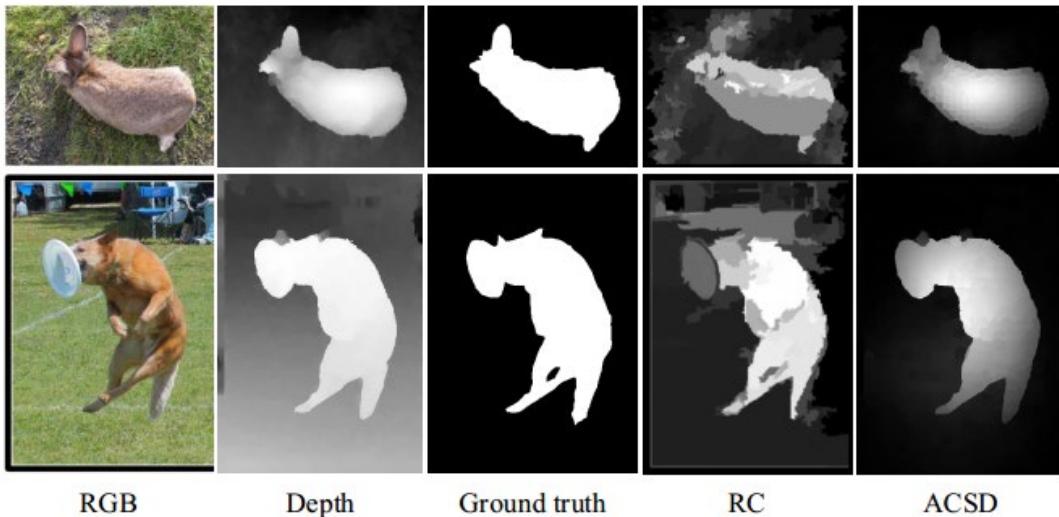


- ◆ Saliency detection aims to detecting the salient regions automatically, which has been applied in image/video segmentation, image/video retrieval, image retargeting, video coding, quality assessment, action recognition, and video summarization.
- ◆ The last decade has witnessed the remarkable progress of image saliency detection, and a plenty of methods have been proposed based on some priors or techniques, such as uniqueness prior, background prior, compactness prior, sparse coding, random walks, and deep learning.

Introduction

- What is saliency detection?

RGBD saliency detection

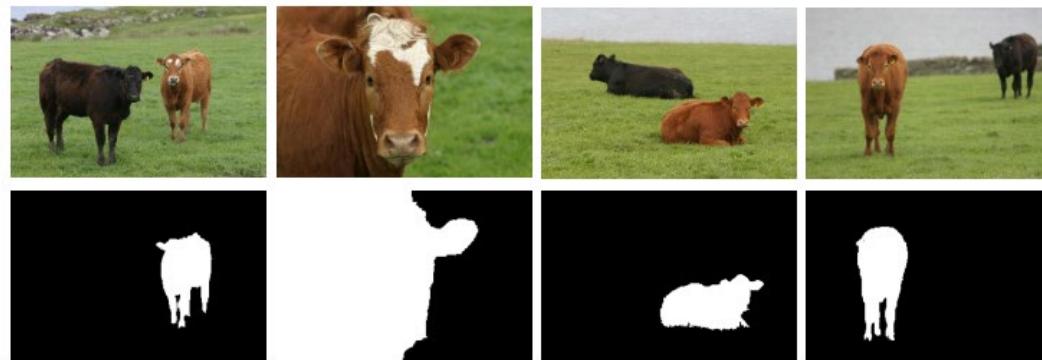


- ◆ In fact, the human visual system can not only perceive the appearance of the object, but also be affected by the depth information from the scene. Depth map provides better shape representation and other useful attributes for many vision tasks.
- ◆ Generally, depth information can be utilized in two manners: directly incorporating as an additional feature and designing as the depth measure.

Introduction

- What is saliency detection?

Co-saliency detection

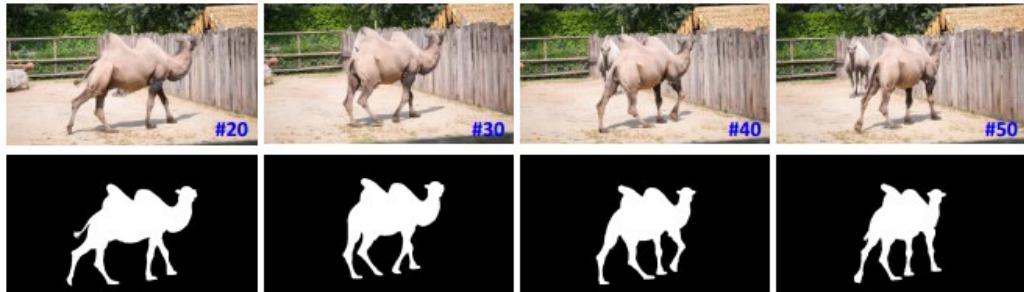


- ◆ co-saliency detection aims at detecting the common and salient regions from an image group containing multiple related images, while the categories, intrinsic attributes, and locations are entirely unknown.
- ◆ Therefore, the **inter-image correspondence** among multiple images plays a useful role in representing the common attribute.

Introduction

- What is saliency detection?

Video saliency detection



Motion cue

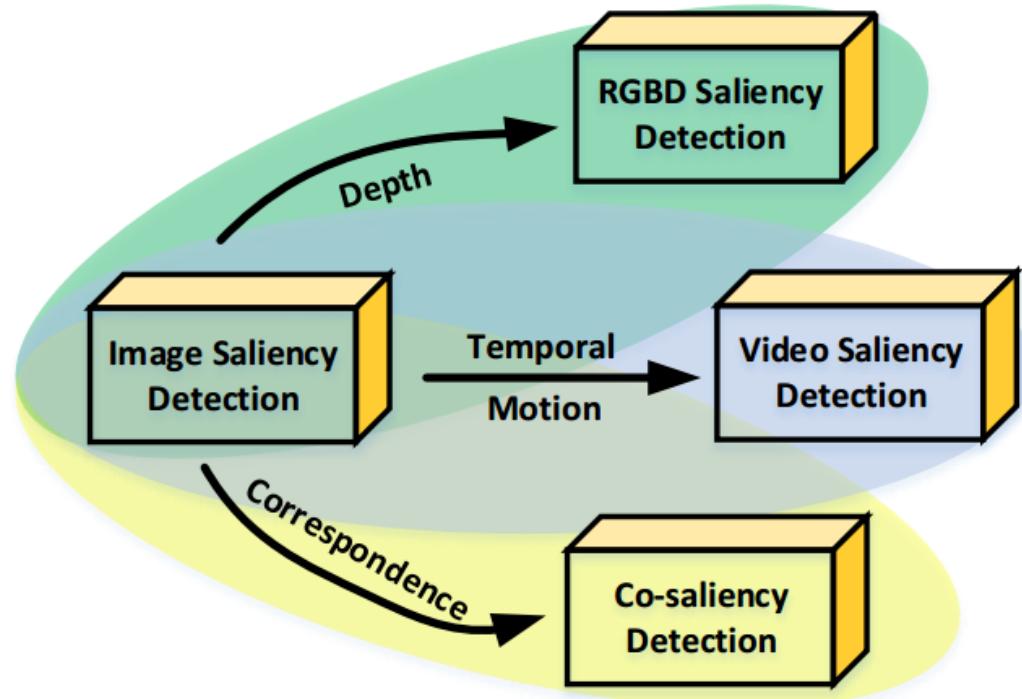
Inter-frame correspondence

Spatiotemporal constraint

- ◆ Video saliency detection aims at continuously locating the motion-related salient object from the given video sequences by considering the spatial and temporal information jointly.
- ◆ We divide the video saliency detection methods into two categories, i.e., low-level cues based method and learning based method.

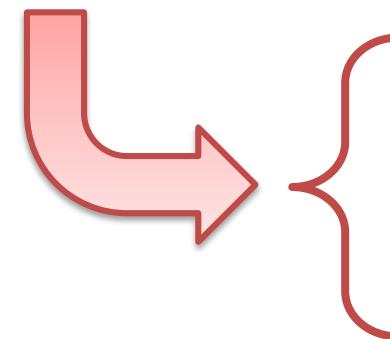
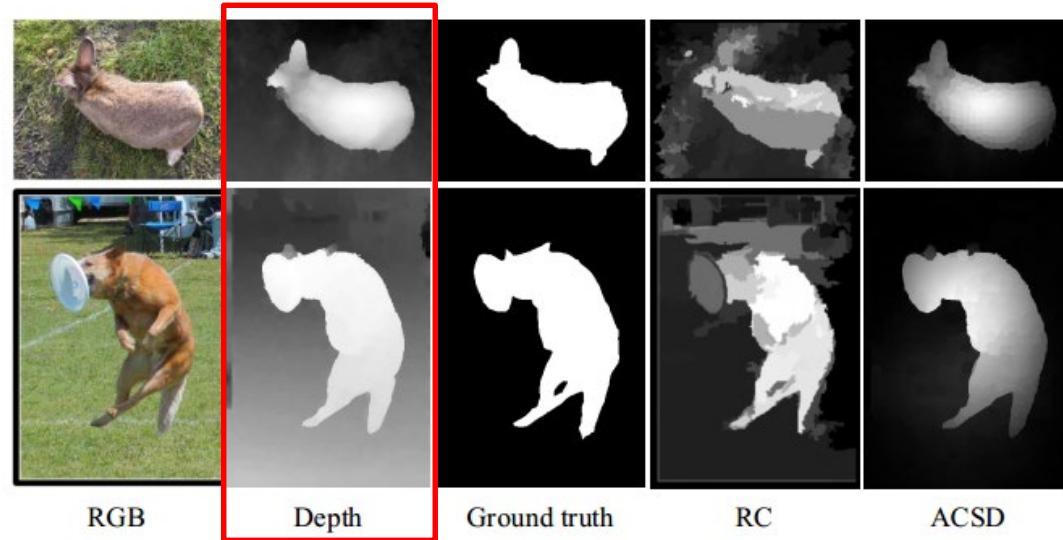
Introduction

- Relationship and Comprehensive Information



The image saliency detection model is the basis for other three models. With the acquisition technology development, more comprehensive information is available, such as the **depth cue for RGBD data**, the **inter-image constraint for image group**, and the **temporal relationship for video data**.

RGBD Saliency Detection



- shape
- contour
- internal consistency
- surface normal
-

Going from RGB to RGBD saliency: A depth-guided transformation model

**Runmin Cong, Jianjun Lei, Huazhu Fu, Junhui Hou,
Qingming Huang, Sam Kwong**

IEEE Transactions on Cybernetics, 2019

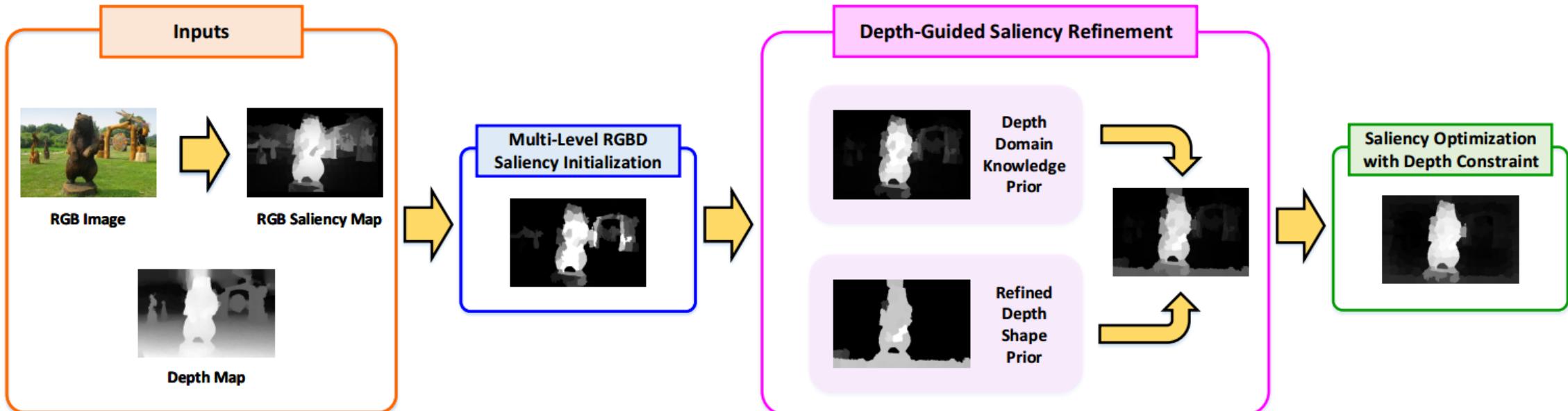
Going from RGB to RGBD saliency: A depth-guided transformation model

- **Motivation**

- Depth information has demonstrated to be useful for saliency detection. However, the existing methods for RGBD saliency detection mainly focus on designing straightforward and comprehensive models, while ignoring the transferable ability of the existing RGB saliency detection models.
 - We propose a novel depth-guided transformation model going from RGB saliency to RGBD saliency. The proposed model includes three components, i.e., **multi-level RGBD saliency initialization, depth-guided saliency refinement, and saliency optimization with depth constraints.**

Going from RGB to RGBD saliency: A depth-guided transformation model

- Framework



Going from RGB to RGBD saliency: A depth-guided transformation model

- **Contributions**

1. The most advantage of our method is to fully exploit the depth cue and provide **a general transformation model** going from RGB saliency to RGBD saliency.
2. A **multi-level RGBD saliency initialization** is proposed to integrate the global compactness and the local geodesic saliency cues.
3. To capture more accurate and complete shape information from the depth map, a **refined depth shape prior** is proposed.
4. To improve the accuracy and consistency, an **optimization strategy with depth constraints** is designed, which introduces the depth consistency relationship as an additional term in the energy optimization function.

Going from RGB to RGBD saliency: A depth-guided transformation model

- **Multi-level RGBD Saliency Initialization**
 - In order to exploit the depth feature and guarantee the basic performance of the proposed transformation model, a **multi-level RGBD saliency model** is proposed to generate the RGBD saliency initialization, which **provides an initialization of the transformation framework by using the explicit depth feature.**
 - The global compactness saliency cue is worked as a robust global representation combining the color compactness and depth compactness, and the local geodesic saliency cue is utilized to measure the saliency from the local perspective with the novel depth weight and graph relationship.

Going from RGB to RGBD saliency: A depth-guided transformation model

- **Multi-level RGBD Saliency Initialization**

- **Global compactness saliency cue.** we integrate the color compactness and depth compactness into a formulation to define the global saliency.

$$S_c(r_i) = 1 - \frac{\sum_{j=1}^N n_j \cdot (a_{ij}^c \cdot \|\mathbf{b}_j - \mathbf{u}_i\|_2 + a_{ij}^d \cdot \|\mathbf{b}_j - \mathbf{p}_0\|_2)}{\sum_{j=1}^N n_j \cdot (a_{ij}^c + a_{ij}^d)}$$

$$\begin{cases} a_{ij}^c = \exp(-\|\mathbf{c}_i - \mathbf{c}_j\|_2 / \sigma^2) \\ a_{ij}^d = \exp(-\lambda_d \cdot |d_i - d_j| / \sigma^2) \end{cases}$$

$$\mathbf{u}_i = [u_i^x, u_i^y] = [\frac{\sum_{j=1}^N a_{ij}^c \cdot n_j \cdot b_j^x}{\sum_{j=1}^N a_{ij}^c \cdot n_j}, \frac{\sum_{j=1}^N a_{ij}^c \cdot n_j \cdot b_j^y}{\sum_{j=1}^N a_{ij}^c \cdot n_j}]$$

- **Local geodesic saliency cue.** we calculate the geodesic saliency with the assistance of novel depth weight and optimized graph relationship from the local perspective.

$$S_G(r) = \min_{r_1=r, r_2, \dots, r_N=B} \sum_{i=1}^N w_{i,i+1}, \quad (r_i, r_{i+1}) \in \varepsilon.$$

$$\begin{aligned} \varepsilon = & \{(r_i, r_j) \mid r_i \text{ is adjacent to } r_j\} \\ & \cup \{(r_i, BV) \mid r_i \text{ is on image boundary}\} \\ & \cup \{(r_i, BG) \mid r_i \text{ is given background}\}. \end{aligned}$$

$$w_{ij} = \begin{cases} a_{ij}^c \cdot a_{ij}^d, & \text{if } (r_i, r_j) \in \varepsilon \\ 0, & \text{otherwise,} \end{cases}$$

$$S_{ML}(r_i) = \frac{1}{2}(S_{RGB}(r_i) + \mathbf{N}[S_C(r_i) + S_C(r_i) \cdot S_G(r_i)])$$

Going from RGB to RGBD saliency: A depth-guided transformation model

- **Depth-guided Saliency Refinement**

In this model, we exploit the **implicit depth information** to refine the saliency map.

(1) Generally, the salient object is placed near the camera by a photographer when taking a picture.

Thus, object with a large depth magnitude tend to be salient. Moreover, the depth distribution between the foreground and background regions is different. Therefore, **the depth domain knowledge prior, including the depth distance and depth contrast**, is proposed to refine the saliency map.

(2) Although the depth map does not provide rich texture information as color image, it provides effective shape attribute representation. Based on this, **the refined depth shape prior (RDSP) refinement is proposed to capture the shape constraint** from the depth map and refine the salient region.

Going from RGB to RGBD saliency: A depth-guided transformation model

- **Depth-guided Saliency Refinement: Depth domain knowledge prior**

- a) salient object tends to be close the camera with a large depth magnitude;
- b) salient object could be identified by the depth contrast compared with the background regions.

$$S_{DDK}(r_i)$$

$$= \begin{cases} \frac{1}{2}(S_{ML}(r_i) + S_{DC}(r_i)), & \lambda_d \geq \tau_1 \\ S_{ML}(r_i) \cdot d_i, & \tau_2 \leq \lambda_d < \tau_1 \\ S_{ML}(r_i), & \text{otherwise,} \end{cases}$$

When the quality of depth map is reliable, the depth contrast could better describe the depth saliency characteristic, which is directly used to refine the initial saliency map.

When the quality of depth map is tolerable, the depth distance relationship is used to weight the initial saliency map.

When the depth map quality is poor, which is unable to provide enough effective and accurate auxiliary information for saliency detection, we only retain the initial saliency result.

$$S_{DC}(r_i) = \sum_{j=1, j \neq i}^N |d_i - d_j| \cdot \exp(-\|\mathbf{b}_i - \mathbf{b}_j\|_2 / \sigma^2)$$

Going from RGB to RGBD saliency: A depth-guided transformation model

- **Depth-guided Saliency Refinement: Refined depth shape prior**
 - **Smoothness Decision:** The depth difference with color constraint between the neighbor superpixel and l-1-loop child seeds should be less than a given threshold;
 - **Consistency Decision:** The depth difference with color constraint between the neighbor superpixel and root seed should be smaller than a given threshold;

$$RDSP_k(r_{cp}) = 1 - \min(|d_{cp_l} - d_{c_{l-1}}|, |d_{cp_l} - d_{rk}|)$$

$$RDSP(r_i) = (\sum_{k=1}^K RDSP_k(r_i))/K$$

$$S_{DR}(r_i) = \mathbf{N}[S_{DDK}(r_i) + RDSP(r_i)]$$



Going from RGB to RGBD saliency: A depth-guided transformation model

- **Saliency Optimization with Depth Constraints**

- From the depth map, in addition to providing effective shape description, the whole object usually has high consistency in the depth map. Therefore, the depth information can be used to improve the consistency and smoothness of the acquired saliency map. In our work, a saliency optimization strategy with the depth constraint is formulated to attain more consistent and accurate saliency result, where the depth consistency relationship is introduced as an additional term in the energy function.

$$E = E_u + E_s + E_c = \sum_i (s_i^* - s_i)^2 + \sum_{(i,j) \in \Omega_s} \omega_{ij}^c \cdot (s_i^* - s_j^*)^2 + \sum_{(i,j) \in \Omega_s} \omega_{ij}^d \cdot (s_i^* - s_j^*)^2$$

data term controls the updating degree between the final saliency map and initial saliency map

color smooth term constrains the spatially adjacent regions, which guarantees the similar color appearance should be assigned to similar saliency scores.

depth consistency term imposes that the adjacent regions with similar depth distribution should be assigned to consistent saliency scores

Going from RGB to RGBD saliency: A depth-guided transformation model

- **Saliency Optimization with Depth Constraints**

$$\mathbf{E} = \mathbf{E}_u + \mathbf{E}_s + \mathbf{E}_c = (\mathbf{s}^* - \mathbf{s})^T \cdot (\mathbf{s}^* - \mathbf{s}) + \mathbf{s}^{*T} \cdot (\mathbf{D}_c - \mathbf{W}_c) \cdot \mathbf{s}^* + \mathbf{s}^{*T} \cdot (\mathbf{D}_d - \mathbf{W}_d) \cdot \mathbf{s}^*$$

$$\mathbf{s} = [s_1, s_2, \dots, s_N]^T \quad \mathbf{s}^* = [s_1^*, s_2^*, \dots, s_N^*]^T$$

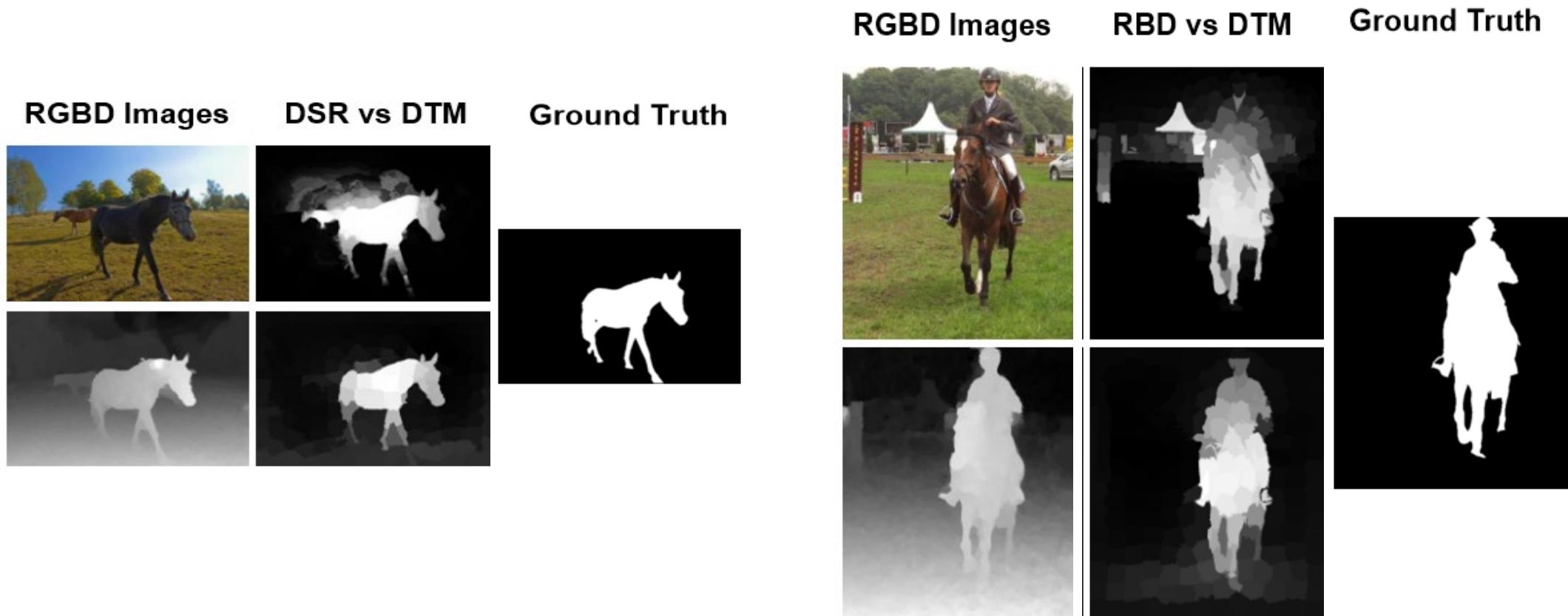
$$\mathbf{W}_c = [\omega_{ij}^c]_{N \times N} \quad \mathbf{W}_d = [\omega_{ij}^d]_{N \times N}$$

$$\mathbf{D}_c = diag(d_1^c, d_2^c, \dots, d_N^c) \quad \mathbf{D}_d = diag(d_1^d, d_2^d, \dots, d_N^d)$$

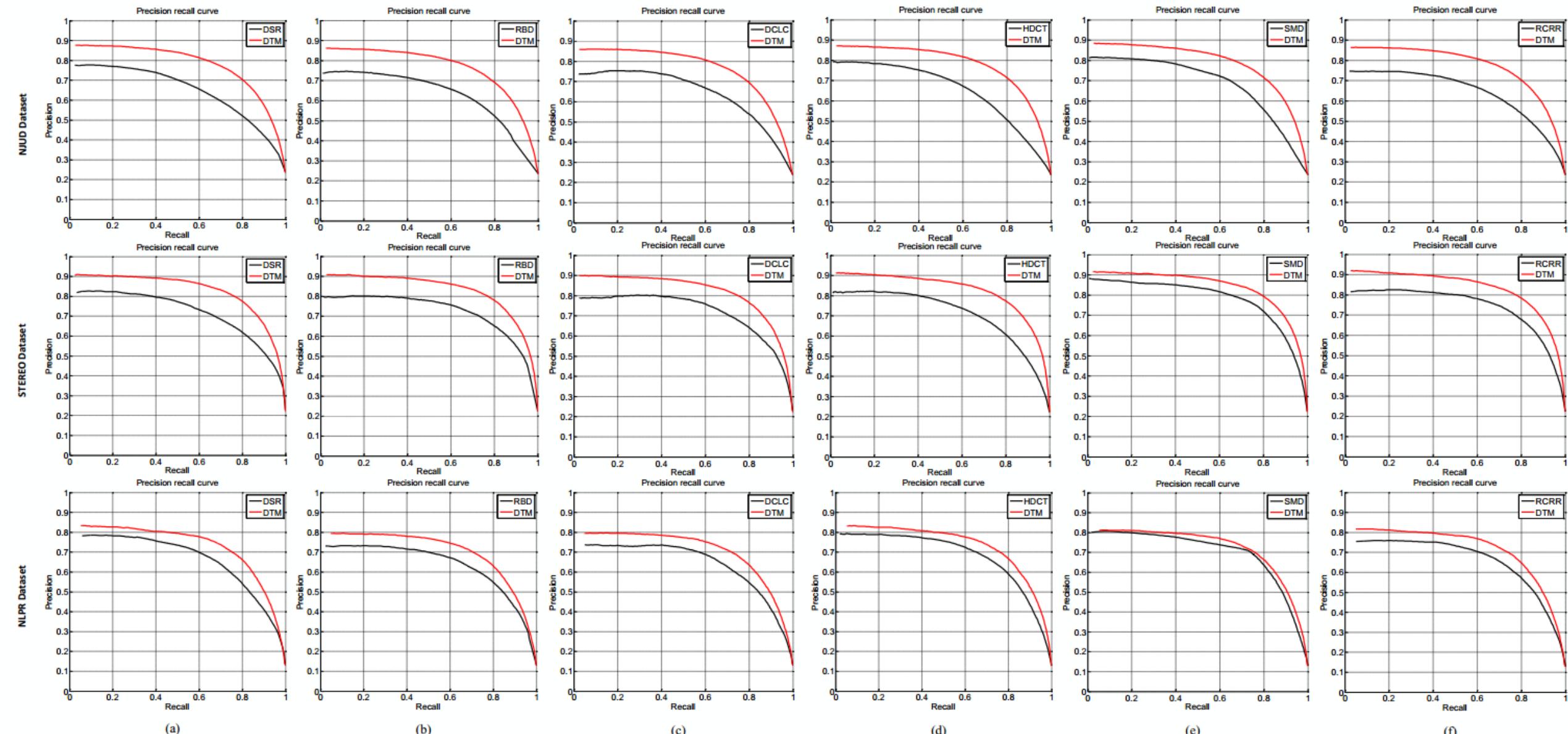
$$\boxed{\mathbf{s}^* = [\mathbf{I} + (\mathbf{D}_c - \mathbf{W}_c) + (\mathbf{D}_d - \mathbf{W}_d)]^{-1} \cdot \mathbf{s}}$$

Going from RGB to RGBD saliency: A depth-guided transformation model

- **Experiments**



Going from RGB to RGBD saliency: A depth-guided transformation model



Going from RGB to RGBD saliency: A depth-guided transformation model

TABLE I

QUANTITATIVE COMPARISONS ON THE NJUD DATASET. ΔPG IS THE PERCENTAGE GAIN BETWEEN THE RGB SALIENCY AND PROPOSED DTM

| | DSR [9] | DTM | ΔPG | RBD [10] | DTM | ΔPG |
|-----------|-----------|--------|-------------|-----------|--------|-------------|
| F_β | 0.6457 | 0.7566 | 17.2% | 0.6433 | 0.7490 | 16.4% |
| AUC | 0.8634 | 0.9174 | 6.3% | 0.8498 | 0.9138 | 7.5% |
| S_m | 0.6321 | 0.7063 | 11.7% | 0.6542 | 0.7059 | 7.9% |
| | DCLC [11] | DTM | ΔPG | HDCT [12] | DTM | ΔPG |
| F_β | 0.6527 | 0.7514 | 15.1% | 0.6581 | 0.7602 | 15.5% |
| AUC | 0.8526 | 0.9081 | 6.5% | 0.8655 | 0.9220 | 6.5% |
| S_m | 0.6188 | 0.7010 | 13.3% | 0.6391 | 0.7099 | 11.1% |
| | SMD [13] | DTM | ΔPG | RCRR [14] | DTM | ΔPG |
| F_β | 0.6900 | 0.7633 | 10.6% | 0.6508 | 0.7548 | 16.0% |
| AUC | 0.8535 | 0.9204 | 6.6% | 0.8500 | 0.9153 | 7.7% |
| S_m | 0.6782 | 0.7170 | 5.7% | 0.6412 | 0.7079 | 10.4% |

TABLE II

QUANTITATIVE COMPARISONS ON THE STEREO DATASET. ΔPG IS THE PERCENTAGE GAIN BETWEEN THE RGB SALIENCY AND PROPOSED DTM

| | DSR [9] | DTM | ΔPG | RBD [10] | DTM | ΔPG |
|-----------|-----------|--------|-------------|-----------|--------|-------------|
| F_β | 0.6974 | 0.7973 | 14.3% | 0.7157 | 0.7997 | 11.7% |
| AUC | 0.9126 | 0.9470 | 3.8% | 0.9149 | 0.9497 | 3.8% |
| S_m | 0.6674 | 0.7335 | 9.9% | 0.7136 | 0.7449 | 4.4% |
| | DCLC [11] | DTM | ΔPG | HDCT [12] | DTM | ΔPG |
| F_β | 0.7150 | 0.7925 | 10.8% | 0.7005 | 0.7954 | 13.5% |
| AUC | 0.9052 | 0.9416 | 4.0% | 0.9087 | 0.9495 | 4.5% |
| S_m | 0.6614 | 0.7314 | 10.6% | 0.6775 | 0.7372 | 8.8% |
| | SMD [13] | DTM | ΔPG | RCRR [14] | DTM | ΔPG |
| F_β | 0.7620 | 0.8077 | 6.0% | 0.7360 | 0.7996 | 8.6% |
| AUC | 0.9260 | 0.9532 | 2.9% | 0.9066 | 0.9490 | 4.7% |
| S_m | 0.7398 | 0.7528 | 1.8% | 0.7039 | 0.7430 | 5.6% |

TABLE III

QUANTITATIVE COMPARISONS ON THE NLPR DATASET. ΔPG IS THE PERCENTAGE GAIN BETWEEN THE RGB SALIENCY AND PROPOSED DTM

| | DSR [9] | DTM | ΔPG | RBD [10] | DTM | ΔPG |
|-----------|-----------|--------|-------------|-----------|--------|-------------|
| F_β | 0.6743 | 0.7326 | 8.6% | 0.6542 | 0.7090 | 8.4% |
| AUC | 0.9252 | 0.9340 | 1.0% | 0.9201 | 0.9295 | 1.0% |
| S_m | 0.7037 | 0.7320 | 4.0% | 0.7124 | 0.7260 | 1.9% |
| | DCLC [11] | DTM | ΔPG | HDCT [12] | DTM | ΔPG |
| F_β | 0.6662 | 0.7134 | 7.1% | 0.6914 | 0.7326 | 6.0% |
| AUC | 0.8992 | 0.9246 | 2.8% | 0.9385 | 0.9397 | 0.1% |
| S_m | 0.6829 | 0.7245 | 6.1% | 0.7108 | 0.7338 | 3.2% |
| | SMD [13] | DTM | ΔPG | RCRR [14] | DTM | ΔPG |
| F_β | 0.7138 | 0.7290 | 2.1% | 0.6783 | 0.7252 | 6.9% |
| AUC | 0.9229 | 0.9334 | 1.1% | 0.9017 | 0.9270 | 2.8% |
| S_m | 0.7303 | 0.7355 | 0.7% | 0.6919 | 0.7273 | 5.1% |

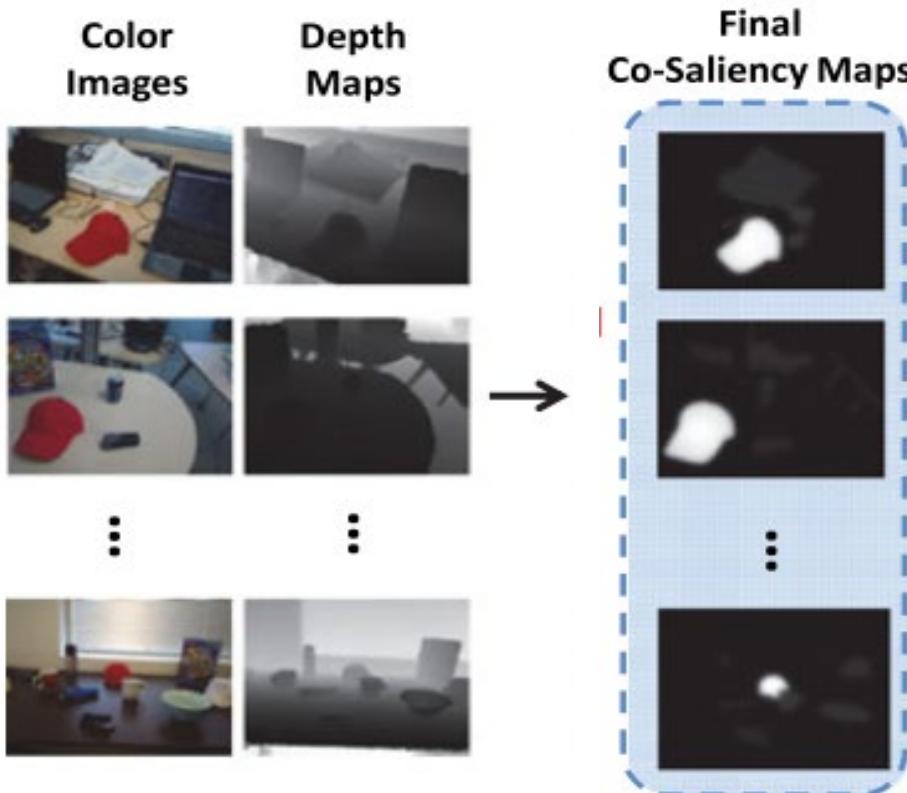
TABLE IV
QUANTITATIVE COMPARISONS OF DIFFERENT RGBD SALIENCY DETECTION METHODS ON THREE DATASETS.

| | NJUD Dataset | | | STEREO Dataset | | | NLPR Dataset | | |
|-----------|--------------|--------|--------|----------------|--------|--------|--------------|--------|--------|
| | F_β | AUC | S_m | F_β | AUC | S_m | F_β | AUC | S_m |
| SS [39] | 0.6128 | 0.8103 | 0.5755 | 0.5478 | 0.7943 | 0.5412 | 0.4712 | 0.8007 | 0.5737 |
| ACSD [42] | 0.7459 | 0.9259 | 0.6987 | 0.7467 | 0.9333 | 0.7082 | 0.6695 | 0.9229 | 0.6825 |
| WSC [43] | 0.6418 | 0.7579 | 0.6325 | 0.6987 | 0.8034 | 0.6727 | 0.6586 | 0.8494 | 0.6955 |
| CDCP [49] | 0.6673 | 0.8699 | 0.6689 | 0.7168 | 0.9065 | 0.7181 | 0.6863 | 0.9175 | 0.7266 |
| MBP [50] | 0.6025 | 0.7231 | 0.5272 | 0.6627 | 0.7701 | 0.5574 | 0.6015 | 0.7852 | 0.6050 |
| LMH [51] | 0.7029 | 0.8489 | 0.5137 | 0.5862 | 0.7360 | 0.4773 | 0.7057 | 0.8947 | 0.6141 |
| DF [52] | 0.6384 | 0.8338 | 0.5881 | 0.6961 | 0.8804 | 0.6279 | 0.6407 | 0.8801 | 0.6610 |
| ours | 0.7633 | 0.9204 | 0.7170 | 0.8077 | 0.9532 | 0.7528 | 0.7290 | 0.9334 | 0.7355 |

TABLE VI
F-MEASURE OF DIFFERENT MODULES ON THE STEREO DATASET.

| Modules | F-measure |
|--|-----------|
| RGB Saliency (DSR) | 0.6974 |
| Multi-level RGBD Saliency Initialization | 0.7050 |
| Depth-guided Saliency Refinement | 0.7697 |
| Saliency Optimization with Depth Constraints | 0.7973 |

RGBD Co-saliency Detection



Problems and Important Issues

how to explore inter-image correspondence among multiple images to constrain the common properties of salient object is a challenge.

how to capture the accurate and effective depth representation to assist in saliency detection is a challenge.

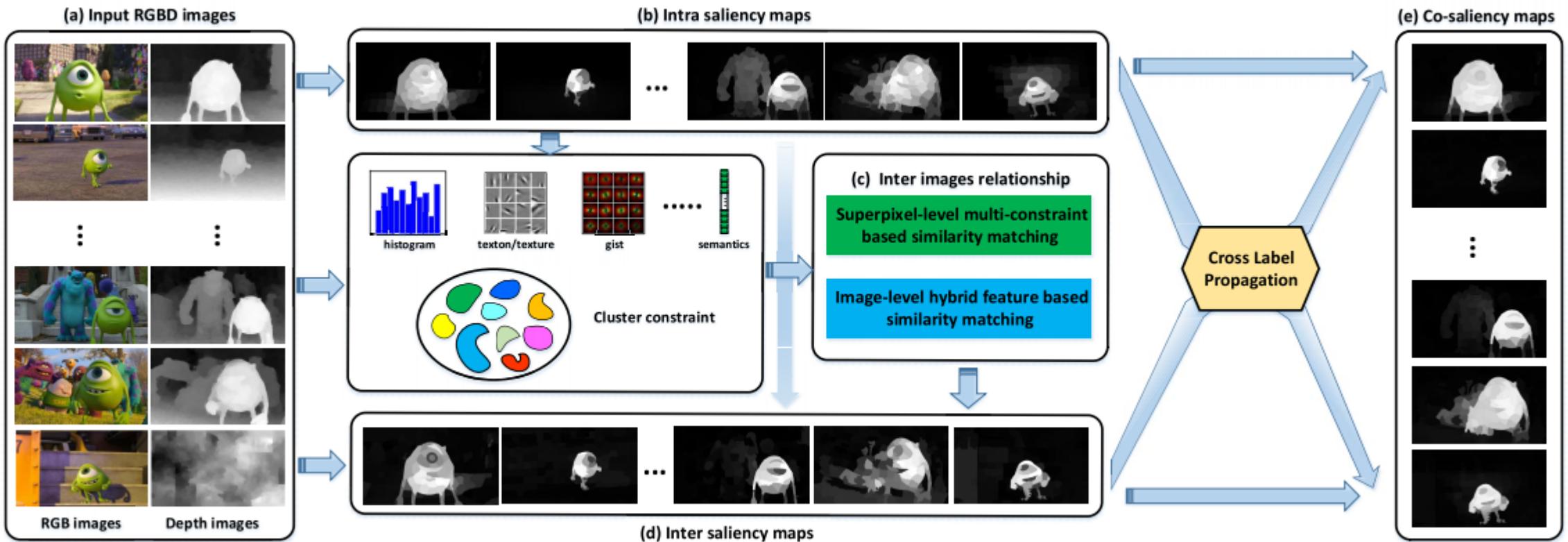
Co-saliency detection for RGBD images based on multi-constraint feature matching and cross label propagation

**Runmin Cong, Jianjun Lei, Huazhu Fu, Qingming Huang,
Xiaochun Cao, Chunping Hou**

IEEE Transactions on Image Processing
vol.27, no.2, pp.568-579, 2018

Co-saliency detection for RGBD images based on multi-constraint feature matching and cross label propagation

- Framework



Co-saliency detection for RGBD images based on multi-constraint feature matching and cross label propagation

- **Contributions**

1. This method is the **first model that detects the co-salient objects from RGBD images**.
The depth information is demonstrated to be served as a useful complement for co-saliency detection.
2. A **multi-constraint feature matching** method is introduced to constrain the inter saliency map generation, which is robust to the complex backgrounds.
3. The **Cross Label Propagation (CLP)** method is proposed to optimize the co-saliency model in a cross manner.
4. We construct a **new RGBD co-saliency dataset**, named **RGBD Cosal150** dataset, for performance evaluation.

Co-saliency detection for RGBD images based on multi-constraint feature matching and cross label propagation

- **Inter Saliency Detection**

- Acquiring the corresponding relationship among multiple images is the key point of co-saliency detection model. In the proposed model, the **matching methods on two levels** are designed to represent the correspondence among multiple images.
- The first one is the superpixel-level similarity matching scheme, which focuses on determining the matching superpixel set for the current superpixel based on three constraints from other images. The second is the image-level similarity measurement, which provides a global relationship on the whole image scale.
- With the corresponding relationship, the inter saliency of a superpixel is defined as the weighted sum of the intra saliency of corresponding superpixels in other images.

Co-saliency detection for RGBD images based on multi-constraint feature matching and cross label propagation

- **Inter Saliency Detection: Superpixel-Level Multi-Constraint Based Similarity Matching**

- At the superpixel level, the correspondence is represented as the **multi-constraint based matching relationship between the superpixels among the multiple images**, which considers the similarity constraint, saliency consistency, and cluster-based constraint.

Similarity constraint:

$$s(r_m^i, r_n^j) = \exp\left(-\frac{\|c_m^i - c_n^j\|_2 + \min(\lambda_d^i, \lambda_d^j) \cdot |d_m^i - d_n^j|}{\sigma^2}\right)$$

the Kmax nearest neighbors in each of other images are determined to form $\Phi_1(r_m^i)$

Saliency constraint:

$$\Phi_2(r_m^i) = \{r_n^j \mid |S_{intra}(r_m^i) - S_{intra}(r_n^j)| \leq T_1\}$$

Cluster-based constraint:

$$\Phi_3(r_m^i) = \{r_n^j \mid \arg \min_{C_q^j, q \in [1, K]} Ed(c_p^i, c_q^j)\}$$

Similarity matching:

$$ml(r_m^i, r_n^j) = \begin{cases} 1, & \text{if } r_n^j \in \{\Phi_1(r_m^i) \cap \Phi_2(r_m^i) \cap \Phi_3(r_m^i)\} \\ 0, & \text{otherwise} \end{cases}$$

Co-saliency detection for RGBD images based on multi-constraint feature matching and cross label propagation

- **Inter Saliency Detection: Image-Level Hybrid Feature Based Similarity Matching**
 - Enlightened by the observation that the **greater similarity between two images means the greater likelihood of finding the matching regions**, a full-image size similarity descriptor is designed as the weighted coefficient for inter saliency calculation.

| | features | description | dim | distance |
|-----|----------------|--------------------|------|---|
| col | \mathbf{h}_c | RGB histogram | 512 | $d_{c1} = \chi^2(\mathbf{h}_c^i, \mathbf{h}_c^j)$ |
| | \mathbf{t} | texton histogram | 15 | $d_{c2} = \chi^2(\mathbf{t}^i, \mathbf{t}^j)$ |
| | \mathbf{s} | semantic feature | 4096 | $d_{c3} = 1 - \cos(\mathbf{s}^i, \mathbf{s}^j)$ |
| | \mathbf{g} | GIST feature | 512 | $d_{c4} = 1 - \cos(\mathbf{g}^i, \mathbf{g}^j)$ |
| dep | \mathbf{h}_d | depth histogram | 512 | $d_d = \chi^2(\mathbf{h}_d^i, \mathbf{h}_d^j)$ |
| sal | \mathbf{h}_s | saliency histogram | 512 | $d_s = \chi^2(\mathbf{h}_s^i, \mathbf{h}_s^j)$ |

$$\varphi^{ij} = 1 - (\alpha_c \cdot \sum_{i=1}^4 d_{ci}/4 + \alpha_d \cdot d_d + \alpha_s \cdot d_s)$$

$$\alpha_d = \begin{cases} \lambda_d^{min}, & \text{if } \lambda_d^{min} = \min(\lambda_d^i, \lambda_d^j) \leq T_2 \\ 1/3, & \text{otherwise} \end{cases}$$

$$\alpha_c = \alpha_s = \frac{1}{2} \cdot (1 - \alpha_d)$$

Co-saliency detection for RGBD images based on multi-constraint feature matching and cross label propagation

- **Inter Saliency Detection**

- After obtaining the corresponding relationship among multiple images through the superpixel-level feature matching and image-level similarity matching, the inter saliency of a superpixel is computed as the **weighted sum of the intra saliency of corresponding superpixels in other images.**

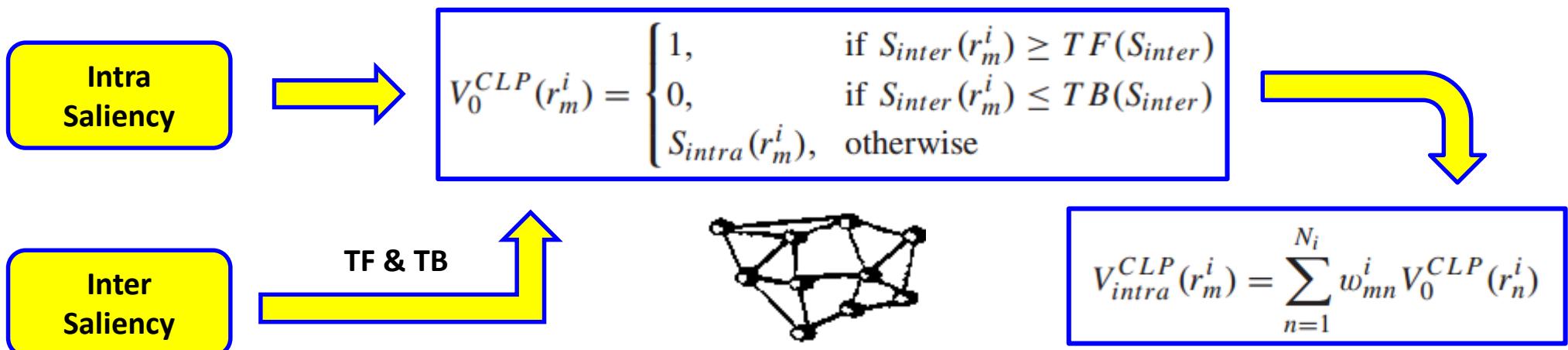
$$S_{inter}(r_m^i) = \frac{1}{N-1} \sum_{j=1, j \neq i}^N \frac{\varphi^{ij}}{N_j} \sum_{n=1}^{N_j} S_{intra}(r_n^j) \cdot ml(r_m^i, r_n^j)$$

where r_m^i denotes the m^{th} superpixel in image I^i , N represents the number of images in the group, N_j is the number of superpixels in the j^{th} image, and φ^{ij} is the similarity measurement between the i^{th} and j^{th} images.

Co-saliency detection for RGBD images based on multi-constraint feature matching and cross label propagation

- Optimization and Propagation

- In the proposed method, the optimization of saliency map is casted as a “**label propagation**” problem, where the uncertain labels are propagated by using **two types of certain seeds**, i.e. background and salient seeds. **The proposed CLP method is used to optimize the intra and inter saliency maps in a cross way, which means the propagative seeds are crosswise interacted.** The cross seeding strategy optimizes the intra and inter saliency maps jointly, and improves the robustness.



Co-saliency detection for RGBD images based on multi-constraint feature matching and cross label propagation

- **Experiments**

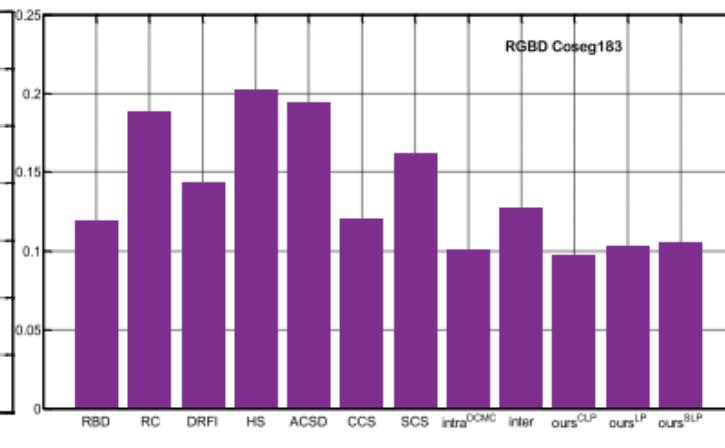
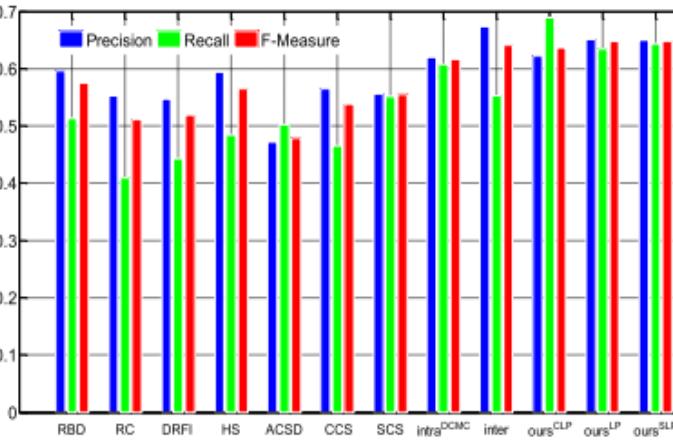
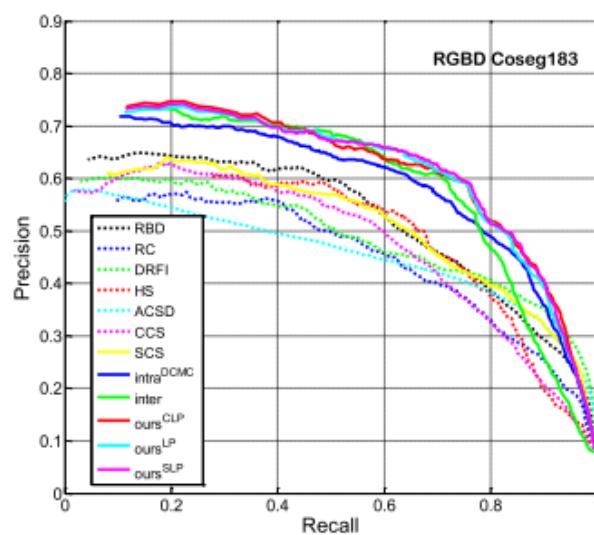
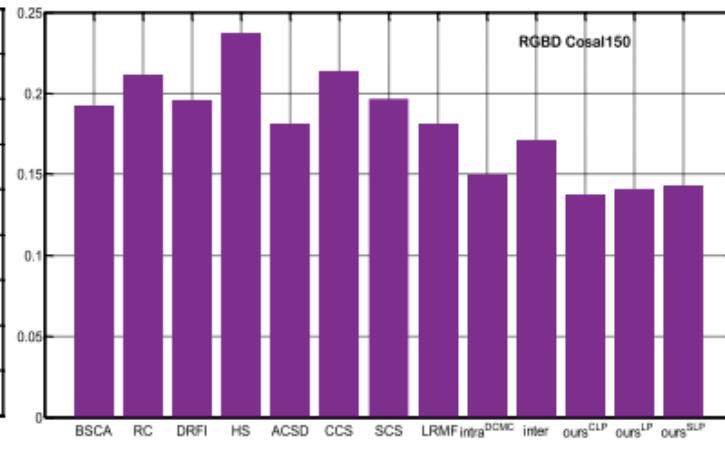
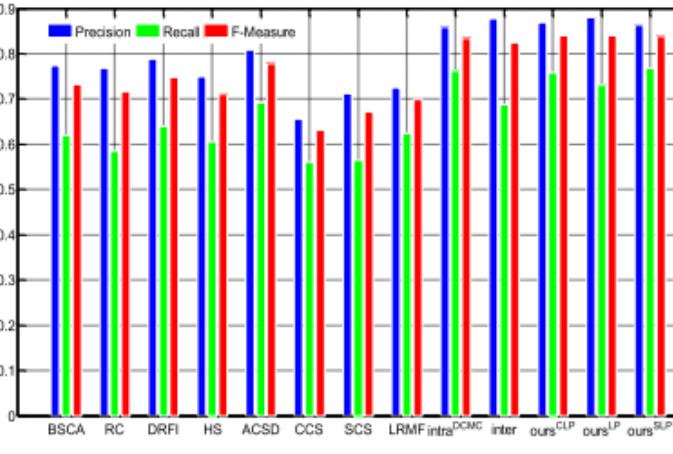
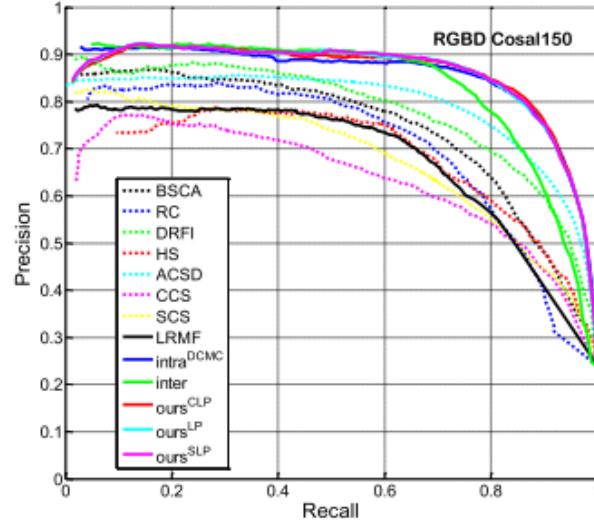


Co-saliency detection for RGBD images based on multi-constraint feature matching and cross label propagation

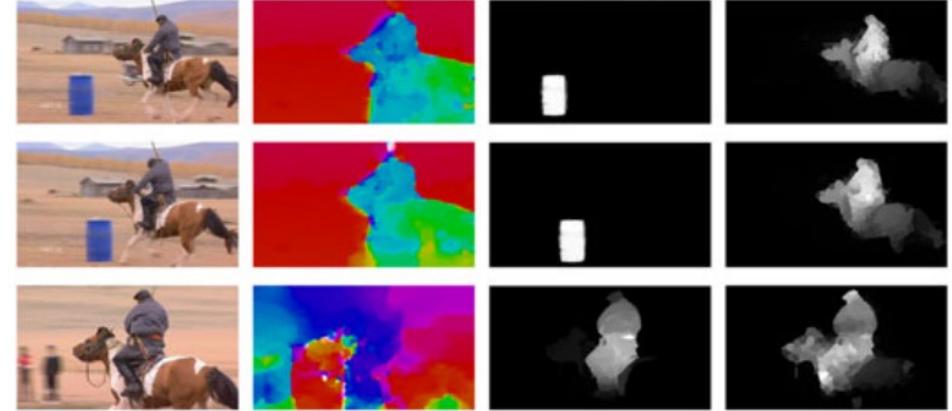
- Experiments

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

$$MAE = \frac{1}{w \times h} \sum_{x=1}^w \sum_{y=1}^h |S(x, y) - G(x, y)|$$



Video Saliency Detection



Problems and Important Issues:

Motion cue plays more important role in discovering the salient object from the clustered and complex scene.



The inter-frame correspondence is used to capture the common attribute of salient objects from the whole video.



The spatiotemporal consistency constrains the smoothness and homogeneity of salient objects from the spatiotemporal domain.



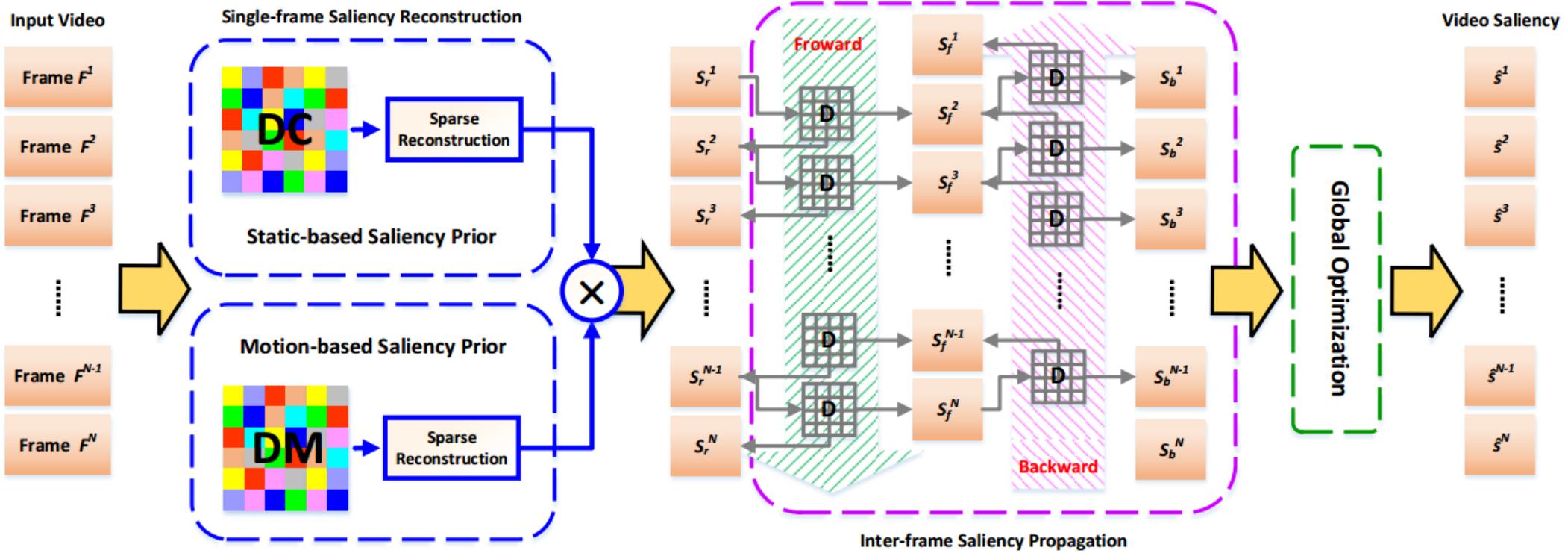
Video saliency detection via sparsity-based reconstruction and propagation

**Runmin Cong, Jianjun Lei, Huazhu Fu, Fatih Porikli,
Qingming Huang, Chunping Hou**

IEEE Transactions on Image Processing, 2019

Video saliency detection via sparsity-based reconstruction and propagation

- Framework



Video saliency detection via sparsity-based reconstruction and propagation

- **Contributions**

1. A novel **sparsity-based saliency reconstruction** is introduced to generate single-frame saliency map, making the best use of the static and motion priors. The motion priors are defined as motion compactness cue and motion uniqueness cue.
2. A new and efficient **sparsity-based saliency propagation** is presented to capture the correspondence in the temporal space and produce inter-frame saliency map. The salient object is sequentially reconstructed by the forward and backward dictionaries.
3. To attain the global consistency of the salient object in the whole video, a **global optimization model**, which integrates unary data term, spatiotemporal smooth term, spatial incompatibility term, and global consistency term, is formulated.

Video saliency detection via sparsity-based reconstruction and propagation

- **Single-frame Saliency Reconstruction**

- For video saliency detection, the detected object should be salient with respect to the background and underlying motion in each frame. To this end, a sparse reconstruction model with two saliency priors is used to detect the salient object in each individual frame. **The first one is the static saliency prior, which utilizes three color saliency cues to construct a color-based reconstruction dictionary (DC). The second one is the motion-based saliency prior, which integrates the motion uniqueness cue and motion compactness cue to build a motion-based dictionary (DM).**
- Given a video sequence $\mathbf{F} = \{F^t\}_{t=1}^N$ including N frames F^t , we firstly derive some homogeneous superpixels $\mathbf{R} = \{r_k^t\}_{k=1}^{N^t}$ using SLIC algorithm for each frame F^t , where N^t is the number of superpixels. In addition, the large displacement optical flow is calculated to represent the pixel-level motion vector. The motion vector v_k^t of superpixel r_k^t is defined as the mean value of pixel-level motion vector in the superpixel.

Video saliency detection via sparsity-based reconstruction and propagation

- **Single-frame Saliency Reconstruction: Static-based saliency prior**

- The static-based saliency prior measures the static saliency in each frame by incorporating the background dictionary into a sparse representation framework. Three color-based cues, including background cue, compactness cue, and uniqueness cue, are integrated to select the background seeds and build the dictionary for reconstruction.
- **Background Cue.** The superpixels along the image boundaries are selected as the background candidate set Φ_{SB}^t that represents the spatial location attribute of the background regions.
- **Static Compactness Cue.** The salient regions incline to have a small spatial variance, whereas the backgrounds usually have a high spatial variance since their superpixels are often distributed over the entire image. Then, the top Q_1 superpixels with larger spatial variances are selected as the compactness-based background candidate set Φ_{SC}^t .
$$v_s(r_k^t) = \frac{\sum_{l=1}^{N^t} a_{kl}^t \cdot n_l^t \cdot \|\mathbf{p}_l^t - \mathbf{u}_k^t\|_2}{\sum_{l=1}^{N^t} a_{kl}^t \cdot n_l^t}$$
$$a_{kl}^t = \exp(-\|\mathbf{l}\mathbf{c}_k^t - \mathbf{l}\mathbf{c}_l^t\|_2/\sigma^2)$$

Video saliency detection via sparsity-based reconstruction and propagation

- **Single-frame Saliency Reconstruction: Static-based saliency prior**

- **Static Uniqueness Cue.** The third cue represents the global appearance of the background regions in which the salient object shows different properties in appearance compared with the background. In our work, a **cluster-based method** is proposed to define the uniqueness cue. First, K-means++ clustering is used to group the superpixels into K clusters. Then, two clusters with the largest Euclidean distance are selected. The selected two clusters correspond to one foreground cluster and one background cluster. Finally, a decision scheme considering the spatial variance and background probability is designed to determine the **uniqueness-based background candidate set** Φ_{SU}^t .

$$\Phi_{SU}^t = \begin{cases} \{C_p^t\}, & \text{if } [v_s(C_p^t) > v_s(C_q^t)] \cap [P_b(C_p^t) > P_b(C_q^t)] \\ \{C_q^t\}, & \text{if } [v_s(C_p^t) \leq v_s(C_q^t)] \cap [P_b(C_p^t) \leq P_b(C_q^t)] \\ \emptyset, & \text{otherwise} \end{cases}$$

Video saliency detection via sparsity-based reconstruction and propagation

- **Single-frame Saliency Reconstruction: Static-based saliency prior**

- **Static-based Saliency Reconstruction.** The final background set is obtained by combining all background candidates as $\Phi_{CB}^t = \Phi_{SB}^t \cup \Phi_{SC}^t \cup \Phi_{SU}^t$. Then, three types of features considering the color components, spatial location, and texture distribution are used to describe each superpixel as $\mathbf{x}_k^t = [\mathbf{c}_k^t \ \mathbf{p}_k^t \ \mathbf{t}_k^t]$. The background dictionary \mathbf{D}_B^t is constructed by the feature representations of the stacking background seeds in Φ_{CB}^t . **Based on the assumption that reconstruction error should be different for foreground and background through a sparse reconstruction model, the image saliency can be measured by the reconstruction error.** Each superpixel r_k^t is encoded by:

$$\alpha_k^{t*} = \arg \min_{\alpha_k^t} \|\mathbf{x}_k^t - \mathbf{D}_B^t \cdot \alpha_k^t\|_2^2 + \lambda \cdot \|\alpha_k^t\|_1$$

- The saliency of superpixel r_k^t can be measured by the reconstruction error:

$$S_s(r_k^t) = \varepsilon_k^t = \|\mathbf{x}_k^t - \mathbf{D}_B^t \cdot \alpha_k^{t*}\|_2^2$$

Video saliency detection via sparsity-based reconstruction and propagation

- **Single-frame Saliency Reconstruction: Motion-based saliency prior**

- Moving target attracts more attention in visual perception, thus, we introduce a motion-based saliency prior to represent the salient object from the perspective of motion space.
- **The spatial distribution of moving object is more concentrated than the background regions in the optical flow data. In addition, the moving object is often different from the background regions in terms of the magnitude of optical flow (MOF), which is consistent with the uniqueness cue in the color space.**
- Based on these observations, we extend the color-related cues to the motion field and determine the background seeds for dictionary construction.

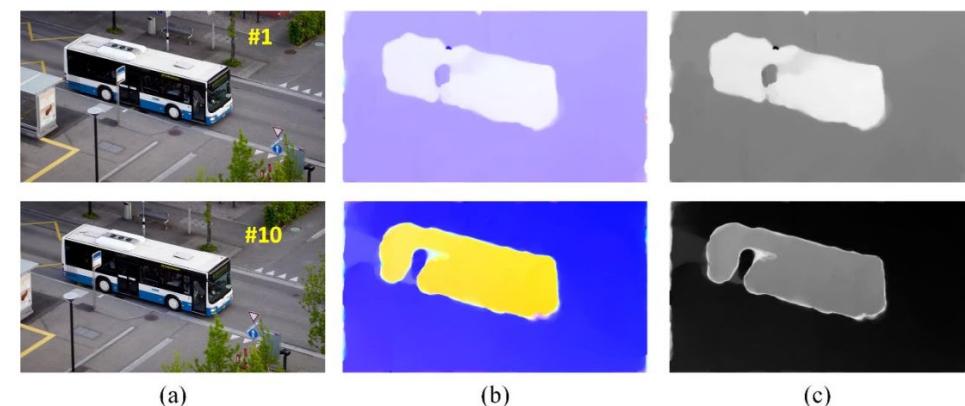


Fig. 7-2 Optical flow data of different video frames. (a) RGB image. (b) Optical flow map. (c) The MOF data.

Video saliency detection via sparsity-based reconstruction and propagation

- **Single-frame Saliency Reconstruction: Motion-based saliency prior**

- **Motion Compactness Cue.** We introduce a “motion compactness” cue to describe the distribution of the optical flow data and determine the background candidates.

$$v_m(r_k^t) = \frac{\sum_{l=1}^{N^t} m_{kl}^t \cdot n_l^t \cdot \|\mathbf{p}_l^t - \tilde{\mathbf{u}}_k^t\|_2}{\sum_{l=1}^{N^t} a_{kl}^t \cdot n_l^t}$$

$$m_{kl}^t = \exp(-\|\mathbf{v}_k^t - \mathbf{v}_l^t\|_2 / \sigma^2)$$

TOP Q_1 superpixels
with larger values

$$\begin{cases} ux_k^t = \frac{\sum_{l=1}^{N^t} m_{kl}^t \cdot n_l^t \cdot x_l^t}{\sum_{l=1}^{N^t} m_{kl}^t \cdot n_l^t} \\ uy_k^t = \frac{\sum_{l=1}^{N^t} m_{kl}^t \cdot n_l^t \cdot y_l^t}{\sum_{l=1}^{N^t} m_{kl}^t \cdot n_l^t} \end{cases}$$



$$\Phi_{MC}^t$$

- **Motion Uniqueness Cue.** In general, the moving target exhibits different motion appearance compared with the background regions in the MOF data. We define a “motion uniqueness” cue in the MOF field by calculating the global contrast of each superpixel.

$$u_m(r_k^t) = \sum_{k=1, k \neq l}^{N^t} |M_f(r_k^t) - M_f(r_l^t)| \cdot e^{-E_d(\mathbf{p}_k^t, \mathbf{p}_l^t) / \sigma^2}$$

TOP Q_1 superpixels
with smaller values



$$\Phi_{MU}^t$$

Video saliency detection via sparsity-based reconstruction and propagation

- **Single-frame Saliency Reconstruction: Motion-based saliency prior**
 - **Motion-based Saliency Reconstruction.** For the motion-based sparse reconstruction, the motion feature is necessarily introduced to represent the motion cue. Furthermore, in order to guarantee the robustness of the feature representation, the basic color components are also embedded into the feature pool. Each superpixel is represented as a 12-dimensional feature vector $\mathbf{x}_k^t = [\mathbf{c}_k^t \ \mathbf{m}_k^t]$. Then, the feature representation of each motion-related background seed in $\Phi_{MB}^t = \Phi_{MC}^t \cup \Phi_{MU}^t$ is used to construct the background dictionary for frame F^t as \mathbf{D}_B^t . At last, as same as the static-based saliency reconstruction, the motion saliency of each superpixel is represented by the reconstruction error, which is denoted as $S_m(r_k^t)$.
- **Single-frame Saliency Map**

$$S_r(r_k^t) = S_s(r_k^t) \cdot S_m(r_k^t)$$

Video saliency detection via sparsity-based reconstruction and propagation

- **Inter-frame Saliency Propagation**

- The sequential relationship across the time axis is crucial to video saliency detection. The salient object in an individual frame should be further discriminated by using the inter-frame information.
- **Considering the high consistency and smoothness of the salient object in appearances and views between two adjacent frames, the previous frame can be employed to build a foreground dictionary and reconstruct the current frame in a forward way. Likewise, the current frame can be reconstructed by the next frame in a backward propagation manner.**
- Therefore, a spatiotemporal saliency model is established via **sparse propagation with a forward-backward strategy** to smooth the salient object and suppress the background.

Video saliency detection via sparsity-based reconstruction and propagation

- **Inter-frame Saliency Propagation: Forward propagation**

- In the forward propagation, the current frame is reconstructed by a foreground dictionary derived from the previous frame, and the video is sequentially processed from the first frame to the last frame.

Top Q_2 superpixels with larger single-frame saliency values in frame F^{t-1} are selected as the foreground seeds

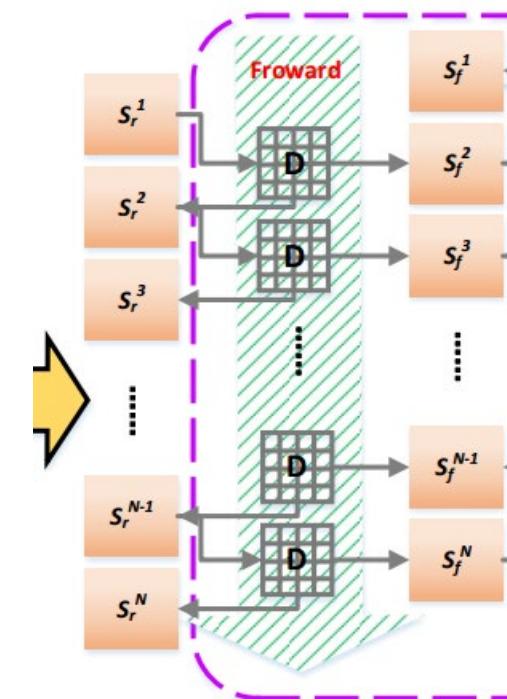


$$x_k^t = [c_k^t \ p_k^t \ t_k^t \ m_k^t \ S_r(r_k^t)]$$

The feature representations of all foreground seeds from frame F^{t-1} are stacked to construct the forward foreground dictionary for frame F^t , which is denoted as \mathbf{D}_F^{t-1} .



$$S_f(r_k^t) = \exp(-\overline{\varepsilon_k^t}/\sigma^2) = \exp(-\|\mathbf{x}_k^t - \mathbf{D}_F^{t-1} \cdot \overline{\alpha_k^{t*}}\|_2^2/\sigma^2)$$



Video saliency detection via sparsity-based reconstruction and propagation

- **Inter-frame Saliency Propagation: Backward propagation**

- In the forward propagation, the current frame is reconstructed by a foreground dictionary derived from the previous frame, and the video is sequentially processed from the first frame to the last frame.

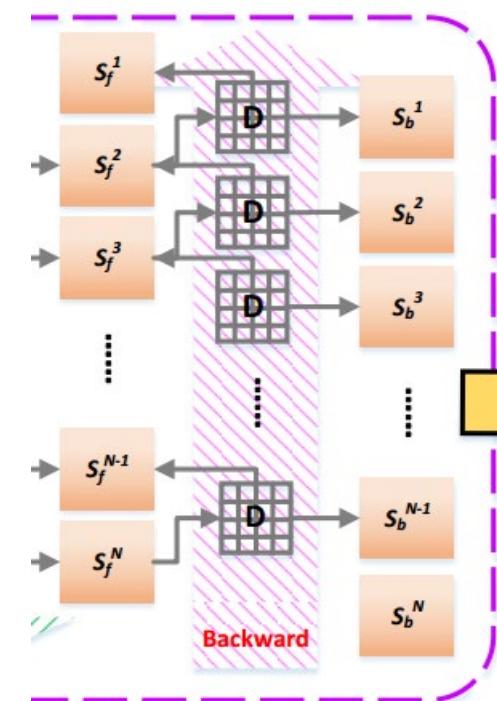
Top $Q_2/2$ superpixels with larger saliency values in the single-frame and forward saliency models are selected, respectively.

$$\downarrow \quad \mathbf{x}_k^t = [\mathbf{c}_k^t \ \mathbf{p}_k^t \ \mathbf{t}_k^t \ \mathbf{m}_k^t \ S_r(r_k^t) \ S_f(r_k^t)]$$

The feature representations of all foreground seeds from frame F^{t+1} are stacked to construct the forward foreground dictionary for frame F^t , which is denoted as \mathbf{D}_F^{t+1} .



$$S_b(r_k^t) = \exp(-\tilde{\varepsilon}_k^t/\sigma^2) = \exp(-\|\mathbf{x}_k^t - \mathbf{D}_F^{t+1} \cdot \tilde{\alpha}_k^{t*}\|_2^2/\sigma^2)$$



Video saliency detection via sparsity-based reconstruction and propagation

- **Global Optimization**

$$E = \eta_1 \cdot E_u + \eta_2 \cdot E_s + \eta_3 \cdot E_i + \eta_4 \cdot E_g$$

- In order to achieve superior and globally consistent saliency map, we propose an efficient optimization model with an energy function that consists of four complementary terms.

$$E_u = \sum_k (\hat{s}_k^t - s_k^t)^2$$

This term encourages the similarity between the final saliency map and initial saliency map;

$$E_s = \sum_{(k,l) \in \Omega_{st}} \omega_{kl} \cdot (\hat{s}_k^t - \hat{s}_l^t)^2$$

This term favors that all the similar and spatiotemporally adjacent superpixels across the whole video should be assigned to consistent saliency scores;

$$E_i = \sum_{(k,l) \in \Omega_s} \omega_{kl} \cdot \hat{s}_k^t \cdot \hat{s}_l^t$$

This term enforces that the same region should not have high foreground and background probabilities simultaneously;

$$E_g = \sum_k \kappa_k \cdot \hat{s}_k^t$$

This term is proposed to constrain the consistency from the global perspective, which imposes the appearance of salient object approximate to a global video foreground model;

Video saliency detection via sparsity-based reconstruction and propagation

- **Global Optimization**

- In order to achieve superior and globally consistent saliency map, we propose an efficient optimization model with an energy function that consists of four complementary terms.

$$E = \eta_1 \cdot E_u + \eta_2 \cdot E_s + \eta_3 \cdot E_i + \eta_4 \cdot E_g$$



$$\begin{aligned} \mathbf{E} = & \eta_1 \cdot (\hat{\mathbf{s}} - \mathbf{s})^T \cdot (\hat{\mathbf{s}} - \mathbf{s}) + \eta_2 \cdot \hat{\mathbf{s}}^T \cdot (\mathbf{D}_{st} - \mathbf{W}_{st}) \cdot \hat{\mathbf{s}} \\ & + \eta_3 \cdot \hat{\mathbf{s}}^T \cdot \mathbf{W}_s \cdot \hat{\mathbf{s}} + \eta_4 \cdot \hat{\mathbf{s}}^T \cdot \mathbf{K} \cdot \hat{\mathbf{s}} \end{aligned}$$

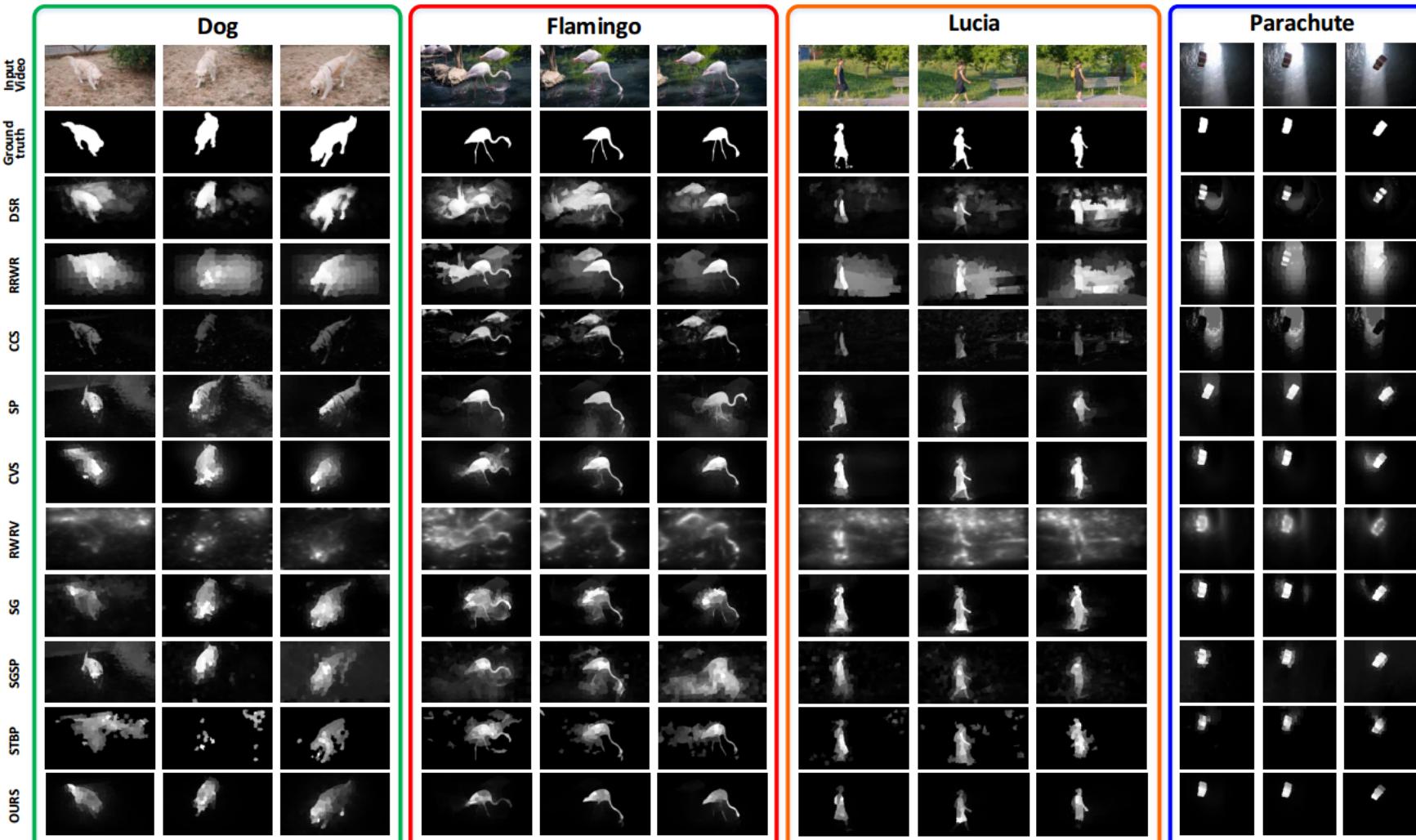
$$\begin{aligned} \mathbf{s} &= [s_k]_{N_a \times 1} \quad \hat{\mathbf{s}} = [\hat{s}_k]_{N_a \times 1} \quad N_a = \sum_{i=1}^N N^i \quad \mathbf{W}_{st} = [\omega_{kl}]_{N_a \times N_a}^{(k,l) \in \Omega_{st}} \quad \mathbf{W}_s = [\omega_{kl}]_{N_a \times N_a}^{(k,l) \in \Omega_s} \\ \mathbf{D}_{st} &= diag(d_1, d_2, \dots, d_{N_a}) \quad d_i = \sum_{j=1, (i,j) \in \Omega_{st}}^{N_a} \omega_{ij} \quad \mathbf{K} = diag(\kappa_1, \kappa_2, \dots, \kappa_{N_a}) \end{aligned}$$



$$\boxed{\hat{\mathbf{s}} = [\eta_1 \cdot \mathbf{I} + \eta_2 \cdot (\mathbf{D}_{st} - \mathbf{W}_{st}) + \eta_3 \cdot \mathbf{W}_s + \eta_4 \cdot \mathbf{K}]^{-1} \cdot (\eta_1 \cdot \mathbf{s})}$$

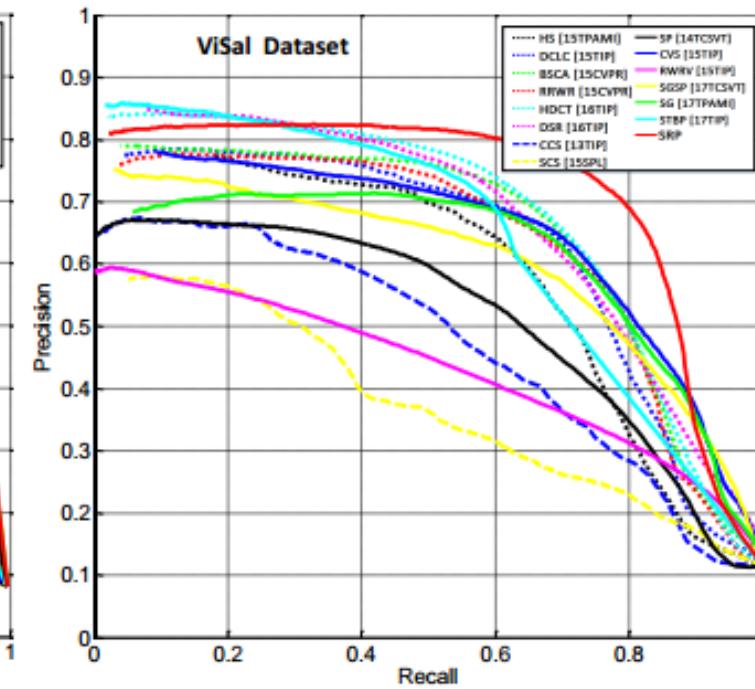
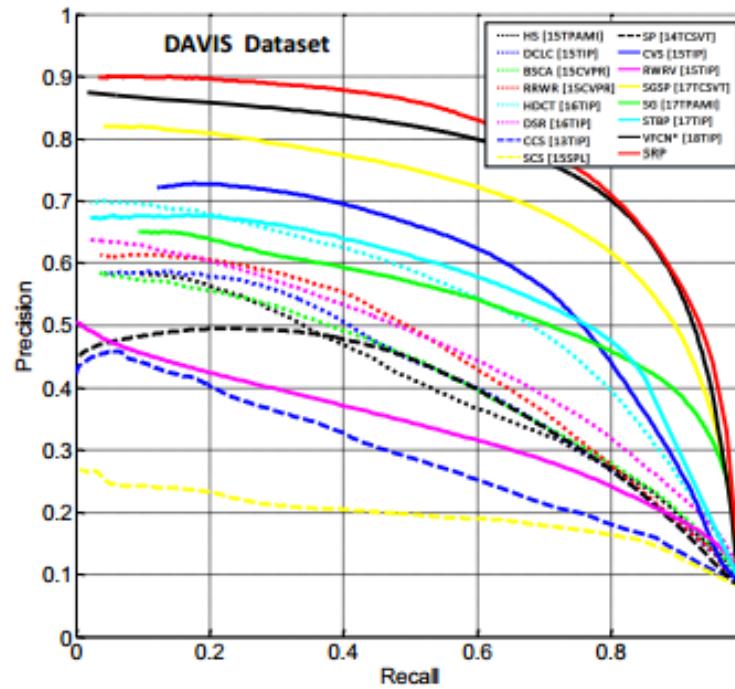
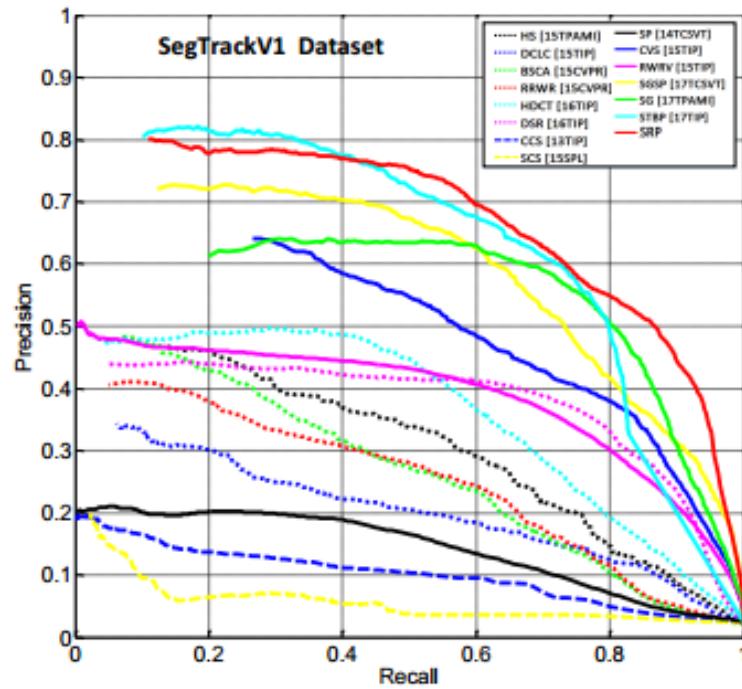
Video saliency detection via sparsity-based reconstruction and propagation

- **Experiments**



Video saliency detection via sparsity-based reconstruction and propagation

• Experiments



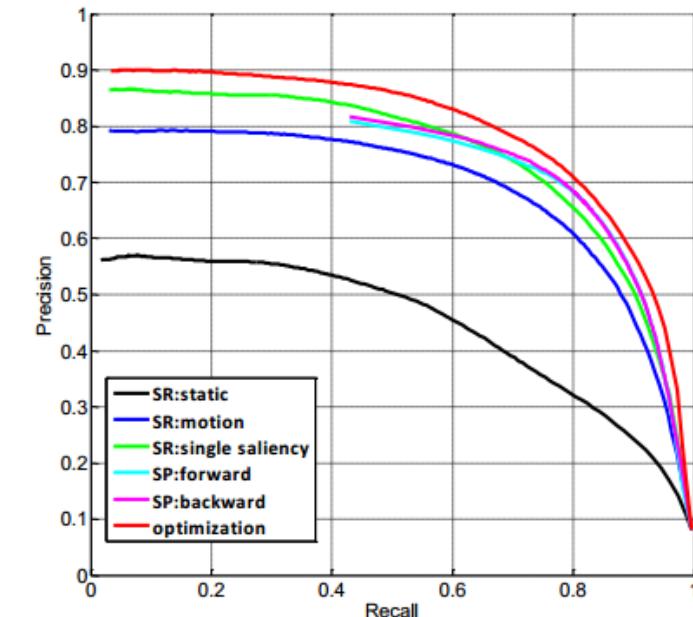
Video saliency detection via sparsity-based reconstruction and propagation

- Experiments

| | SegTrackV1 Dataset | | DAVIS Dataset | | ViSal Dataset | |
|------------|--------------------|---------------|---------------|---------------|---------------|---------------|
| | F-measure | MAE | F-measure | MAE | F-measure | MAE |
| DCLC [9] | 0.2755 | 0.1496 | 0.4783 | 0.1350 | 0.6700 | 0.1265 |
| DSR [11] | 0.4445 | 0.1305 | 0.4972 | 0.1303 | 0.6923 | 0.1061 |
| RRWR [14] | 0.3267 | 0.1963 | 0.5089 | 0.1693 | 0.6707 | 0.1690 |
| HS [15] | 0.3821 | 0.3142 | 0.4523 | 0.2505 | 0.6442 | 0.2019 |
| BSCA [16] | 0.3579 | 0.2366 | 0.4680 | 0.1957 | 0.6949 | 0.1703 |
| HDCT [17] | 0.4681 | 0.1268 | 0.5664 | 0.1346 | 0.7047 | 0.1282 |
| CCS [40] | 0.1486 | 0.1437 | 0.3476 | 0.1510 | 0.5317 | 0.1427 |
| SCS [41] | 0.1137 | 0.2664 | 0.2307 | 0.2567 | 0.4384 | 0.2523 |
| SP [26] | 0.2159 | 0.1195 | 0.4616 | 0.1430 | 0.5723 | 0.1510 |
| CVS [27] | 0.5370 | 0.1085 | 0.6212 | 0.1004 | 0.6676 | 0.1139 |
| RWRV [49] | 0.4458 | 0.1511 | 0.3776 | 0.2001 | 0.4662 | 0.1903 |
| SG [50] | 0.6218 | 0.0810 | 0.5553 | 0.1034 | 0.6640 | 0.1129 |
| SGSP [52] | 0.6275 | 0.1258 | 0.6911 | 0.1374 | 0.6226 | 0.1772 |
| STBP [53] | 0.6583 | 0.0342 | 0.5848 | 0.1015 | 0.6815 | 0.0987 |
| VFCN* [55] | — | — | 0.7488 | 0.0588 | — | — |
| SRP | 0.6830 | 0.0949 | 0.7652 | 0.0688 | 0.7517 | 0.0924 |

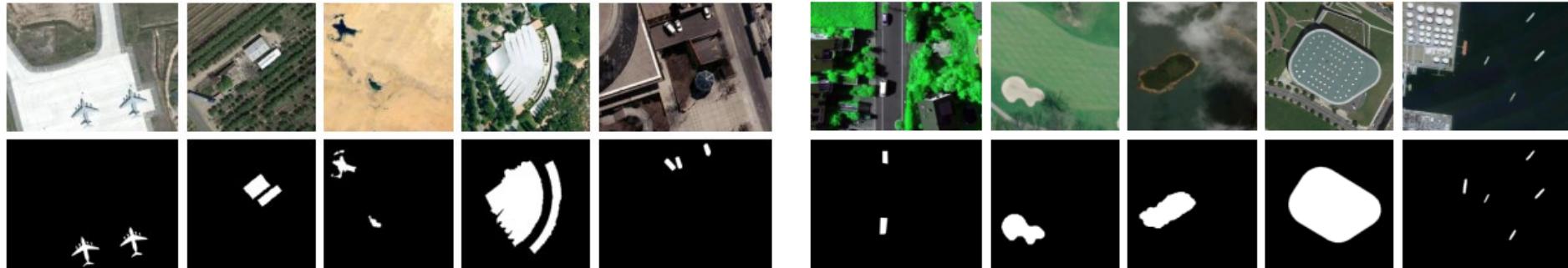
$$F_{\beta} = \frac{(1 + \beta^2) Precision \times Recall}{\beta^2 \times Precision + Recall}$$

$$MAE = \frac{1}{w \times h} \sum_{x=1}^w \sum_{y=1}^h |S(x, y) - G(x, y)|$$



| | Modules | F-measure | MAE |
|---------------------|----------------------|-----------|--------|
| SR | Static Saliency | 0.5029 | 0.1206 |
| | Motion Saliency | 0.6971 | 0.0807 |
| | Single Saliency | 0.7358 | 0.0712 |
| SP | Forward Propagation | 0.7318 | 0.0924 |
| | Backward Propagation | 0.7381 | 0.0793 |
| Global Optimization | | 0.7652 | 0.0688 |

Saliency Detection in Optical Remote Sensing Image



1

Optical RSI may include diversely scaled objects, various scenes and object types, cluttered backgrounds, and shadow noises.

2

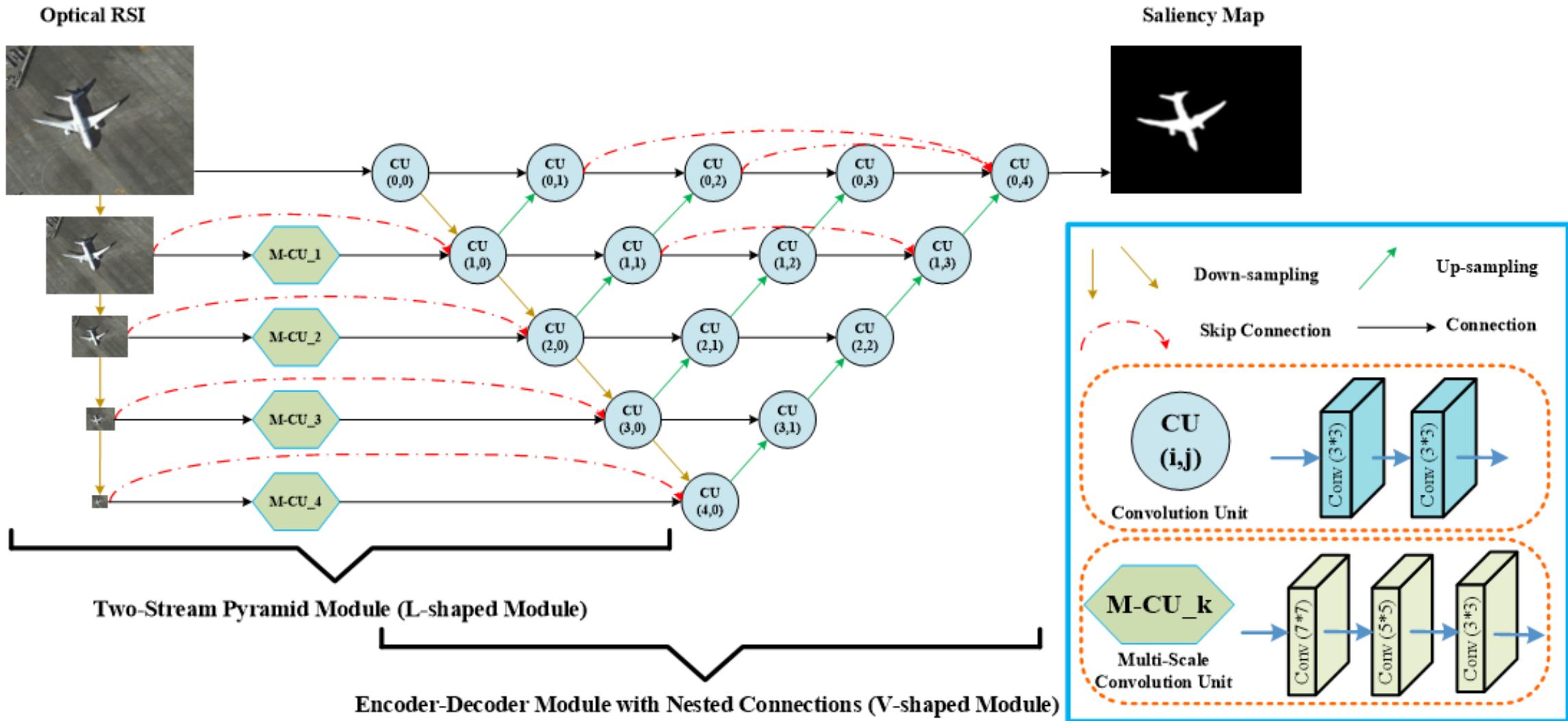
Sometimes, there is even no salient region in a real outdoor scene, such as the desert, forest, and sea.

Nested network with two-stream pyramid for salient object detection in optical remote sensing images

**Chongyi Li, Runmin Cong (co-first and corresponding author),
Junhui Hou, Sanyi Zhang, Yue Qian, Sam Kwong**

IEEE Transaction on Geoscience and Remote Sensing, 2019

Nested Network with Two-Stream Pyramid for Salient Object Detection in Optical Remote Sensing Images



Nested Network with Two-Stream Pyramid for Salient Object Detection in Optical Remote Sensing Images

• Contributions

1. An **end-to-end network** for salient object detection in optical RSIs is proposed, including a **two-stream pyramid module** (L-shaped module) and an **encoder-decoder module with nested connections** (V-shaped module), which generalizes well to varying scenes and object patterns.
2. The **L-shaped module** learns a set of **complementary features** to address the scale variability of salient objects and capture local details, and the **V-shaped module automatically determines the discriminative features** to suppress cluttered backgrounds and highlight salient objects.
3. A **challenging optical RSI dataset** for salient object detection is constructed, including 800 images with the pixel-wise ground truth. Moreover, the proposed method achieves the **best performance** against fourteen state-of-the-art salient object detection methods.

Nested Network with Two-Stream Pyramid for Salient Object Detection in Optical Remote Sensing Images

- **ORSSD Dataset**

- We collected **800 optical RSIs** to construct a dataset for salient object detection, named ORSSD dataset, and the manually **pixel-wise annotation** for each image is provided. The ORSSD dataset is very challenging, because a) **the spatial resolution is diverse**, such as 1264×987 , 800×600 , and 256×256 , b) **the background is cluttered and complicated**, including some shadows, trees, and buildings, c) **the type of salient objects is various**, including airplane, ship, car, river, pond, bridge, stadium, beach, etc, and d) **the number and size of salient objects are variable**, even in some scenes there are no salient object, such as the desert and thick forest.
- In experiments, we randomly selected 600 images from ORSSD dataset for training and the rest 200 images as the testing dataset. The ORSSD dataset is available from our project https://lichongyi.github.io/proj_optical_saliency.html.

Nested Network with Two-Stream Pyramid for Salient Object Detection in Optical Remote Sensing Images

• Experiments

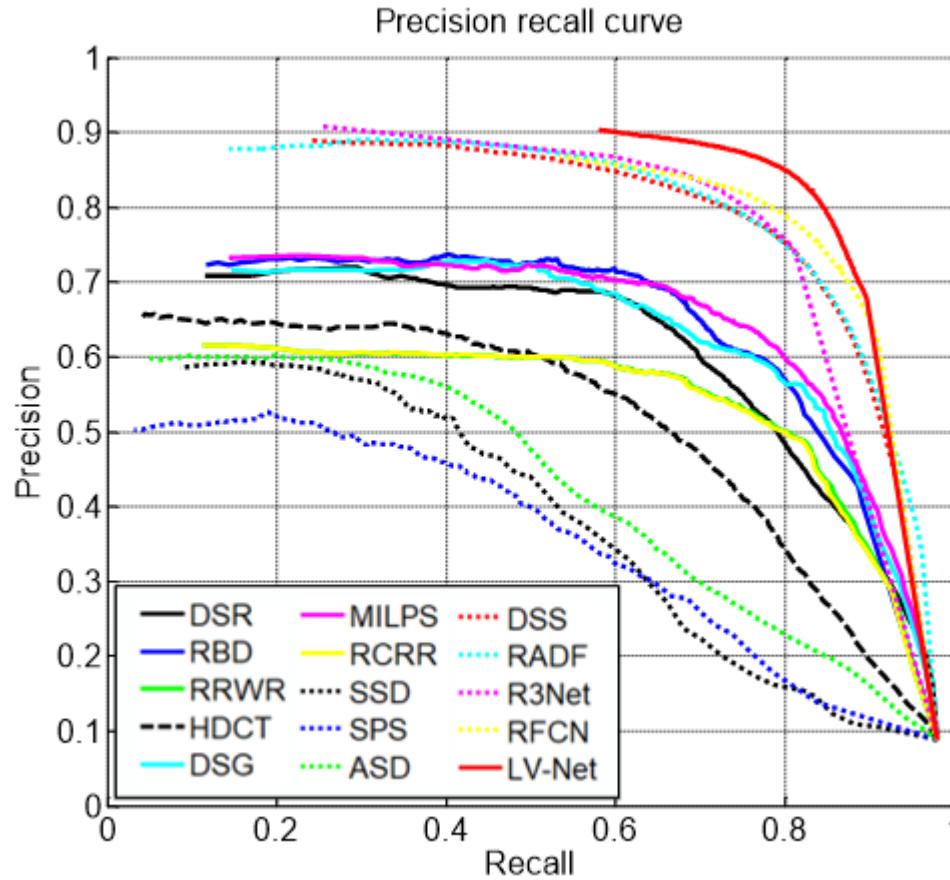
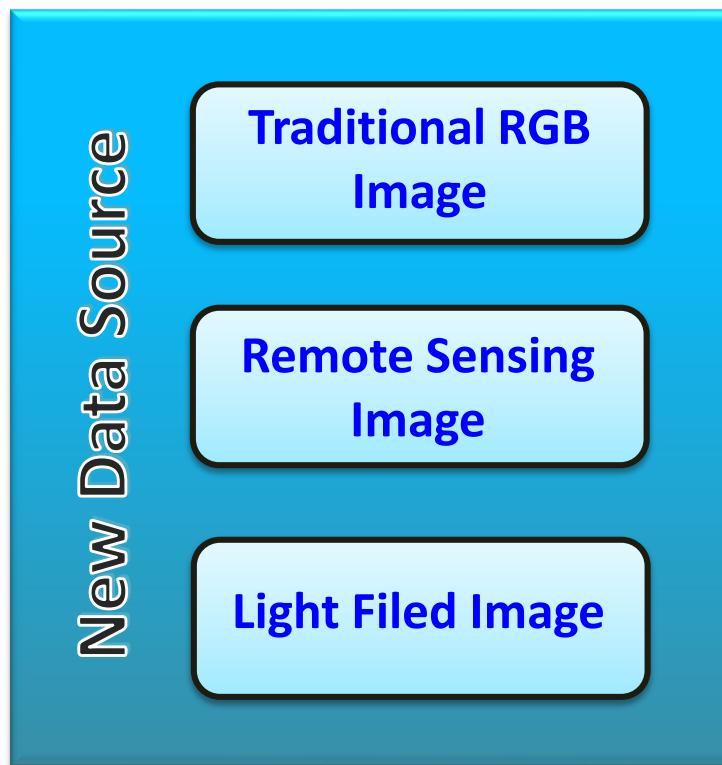


TABLE II
QUANTITATIVE COMPARISONS WITH DIFFERENT METHODS ON THE
TESTING SUBSET OF ORSSD DATASET.

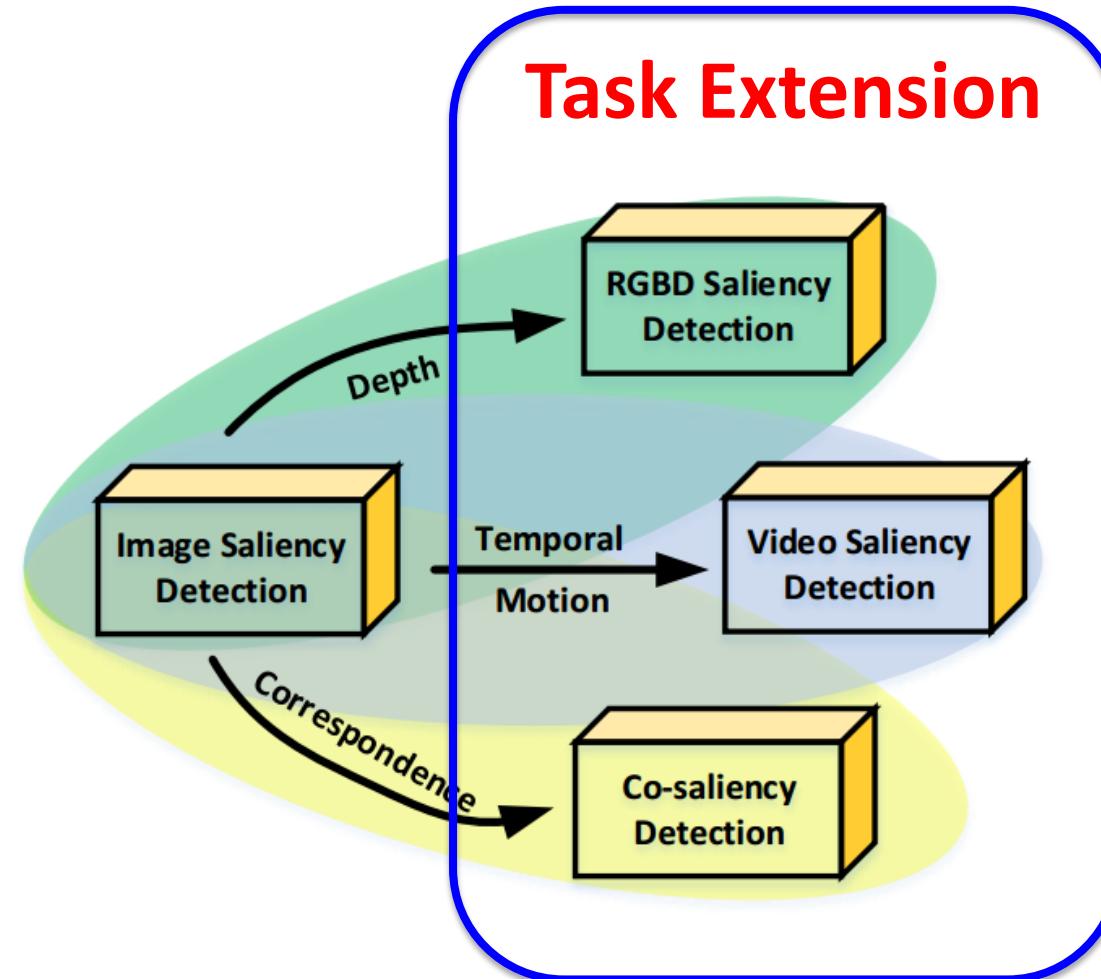
| Method | Precision | Recall | F_β | MAE | S_m |
|------------|---------------|---------------|---------------|---------------|---------------|
| DSR [20] | 0.6829 | 0.5972 | 0.6610 | 0.0859 | 0.7082 |
| RBD [18] | 0.7080 | 0.6268 | 0.6874 | 0.0626 | 0.7662 |
| RRWR [48] | 0.5782 | 0.6591 | 0.5950 | 0.1324 | 0.6835 |
| HDCT [49] | 0.6071 | 0.4969 | 0.5775 | 0.1309 | 0.6197 |
| DSG [50] | 0.6843 | 0.6007 | 0.6630 | 0.1041 | 0.7195 |
| MILPS [51] | 0.6954 | 0.6549 | 0.6856 | 0.0913 | 0.7361 |
| RCRR [15] | 0.5782 | 0.6552 | 0.5944 | 0.1277 | 0.6849 |
| SSD [29] | 0.5188 | 0.4066 | 0.4878 | 0.1126 | 0.5838 |
| SPS [31] | 0.4539 | 0.4154 | 0.4444 | 0.1232 | 0.5758 |
| ASD [33] | 0.5582 | 0.4049 | 0.5133 | 0.2119 | 0.5477 |
| DSS [24] | 0.8125 | 0.7014 | 0.7838 | 0.0363 | 0.8262 |
| RADF [25] | 0.8311 | 0.6724 | 0.7881 | 0.0382 | 0.8259 |
| R3Net [16] | 0.8386 | 0.6932 | 0.7998 | 0.0399 | 0.8141 |
| RFCN [28] | 0.8239 | 0.7376 | 0.8023 | 0.0293 | 0.8437 |
| LV-Net | 0.8672 | 0.7653 | 0.8414 | 0.0207 | 0.8815 |

Conclusion

Data Extension



Task Extension



Future work

1. **New attempts in learning based saliency detection methods, such as small samples training, weakly supervised learning, and cross-domain learning.**

Limited by the labelled training data, more work, such as designing a special network, can be explored in the future to achieve high-precision detection with small training samples. In addition, weakly supervised salient object detection method is a good choice to address the insufficient pixel-level saliency annotations. Furthermore, the cross-domain learning is another direction that needs to be addressed for learning based RGBD saliency detection method.

Future work

2. **Extending the saliency detection task in different data sources, such as light filed image, RGBD video, and remote sensing image.** In the light filed image, the focusness prior, multi-view information, and depth cue should be considered jointly. For the RGBD video data, the depth constraint should be introduced to assist in the spatiotemporal saliency. In the remote sensing image, due to the high angle shot photographed, some small targets and shadows are included. Thus, how to suppress the interference effectively and highlight the salient object accurately should be further investigated in the future.

Publications

- **Runmin Cong**, Jianjun Lei, Huazhu Fu, Weisi Lin, Qingming Huang, Xiaochun Cao, Chunping Hou, “An iterative co-saliency framework for RGBD images,” IEEE Transactions on Cybernetics, vol. 49, no. 1, pp. 233-246, 2019. (SCI, IF=10.387)
- **Runmin Cong**, Jianjun Lei, Huazhu Fu, Qingming Huang, Xiaochun Cao, Chunping Hou, “Co-saliency detection for RGBD images based on multi-constraint feature matching and cross label propagation,” IEEE Transactions on Image Processing, vol. 27, no. 2, pp. 568-579, 2018. (SCI, IF=6.790)
- **Runmin Cong**, Jianjun Lei, Huazhu Fu, Fatih Porikli, Qingming Huang, Chunping Hou, “Video saliency detection via sparsity-based reconstruction and propagation,” IEEE Transactions on Image Processing, DOI: 10.1109/TIP.2019.2910377, 2019. (SCI, IF=6.790)
- **Runmin Cong**, Jianjun Lei, Huazhu Fu, Qingming Huang, Xiaochun Cao, Nam Ling, “HSCS: Hierarchical sparsity based co-saliency detection for RGBD images,” IEEE Transactions on Multimedia, vol. 21, no. 7, pp. 1660-1671, 2019. (SCI, IF=5.452)
- **Runmin Cong**, Jianjun Lei, Huazhu Fu, Ming-Ming Cheng, Weisi Lin, Qingming Huang, “Review of visual saliency detection with comprehensive information,” IEEE Transactions on Circuits and Systems for Video Technology, DOI: 10.1109/TCSVT.2018.2870832, 2019. (SCI, IF=4.046)

Publications

- **Runmin Cong**, Jianjun Lei, Huazhu Fu, Junhui Hou, Qingming Huang, Sam Kwong, “Going from RGB to RGBD saliency: A depth-guided transformation model,” IEEE Transactions on Cybernetics, 2019. (SCI, IF=10.387)
- **Runmin Cong**, Jianjun Lei, Changqing Zhang, Qingming Huang, Xiaochun Cao, Chunping Hou, “Saliency detection for stereoscopic images based on depth confidence analysis and multiple cues fusion,” IEEE Signal Processing Letters, vol. 23, no. 6, pp. 819-823, 2016. (SCI, IF=3.268)
- Chongyi Li, **Runmin Cong#**, Junhui Hou, Sanyi Zhang, Yue Qian, Sam Kwong, “Nested network with two-stream pyramid for salient object detection in optical remote sensing images,” IEEE Transaction on Geoscience and Remote Sensing, 2019. (# co-first and corresponding author, SCI, IF=5.630)
- **Runmin Cong**, Jianjun Lei, Huazhu Fu, Wenguan Wang, Qingming Huang, Lijie Niu, “Research progress of video saliency detection,” Journal of Software, vol. 29, no. 8, pp. 2527-2544, 2018. (EI, in Chinese)
- Yonghua Zhang, Liang Li, **Runmin Cong**, Xiaojie Guo, Hui Xu, Jiawan Zhang, “Co-saliency detection via hierarchical consistency measure,” IEEE ICME, pp. 1-6, 2018. (Best Student Paper Runner-up, CCF B)
- **Runmin Cong**, Hao Chen, Hongyuan Zhu, Huazhu Fu, “Foreground detection and segmentation in RGB-D images,” in Paul Rosin, Yukun Lai, Yonghuai Liu, Ling Shao, RGB-D Image Analysis and Processing, Springer, 2018. (Book Chapter)



Code and Result: <https://rmcong.github.io/>