全场景视觉显著性计算

FULL-SCENE VISUAL SALIENCY COMPUTING

报告人：丛润民

山东大学控制科学与工程学院

机器智能与系统控制教育部重点实验室

2023-09-22

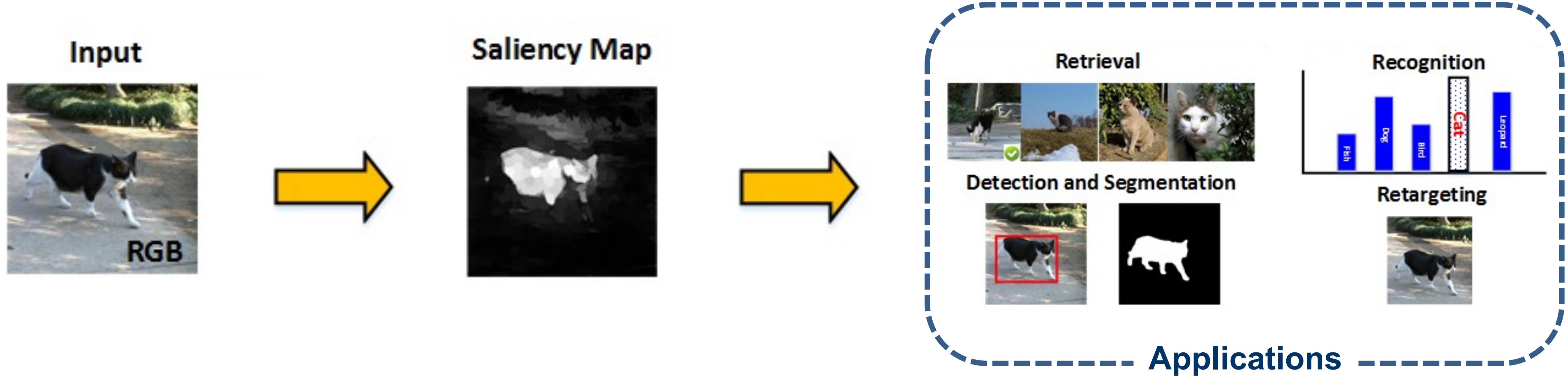
ICIG 2023
NANJING, CHINA | SEPTEMBER 22 - 24, 2023

山东大学
SHANDONG UNIVERSITY

Outline

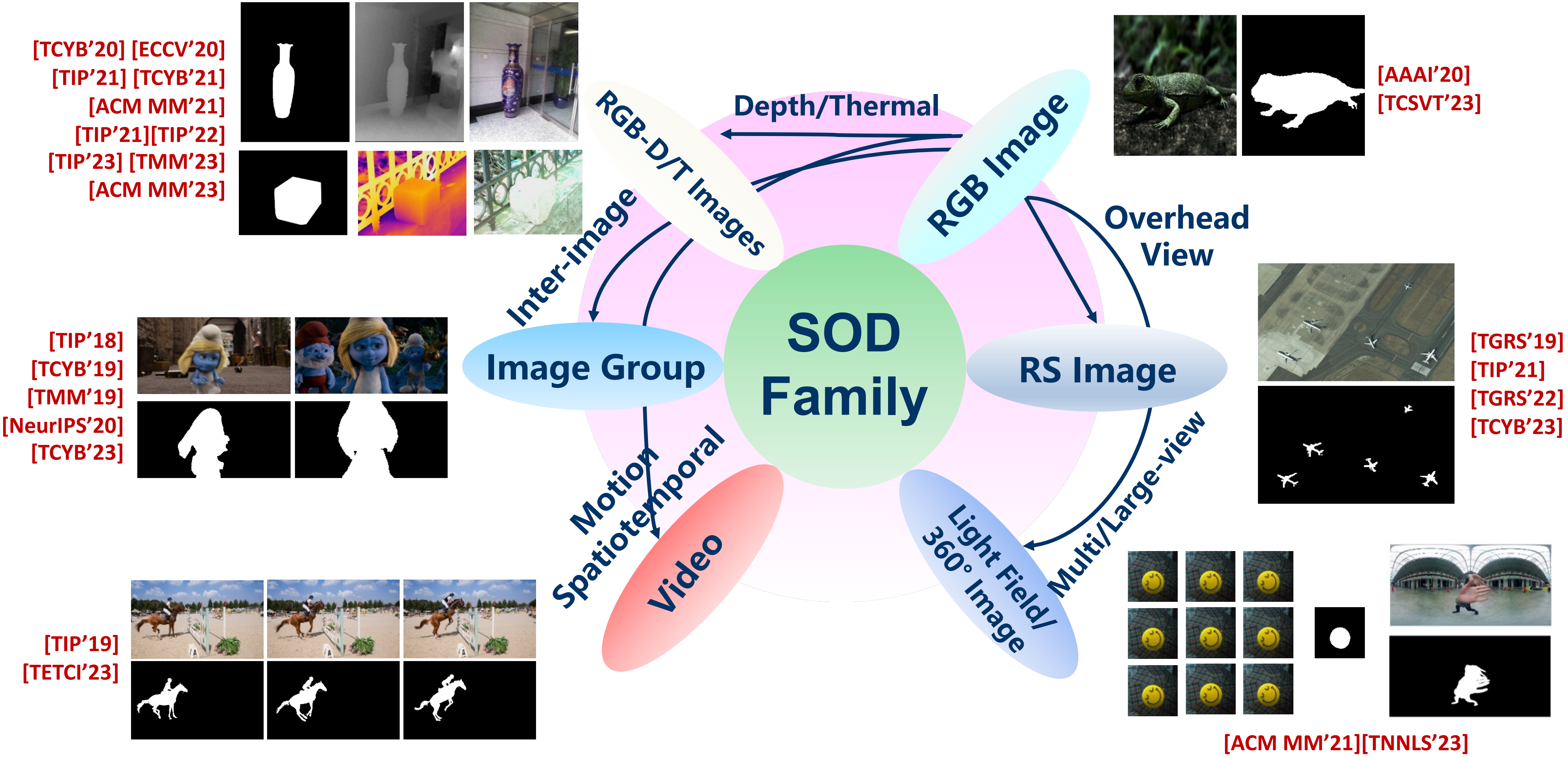
- Introduction
- Technical Methods
  - SOD for Single-modality Data (TCSVT'23)
  - SOD for Cross-modality Data (TIP'21)
  - SOD for Optical Remote Sensing Data (TGRS'22)
- Future Work

Introduction



Simulating the human visual attention mechanism, salient object detection aims at 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.

Introduction



单模态视觉显著性计算

SOD FOR SINGLE-MODALITY DATA

A Weakly Supervised Learning Framework for Salient Object Detection via Hybrid Labels

Runmin Cong, Qi Qin, Chen Zhang, Qiuping Jiang,
Shiqi Wang, Yao Zhao and Sam Kwong

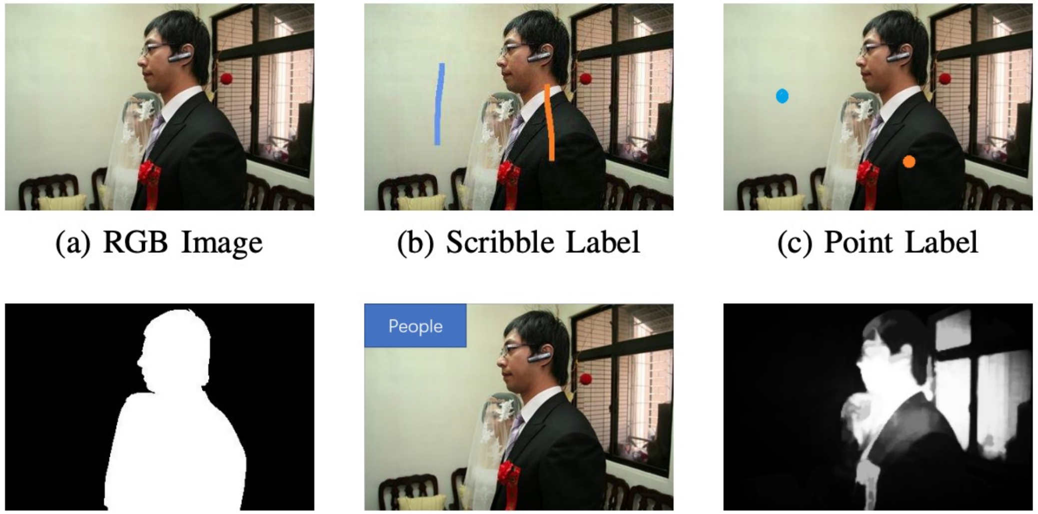
IEEE Transactions on Circuits and Systems for Video Technology, 2023

https://rmcong.github.io/proj_Hybrid-Label-SOD.html

Motivation

Full Supervised

Pixel Wise Annotation



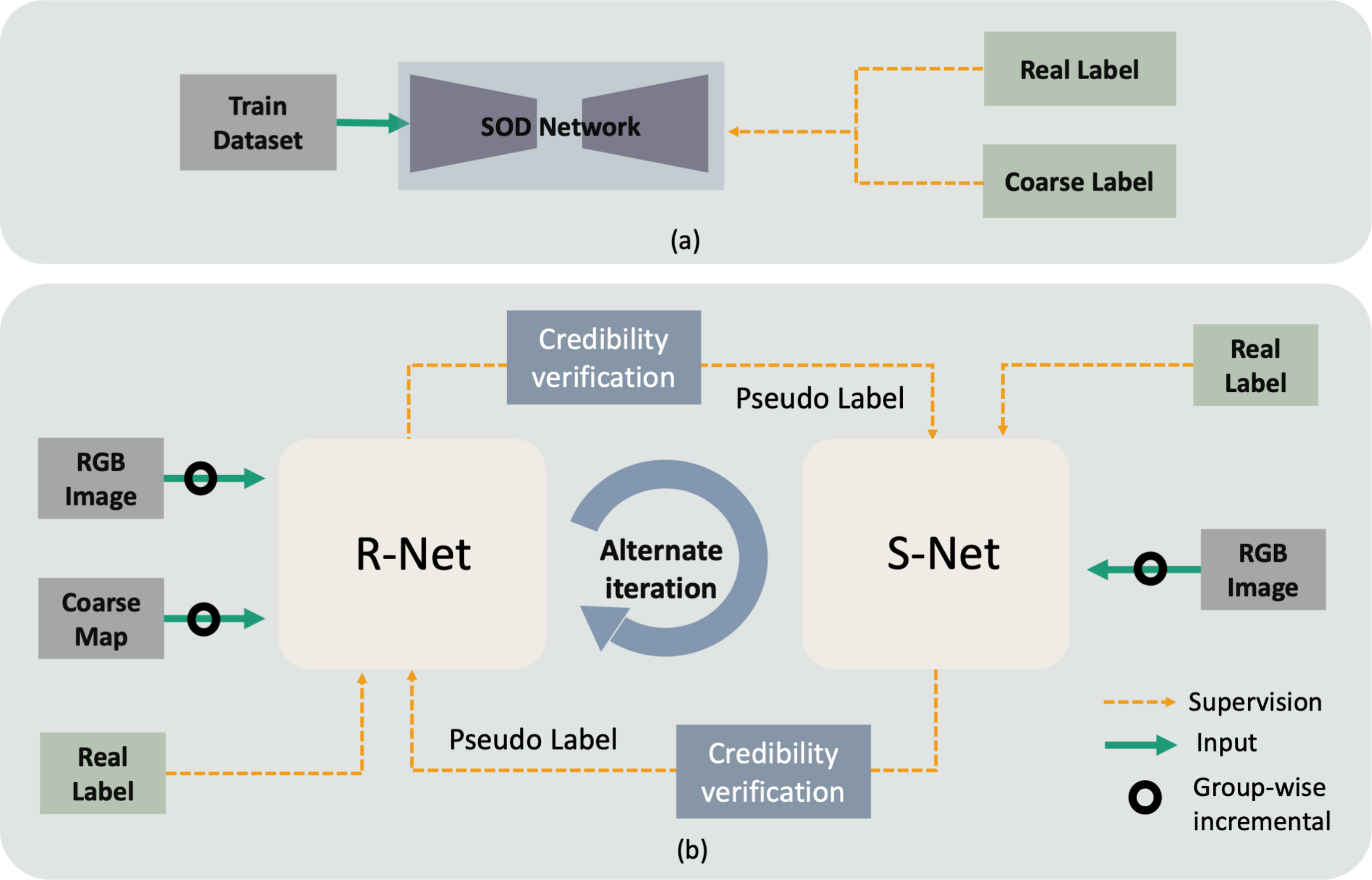
(a) RGB Image (b) Scribble Label (c) Point Label

(d) Pixel-level Label (e) Image-level Label (f) Coarse Label

According to the given labeled data, weakly supervised/unsupervised SOD methods can be roughly divided into the following categories:

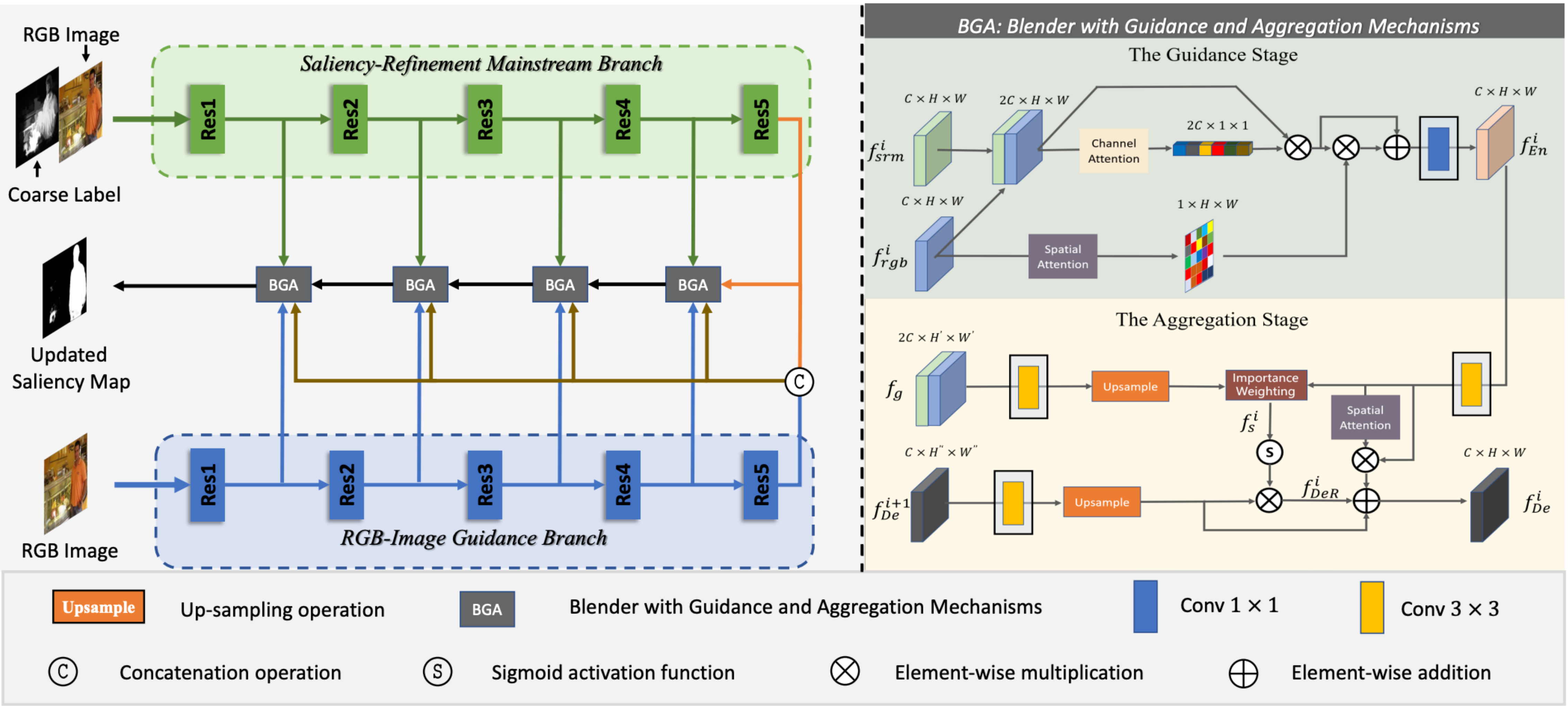
- ✓ scribble label supervision
- ✓ point label supervision
- ✓ image-level label supervision
- ✓ coarse label supervision

Our Method

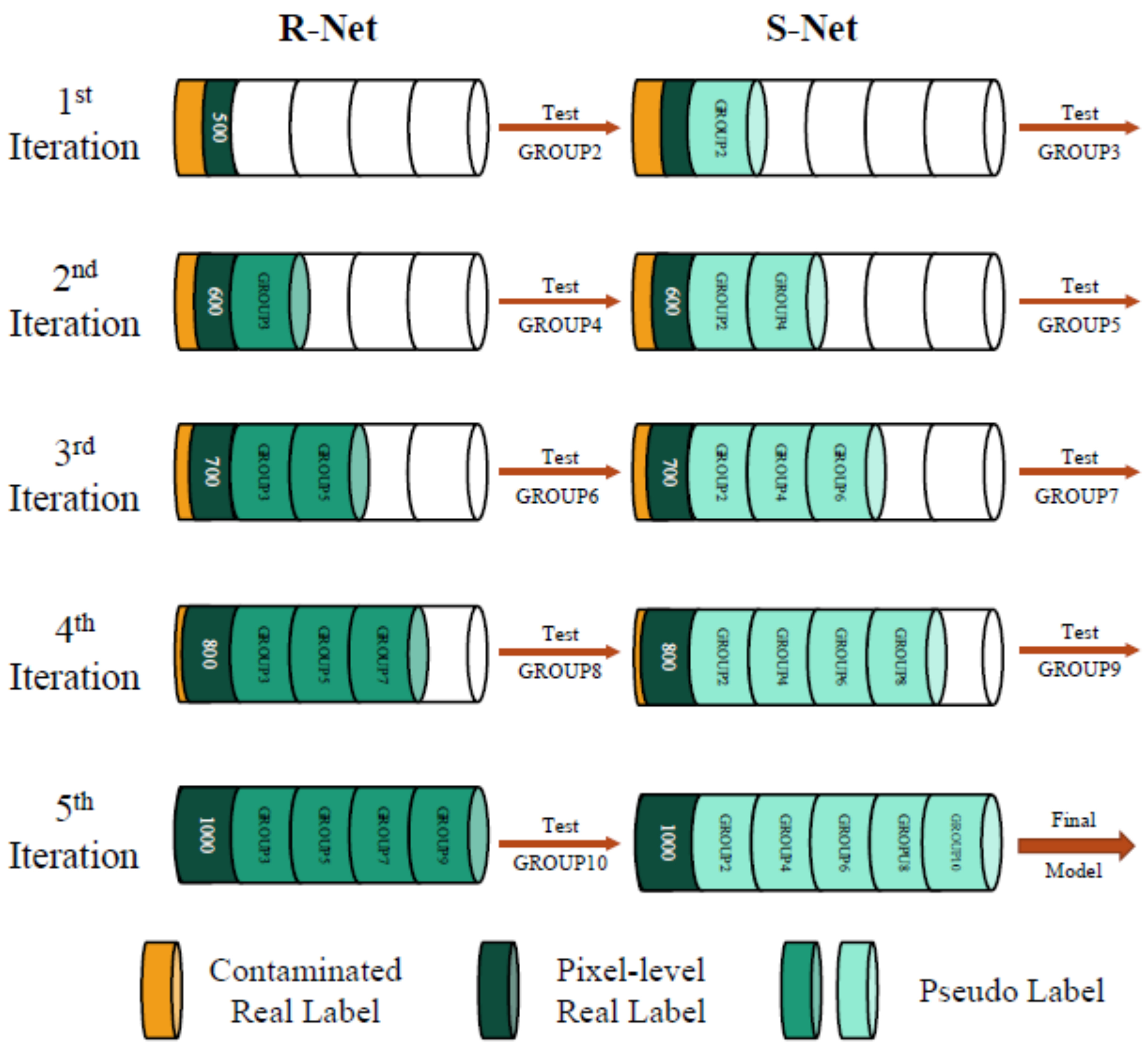


(a) A **simple solution** for training the SOD model with coarse and real labels. (b) The proposed **alternate learning framework** for weakly-supervised SOD task under the hybrid label, consisting of a **Refine Network (R-Net)** and a **Saliency Network (S-Net).** These two networks cooperate with each other and train alternately.

Refinement Network (R-Net)



Training Strategy



Alternate iteration mechanism

Group-wise incremental mechanism

Credibility verification mechanism

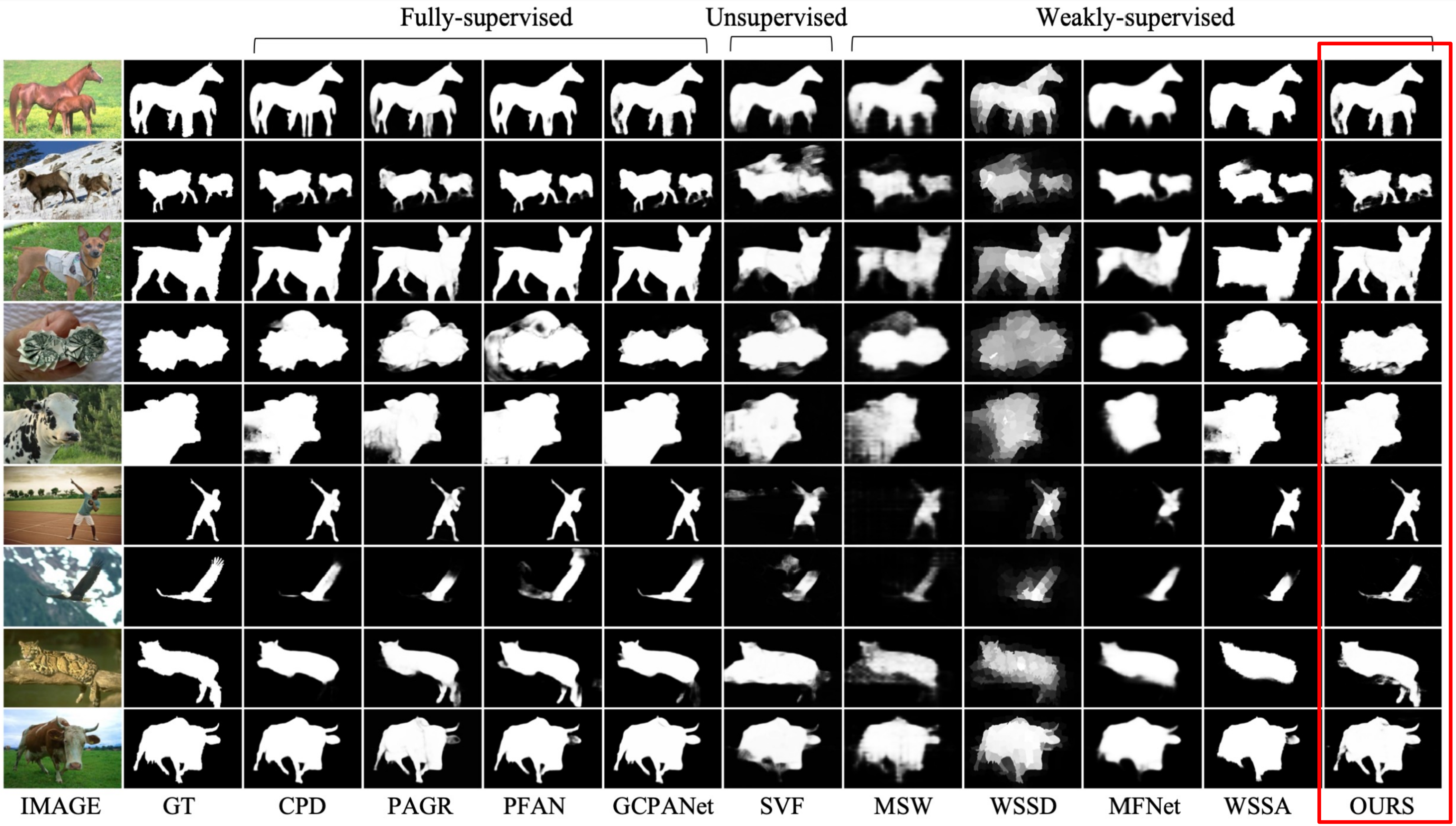
Experiments

- We implement the proposed model via PyTorch toolbox and train it on an RTX 3090 GPU in my training strategy .
- Most of the previous state-of-the-art SOD models are trained on the large-scale **DUTS** dataset, which contains **10,553** training images (DUTS-TR) and **5,019** testing images (DUTS-TE).
- We use the **MB method** to generate the saliency maps (Coarse Label) for all the images in the DUTS-TR dataset.
- We select the first **1,000** samples in the DUTS-TR dataset as the **real-labeled** data, providing the **pixel-wise** real ground truth.

Experiments

	SUP	YEAR	DUTS-TE			ECSSD			HKU-IS			PASCAL-S			THUR		
			$F_\beta^{max}\uparrow$	$S_m\uparrow$	$MAE\downarrow$	$F_\beta^{max}\uparrow$	$S_m\uparrow$	$MAE\downarrow$	$F_\beta^{max}\uparrow$	$S_m\uparrow$	$MAE\downarrow$	$F_\beta^{max}\uparrow$	$S_m\uparrow$	$MAE\downarrow$	$F_\beta^{max}\uparrow$	$S_m\uparrow$	$MAE\downarrow$
DGRL	F	2018	0.805	0.842	0.050	0.913	0.903	0.041	0.900	0.894	0.036	0.837	0.836	0.072	0.746	0.813	0.076
PiCANet	F	2018	0.840	0.863	0.040	0.928	0.916	**0.035**	0.913	0.905	0.031	0.848	0.846	0.065	-	-	-
PAGR	F	2018	0.816	0.838	0.056	0.904	0.889	0.061	0.897	0.887	0.048	0.822	0.819	0.092	0.769	0.830	0.070
MLMSNet	F	2019	0.825	0.861	0.049	0.917	0.911	0.045	0.910	0.906	0.039	0.841	0.845	0.074	0.752	0.819	0.079
CPD	F	2019	0.840	0.869	0.043	0.926	0.918	0.037	0.911	0.905	0.034	0.842	0.847	0.072	0.774	0.834	0.068
AFNet	F	2019	0.836	0.867	0.046	0.924	0.913	0.042	0.909	0.905	0.036	0.848	0.849	0.071	-	-	-
BASNet	F	2019	0.838	0.866	0.048	0.931	0.916	0.037	0.919	0.909	0.032	0.842	0.836	0.077	-	-	-
PFAN	F	2019	0.850	0.874	0.041	0.914	0.904	0.045	0.918	0.914	0.032	**0.866**	0.862	0.065	0.722	0.781	0.104
GCPANet	F	2020	**0.866**	**0.891**	**0.038**	0.936	**0.927**	**0.035**	**0.926**	**0.920**	0.031	0.859	**0.866**	**0.062**	**0.784**	**0.840**	0.070
MINet	F	2020	0.863	0.881	0.039	**0.937**	0.923	0.036	0.922	0.914	**0.030**	0.856	0.855	**0.062**	0.778	0.836	**0.066**
SVF	Un	2017	-	-	-	0.832	0.832	0.091	-	-	-	0.734	0.757	0.134	-	-	-
MNL	Un	2018	0.725	-	0.075	0.810	-	0.091	0.820	-	0.065	0.747	-	0.157	-	-	-
WSS	I	2017	0.633	-	0.100	0.767	-	0.108	0.773	-	0.078	0.697	-	0.184	-	-	-
ASMO	I	2018	0.568	-	0.115	0.762	-	0.068	0.762	-	0.088	0.653	-	0.205	-	-	-
MSW	M	2019	0.705	0.752	0.091	0.851	0.820	0.099	0.828	0.812	0.086	0.759	0.762	0.136	-	-	-
MFNet	I	2021	0.733	0.775	0.076	0.858	0.835	0.084	0.859	0.847	0.058	0.764	0.768	0.117	0.731	0.795	0.075
WSSD	Sub	2021	-	-	-	0.873	0.827	0.119	0.884	0.870	0.082	0.820	0.814	0.128	0.703	0.768	0.114
WSSA	S	2020	0.755	0.803	0.062	0.871	0.865	0.059	0.864	0.865	0.047	0.788	0.796	0.094	0.736	0.800	0.077
Ours	H		**0.803**	**0.837**	**0.050**	**0.899**	**0.886**	**0.051**	**0.892**	**0.887**	**0.038**	**0.827**	**0.828**	**0.076**	**0.755**	**0.813**	**0.069**

Experiments



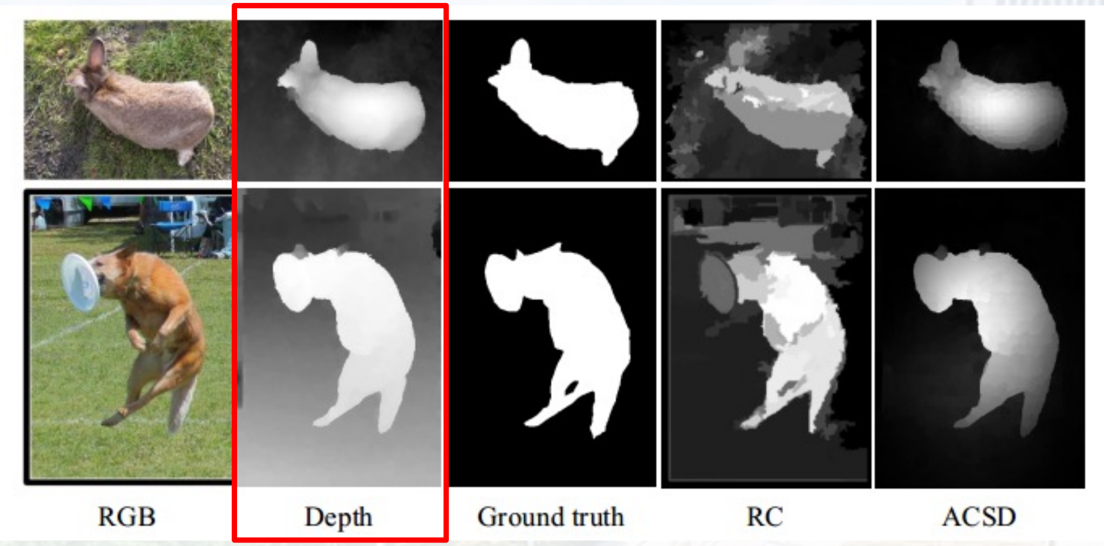
Contributions

- ✓ For the first time, we launch **a new weakly-supervised SOD task** based on **hybrid labels**, with a large number of coarse labels and a small number of real labels as supervision. To this end, we decouple this task into **two sub-tasks** of coarse label refinement and salient object detection, and design the corresponding R-Net and S-Net.
- ✓ We design a BGA in the R-Net to achieve **two-stage feature decoding**, where the **guidance stage** is used to introduce the guidance information from the RGB-image guidance branch to guarantee a relatively robust performance baseline, and the **aggregation stage** is to dynamically integrate different levels of features according to their modification or supplementation roles.
- ✓ In order to guarantee the effectiveness and efficiency of network training, from the perspective of quantity allocation, training method and reliability judgment, we design the **alternate iteration mechanism, group-wise incremental mechanism, and credibility verification mechanism**.

跨模态视觉显著性计算

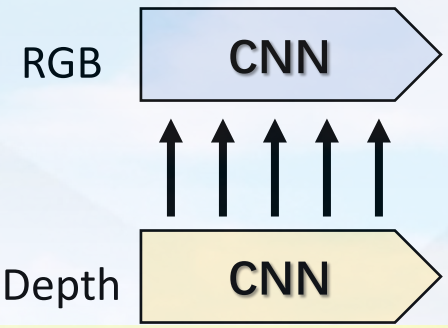
SOD FOR CROSS-MODALITY DATA

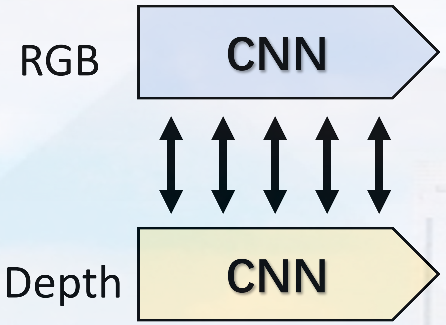
RGB-D Salient Object Detection

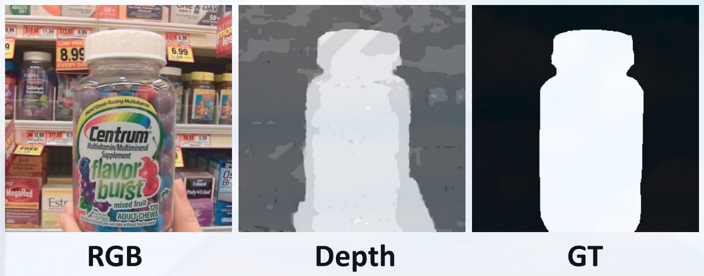


- shape
- contour
- internal consistency
- surface normal
- ……

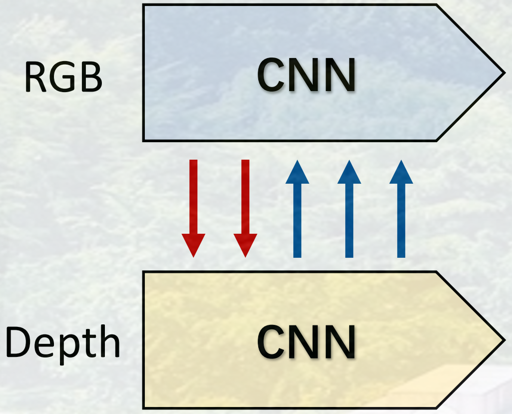
depth quality perception





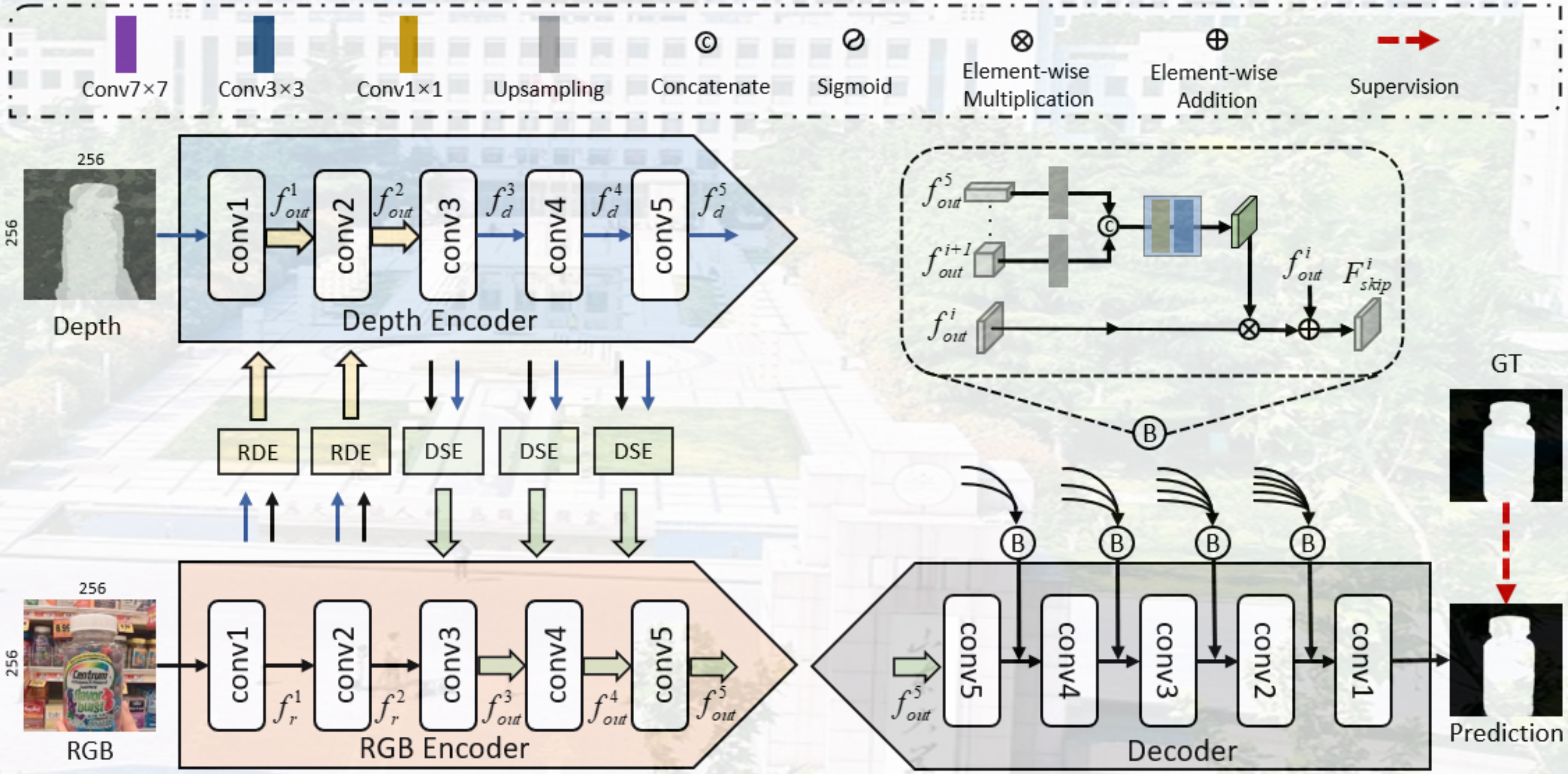


How can we fully exploit the strengths of both modalities and provide clear guidance?



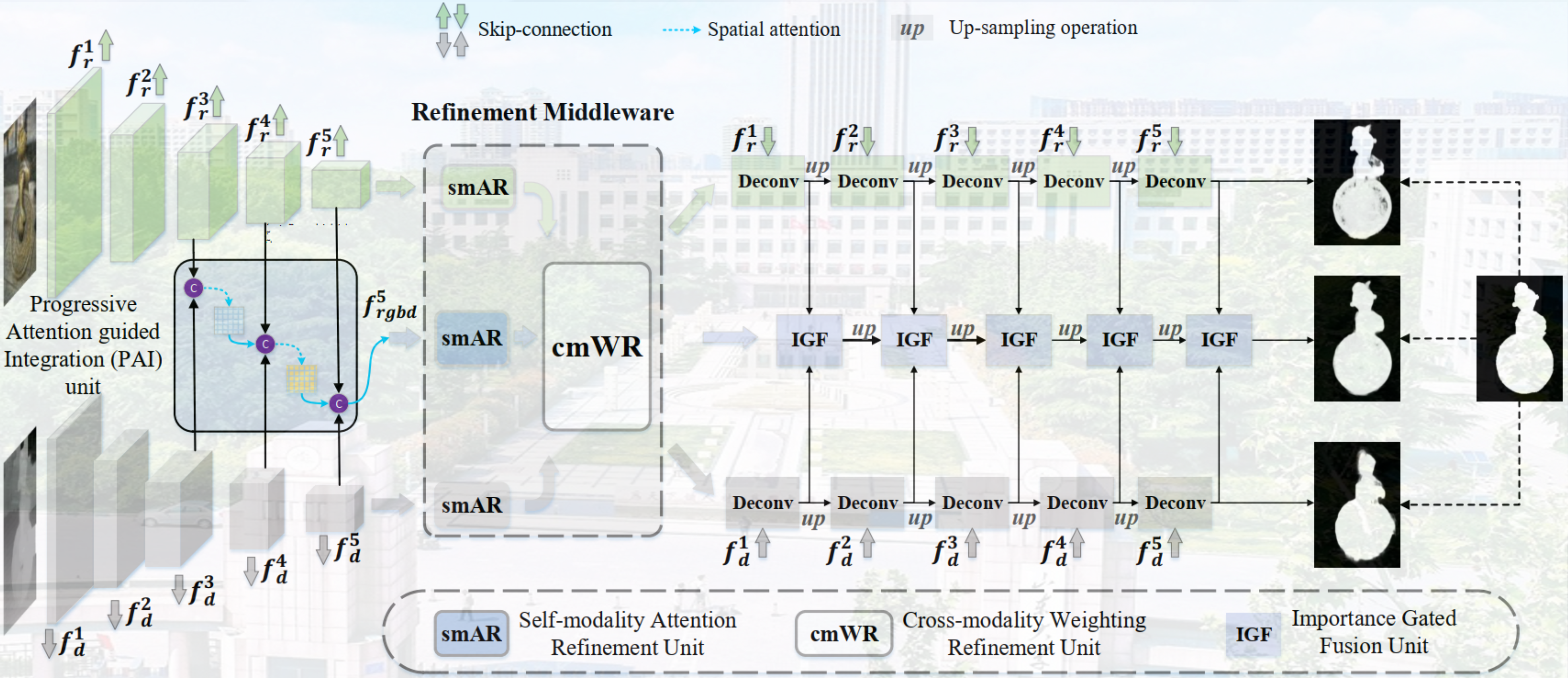
(c)

discrepant interaction mode



R. Cong, et al. Cross-Modality Discrepant Interaction Network for RGB-D Salient Object Detection, ACM MM 2021

cross-modality interaction and refinement mode under three-stream structure



R. Cong, et al. CIR-Net: Cross-modality Interaction and Refinement for RGB-D Salient Object Detection, *TIP 2022*

Highly Cited Paper

DPANet: Depth Potentiality-Aware Gated Attention Network for RGB-D Salient Object Detection

Zuyao Chen‡, Runmin Cong‡, Qianqian Xu, and Qingming Huang

IEEE Transactions on Image Processing, 2021

https://rmcong.github.io/proj_DPANet.html

Motivations

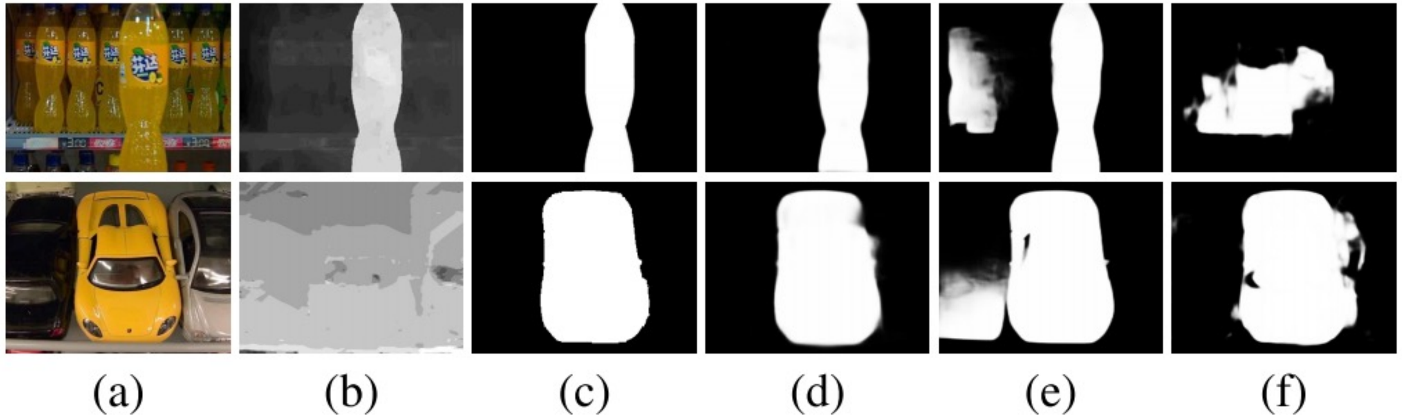
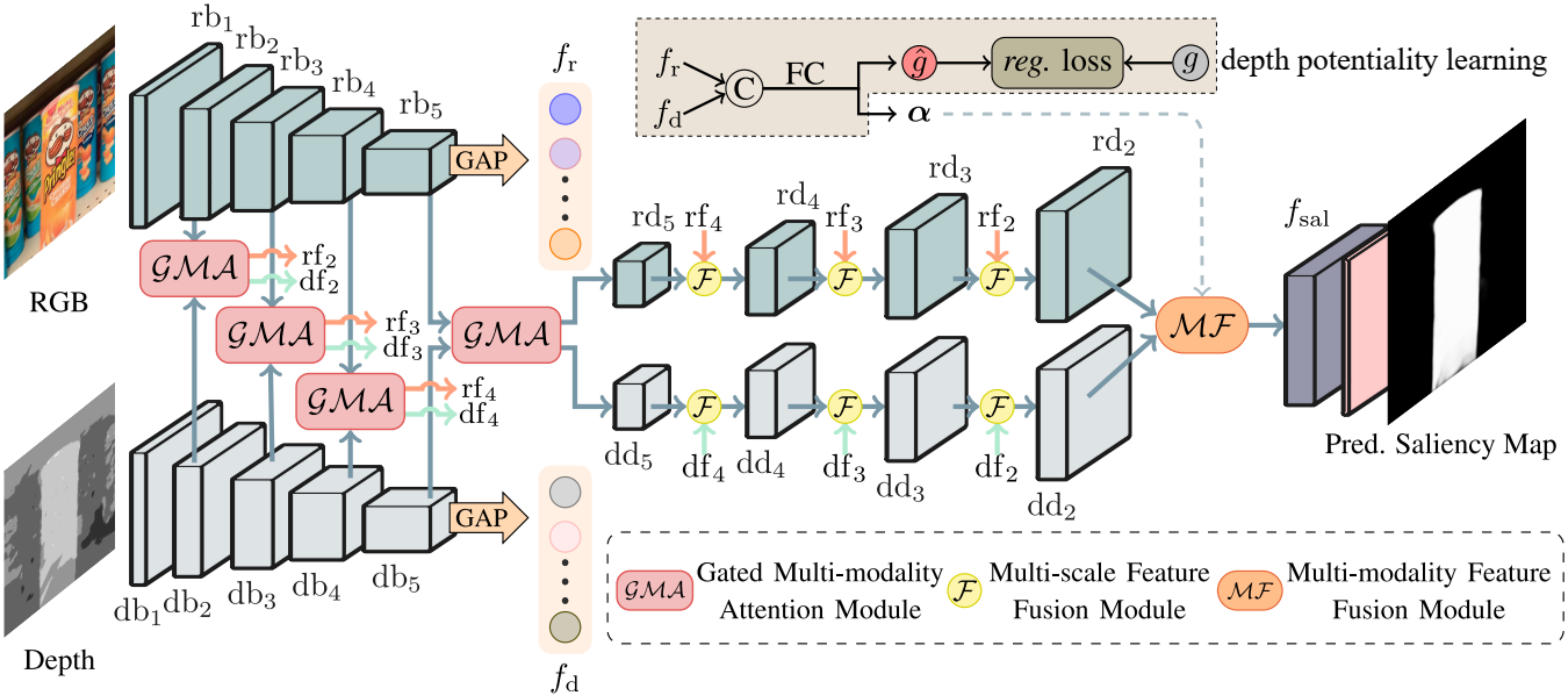


Fig. 1. Sample results of our method compared with others. RGB-D methods are marked in **boldface**. (a) RGB image; (b) Depth map; (c) Ground truth; (d) **Ours**; (e) BASNet [14]; (f) **CPFP** [33].

- how to effectively **integrate** the complementary information from RGB image and its corresponding depth map;
- how to **prevent** the contamination from unreliable depth information;

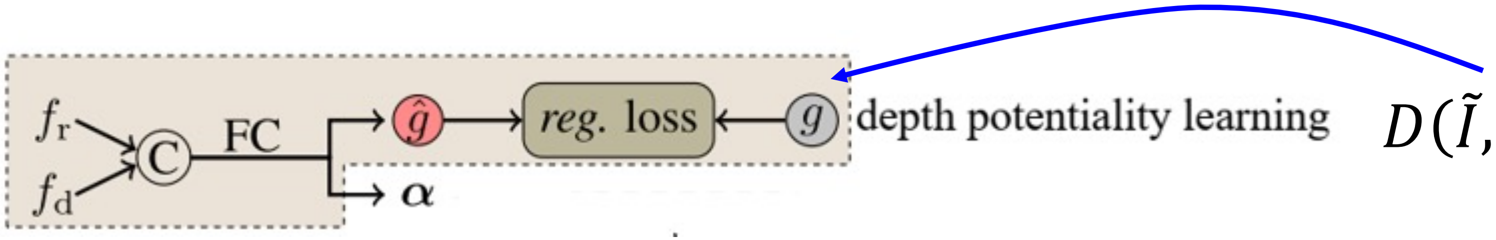
Our Method



RGB
Depth
rb1 rb2 rb3 rb4 rb5
db1 db2 db3 db4 db5
GAP
GAP
GMA
rf2 df2
rf3 df3
rf4 df4
fr
fd
FC
reg. loss
depth potentiality learning
α
rd5 rf4 rd4 rf3 rd3 rf2 rd2
dd5 df4 dd4 df3 dd3 df2 dd2
MF
fsal
Pred. Saliency Map
Gated Multi-modality Attention Module
Multi-scale Feature Fusion Module
Multi-modality Feature Fusion Module

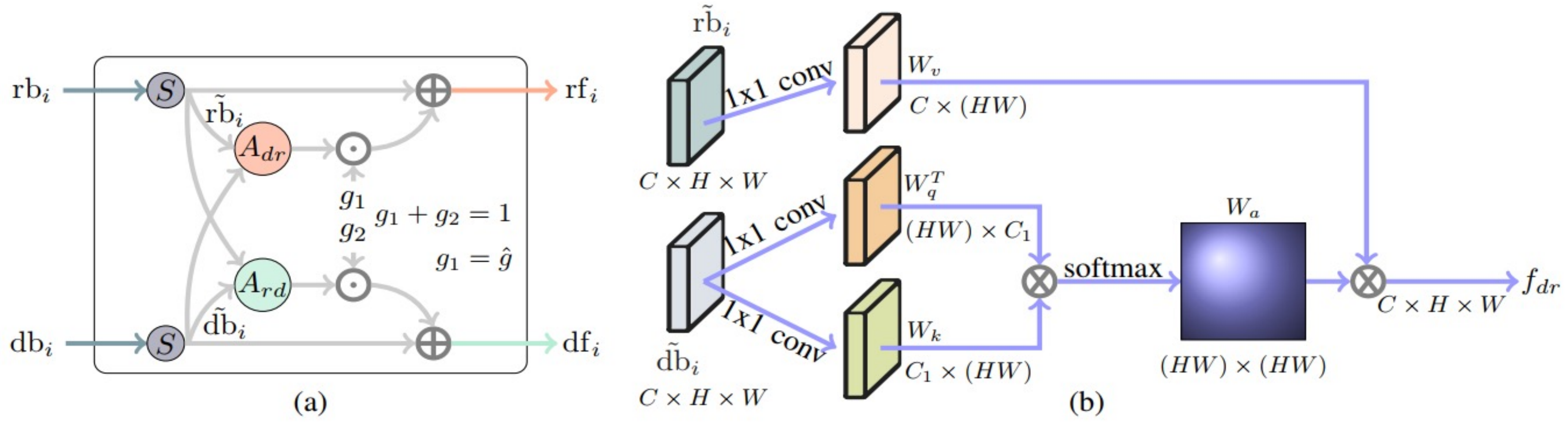
Depth Potentiality Perception

- Most previous works generally integrate the multi-modal features from RGB and corresponding depth information indiscriminately. However, **there exist some contaminations when depth maps are unreliable**.
- Since we do not hold any labels for depth map quality assessment, **we model the depth potentiality perception as a saliency-oriented prediction task**, that is, we train a model to automatically learn the relationship between the binary depth map and the corresponding saliency mask. The above modeling approach is based on the observation that **if the binary depth map segmented by a threshold is close to the ground truth, the depth map is highly reliable, so a higher confidence response should be assigned to this depth input**.



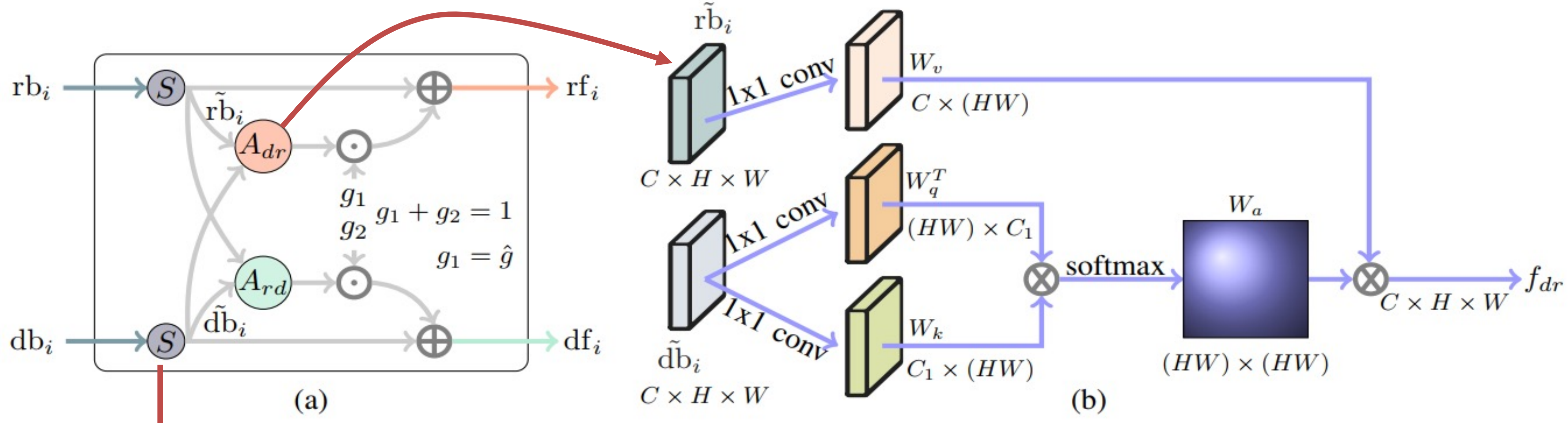
$$D(\tilde{I}, G) = \frac{(1+\gamma)\cdot D_{iou}\cdot D_{cov}}{D_{iou}+\gamma\cdot D_{cov}}$$

Gated Multi-modality Attention Module



- Directly integrating the cross-modal information may induce negative results, such as **contaminations from unreliable depth maps**. Besides, the features of the single modality usually are affluent in spatial or channel aspect with **information redundancy**.
- We design a GMA module that exploits the attention mechanism to **automatically select and strengthen important features** for saliency detection, and **incorporate the gate controller** into the GMA module to prevent the contamination from the unreliable depth map.

Gated Multi-modality Attention Module



single-modal perspective:

spatial attention

reduce the redundancy features and highlight the feature response on the salient regions

cross-modal perspective:

two symmetrical attention sub-modules

capture long-range dependencies

$$\mathrm{rf}_i = \widetilde{\mathrm{rb}}_i + g_1 \cdot f_{dr} \qquad g_1 = \hat{g}$$

$$\mathrm{df}_i = \widetilde{\mathrm{db}}_i + g_2 \cdot f_{rd} \qquad g_1 + g_2 = 1$$

Multi-level Feature Fusion

- Multi-scale Feature Fusion

Low-level features can provide more detail information, such as boundary, texture, and spatial structure, but may be sensitive to the background noises. Contrarily, high-level features contain more semantic information, which is helpful to locate the salient object and suppress the noises. Thus, we adopt a more aggressive yet effective operation, i.e., multiplication.

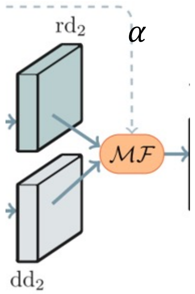
$$f_1 = \delta(up(conv_3(\mathrm{rd}_5)) \odot \mathrm{rf}_4)$$

$$f_2 = \delta(conv_4(\mathrm{rf}_4) \odot up(\mathrm{rd}_5))$$

$$f_F = \delta(conv_5([f_1, f_2]))$$

- Multi-modality Feature Fusion

During the multi-modality feature fusion, we consider two issues: (1) How to select the most useful and complementary information from the RGB and depth features. (2) How to prevent the contamination caused by the unreliable depth map during fusing.



$$f_3 = \boldsymbol{\alpha} \odot \mathrm{rd}_2 + \hat{g} \cdot (1 - \boldsymbol{\alpha}) \odot \mathrm{dd}_2$$

$$f_4 = \mathrm{rd}_2 \odot \mathrm{dd}_2$$

$$f_{sal} = \delta(conv([f_3, f_4]))$$

$\boldsymbol{\alpha}$ is the weight vector learned from RGB and depth information, $\hat{g}$ is the learned weight of the gate as mentioned before.

Loss Function

The final loss is the linear combination of the classification loss and regression loss:

$$\mathcal{L}_{final} = \mathcal{L}_{cls} + \lambda \cdot \mathcal{L}_{reg}$$

classification loss:

$$\mathcal{L}_{cls} = \mathcal{L}_{cls} + \sum_{i=1}^{8} \lambda_i \cdot \mathcal{L}_{aux}^{i}$$

regression loss :

$$\mathcal{L}_{reg} = \begin{cases} 0.5(g - \hat{g})^2, & if\ |g - \hat{g}| < 1 \\ |g - \hat{g}| - 0.5, & otherwise \end{cases}$$

Experiments

Method	RGBD135 Dataset			SSD Dataset			LFSD Dataset			NJUD-test Dataset		
	$F_\beta \uparrow$	$S_m \uparrow$	MAE $\downarrow$	$F_\beta \uparrow$	$S_m \uparrow$	MAE $\downarrow$	$F_\beta \uparrow$	$S_m \uparrow$	MAE $\downarrow$	$F_\beta \uparrow$	$S_m \uparrow$	MAE $\downarrow$
DPANet (ours)	**0.933**	**0.922**	**0.023**	**0.895**	**0.877**	**0.046**	**0.880**	**0.862**	**0.074**	**0.931**	**0.922**	**0.035**
AF-Net (Arxiv19)	0.904	0.892	0.033	0.828	0.815	0.077	0.857	0.818	0.091	0.900	0.883	0.053
DMRA (ICCV19)	0.921	0.911	0.026	0.874	0.857	0.055	0.865	0.831	0.084	0.900	0.880	0.052
CPFP (CVPR19)	0.882	0.872	0.038	0.801	0.807	0.082	0.850	0.828	0.088	0.799	0.798	0.079
PCFN (CVPR18)	0.842	0.843	0.050	0.845	0.843	0.063	0.829	0.800	0.112	0.887	0.877	0.059
PDNet (ICME19)	0.906	0.896	0.041	0.844	0.841	0.089	0.865	0.846	0.107	0.912	0.897	0.060
TAN (TIP19)	0.853	0.858	0.046	0.835	0.839	0.063	0.827	0.801	0.111	0.888	0.878	0.060
MMCI (PR19)	0.839	0.848	0.065	0.823	0.813	0.082	0.813	0.787	0.132	0.868	0.859	0.079
CTMF (TC18)	0.865	0.863	0.055	0.755	0.776	0.100	0.815	0.796	0.120	0.857	0.849	0.085
RS (ICCV17)	0.841	0.824	0.053	0.783	0.750	0.107	0.795	0.759	0.130	0.796	0.741	0.120
EGNet (ICCV19)	0.913	0.892	0.033	0.704	0.707	0.135	0.845	0.838	0.087	0.867	0.856	0.070
BASNet (CVPR19)	0.916	0.894	0.030	0.842	0.851	0.061	0.862	0.834	0.084	0.890	0.878	0.054
PoolNet (CVPR19)	0.907	0.885	0.035	0.764	0.749	0.110	0.847	0.830	0.095	0.874	0.860	0.068
AFNet (CVPR19)	0.897	0.878	0.035	0.847	0.859	0.058	0.841	0.817	0.094	0.890	0.880	0.055
PiCAR (CVPR18)	0.907	0.890	0.036	0.864	0.871	0.055	0.849	0.834	0.104	0.887	0.882	0.060
R^3Net (IJCAI18)	0.857	0.845	0.045	0.711	0.672	0.144	0.843	0.818	0.089	0.805	0.771	0.105

Method	NLPR-test Dataset			STEREO797 Dataset			SIP Dataset			DUT Dataset		
	$F_\beta \uparrow$	$S_m \uparrow$	MAE $\downarrow$	$F_\beta \uparrow$	$S_m \uparrow$	MAE $\downarrow$	$F_\beta \uparrow$	$S_m \uparrow$	MAE $\downarrow$	$F_\beta \uparrow$	$S_m \uparrow$	MAE $\downarrow$
DPANet (ours)	**0.924**	**0.927**	**0.025**	**0.919**	**0.915**	**0.039**	**0.906**	**0.883**	**0.052**	**0.918**	**0.904**	0.047
AF-Net (Arxiv19)	0.904	0.903	0.032	0.905	0.893	0.047	0.870	0.844	0.071	0.862	0.831	0.077
DMRA (ICCV19)	0.887	0.889	0.034	0.895	0.874	0.052	0.883	0.850	0.063	0.913	0.880	0.052
CPFP (CVPR19)	0.888	0.888	0.036	0.815	0.803	0.082	0.870	0.850	0.064	0.771	0.760	0.102
PCFN (CVPR18)	0.864	0.874	0.044	0.884	0.880	0.061	–	–	–	0.809	0.801	0.100
PDNet (ICME19)	0.905	0.902	0.042	0.908	0.896	0.062	0.863	0.843	0.091	0.879	0.859	0.085
TAN (TIP19)	0.877	0.886	0.041	0.886	0.877	0.059	–	–	–	0.824	0.808	0.093
MMCI (PR19)	0.841	0.856	0.059	0.861	0.856	0.080	–	–	–	0.804	0.791	0.113
CTMF (TC18)	0.841	0.860	0.056	0.827	0.829	0.102	–	–	–	0.842	0.831	0.097
RS (ICCV17)	0.900	0.864	0.039	0.857	0.804	0.088	–	–	–	0.807	0.797	0.111
EGNet (ICCV19)	0.845	0.863	0.050	0.872	0.853	0.067	0.846	0.825	0.083	0.888	0.867	0.064
BASNet (CVPR19)	0.882	0.894	0.035	0.914	0.900	0.041	0.894	0.872	0.055	0.912	0.902	**0.041**
PoolNet (CVPR19)	0.863	0.873	0.045	0.876	0.854	0.065	0.856	0.836	0.079	0.883	0.864	0.067
AFNet (CVPR19)	0.865	0.881	0.042	0.905	0.895	0.045	0.891	0.876	0.055	0.880	0.868	0.065
PiCAR (CVPR18)	0.872	0.882	0.048	0.906	0.903	0.051	0.890	0.878	0.060	0.903	0.892	0.062
R^3Net (IJCAI18)	0.832	0.846	0.049	0.811	0.754	0.107	0.641	0.624	0.158	0.841	0.812	0.079

Experiments

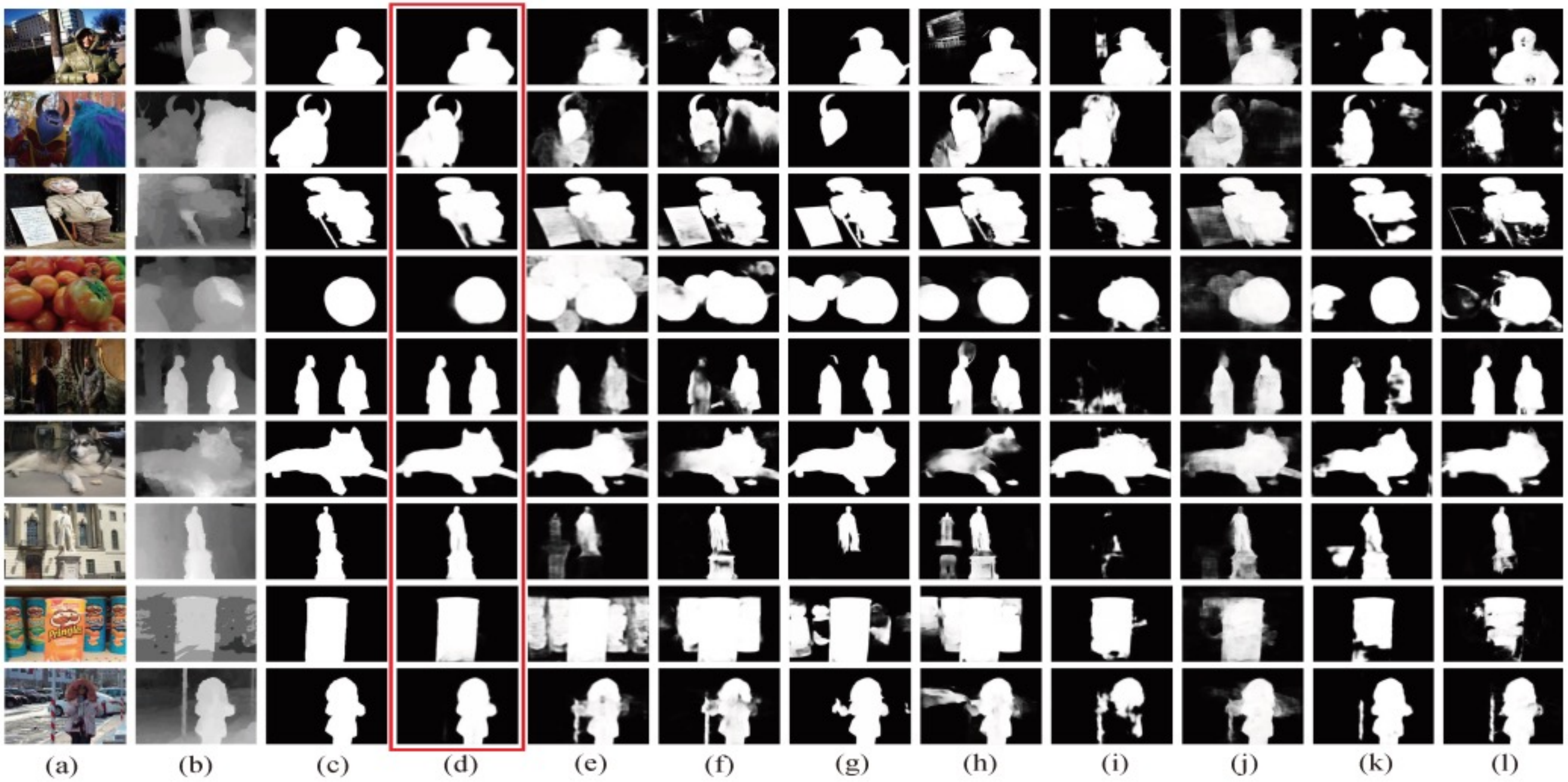


Fig. 4. Qualitative comparison of the proposed approach with some state-of-the-art RGB and RGB-D SOD methods, in which our results are highlighted by a red box. (a) RGB image. (b) Depth map. (c) GT. (d) DPANet. (e) PiCAR. (f) PoolNet. (g) BASNet. (h) EGNet. (i) CPFP. (j) PDNet. (k) DMRA. (l) AF-Net.

Contributions

a) **For the first time, we address the unreliable depth map in the RGB-D SOD network in an end-to-end formulation**, and propose the DPANet by incorporating the depth potentiality perception into the cross-modality integration pipeline.

b) **Without increasing the training label** (i.e., depth quality label), we model a **task-orientated depth potentiality perception module** that can adaptively perceive the potentiality of the input depth map, and further weaken the contamination from unreliable depth information.

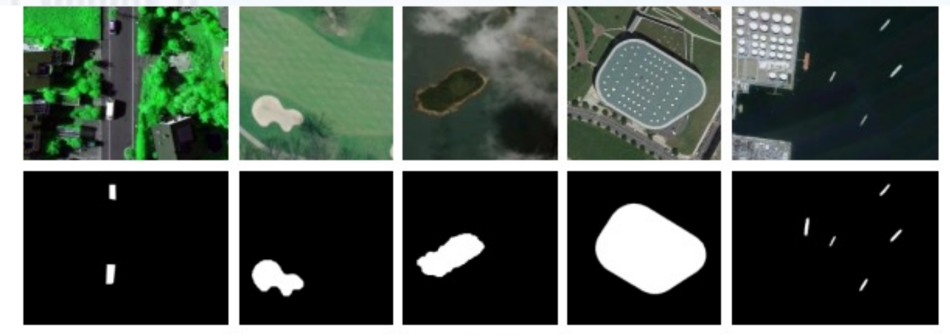
c) We propose a **gated multi-modality attention (GMA) module** to effectively aggregate the cross-modal complementarity of the RGB and depth images.

d) Without any pre-processing or post-processing techniques, the proposed network **outperforms 16 state-of-the-art methods on 8 RGB-D SOD datasets** in quantitative and qualitative evaluations.

遥感数据视觉显著性计算

SOD FOR OPTICAL REMOTE SENSING DATA

Salient Object Detection in Optical RSIs



Challenges

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.

Highly Cited Paper

山东省计算智能最佳论文奖

RRNet: Relational Reasoning Network with Parallel Multiscale Attention for Salient Object Detection in Optical Remote Sensing Images

Runmin Cong, Yumo Zhang, Leyuan Fang, Jun Li, Yao Zhao, and Sam Kwong

***IEEE Transactions on Geoscience and Remote Sensing*, 2022**

https://rmcong.github.io/proj_RRNet.html

Challenges

a) First, salient objects are often corrupted by **background interference and redundancy**.

b) Second, salient objects in RSIs present much more **complex structure and topology** than the ones in NSIs, which poses new **challenges in capturing complete object regions**.

c) Third, for the optical RSI SOD task, there is **only one dataset** (i.e., ORSSD [6]) available for model training and performance evaluation, which contains 800 images totally. This dataset is pioneering, but **its size is still relatively small**.

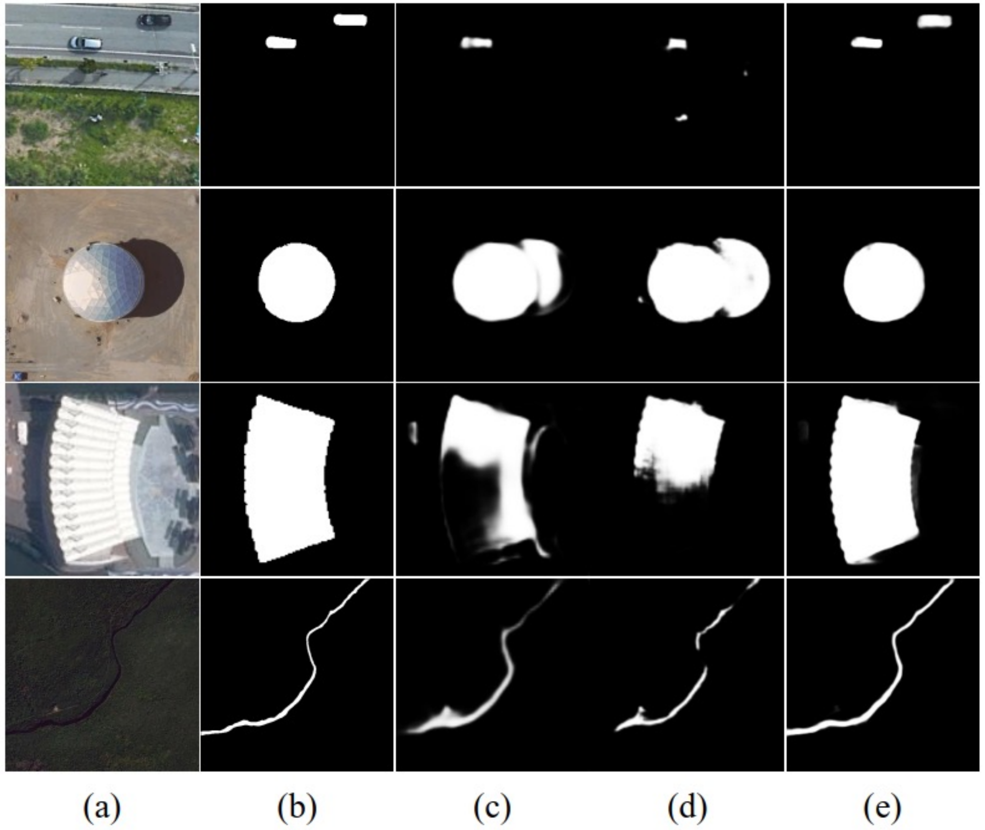
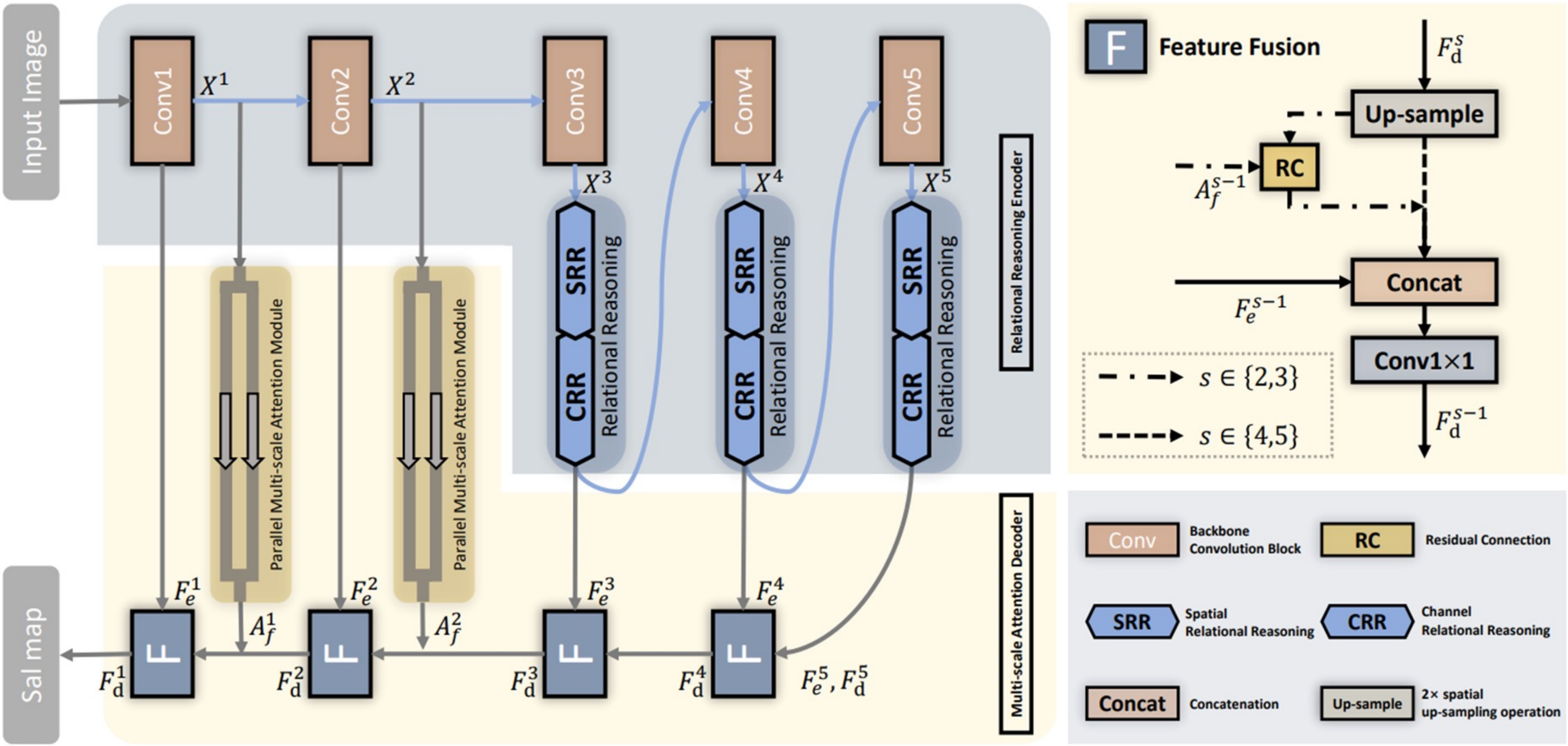


Fig. 1. Visual illustration of SOD results for optical RSIs by applying different methods. (a) Optical RSIs. (b) Ground truth. (c) PFAN [11]. (d) LVNet [6]. (e) Proposed DAFNet.

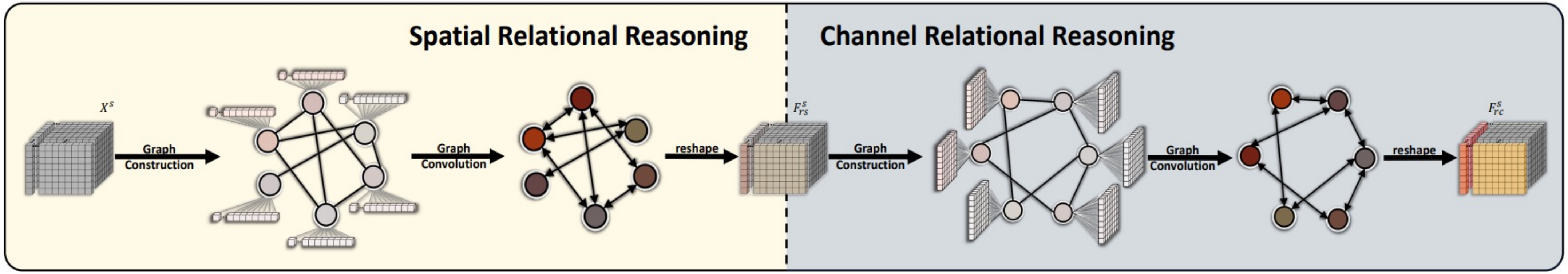
[6] C. Li, R. Cong, J. Hou, S. Zhang, Y. Qian, and S. Kwong, “Nested network with two-stream pyramid for salient object detection in optical remote sensing images,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 11, pp. 9156–9166, 2019

Our Method



Input Image
Conv1
Conv2
Conv3
Conv4
Conv5
X^1
X^2
X^3
X^4
X^5
SRR
CRR
Relational Reasoning
Relational Reasoning Encoder
Parallel Multi-scale Attention Module
Multi-scale Attention Decoder
F_e^1
F_e^2
F_e^3
F_e^4
A_f^1
A_f^2
F
F_d^1
F_d^2
F_d^3
F_d^4
F_e^5, F_d^5
Sal map
Feature Fusion
F_d^s
Up-sample
RC
A_f^{s-1}
Concat
F_e^{s-1}
Conv1×1
F_d^{s-1}
$s \in \{2,3\}$
$s \in \{4,5\}$
Conv
Backbone Convolution Block
RC
Residual Connection
SRR
Spatial Relational Reasoning
CRR
Channel Relational Reasoning
Concat
Concatenation
Up-sample
2× spatial up-sampling operation

Relational Reasoning Encoder



Graph Construction

$$\tilde{\Lambda}(G^s) = diag(conv_{1\times1}(avepool(G^s)))$$

$$\tilde{A}_{ij} = (conv_{1\times1}(G^s))_i \cdot \tilde{\Lambda}(G^s) \cdot (conv_{1\times1}(G^s))_j^T$$

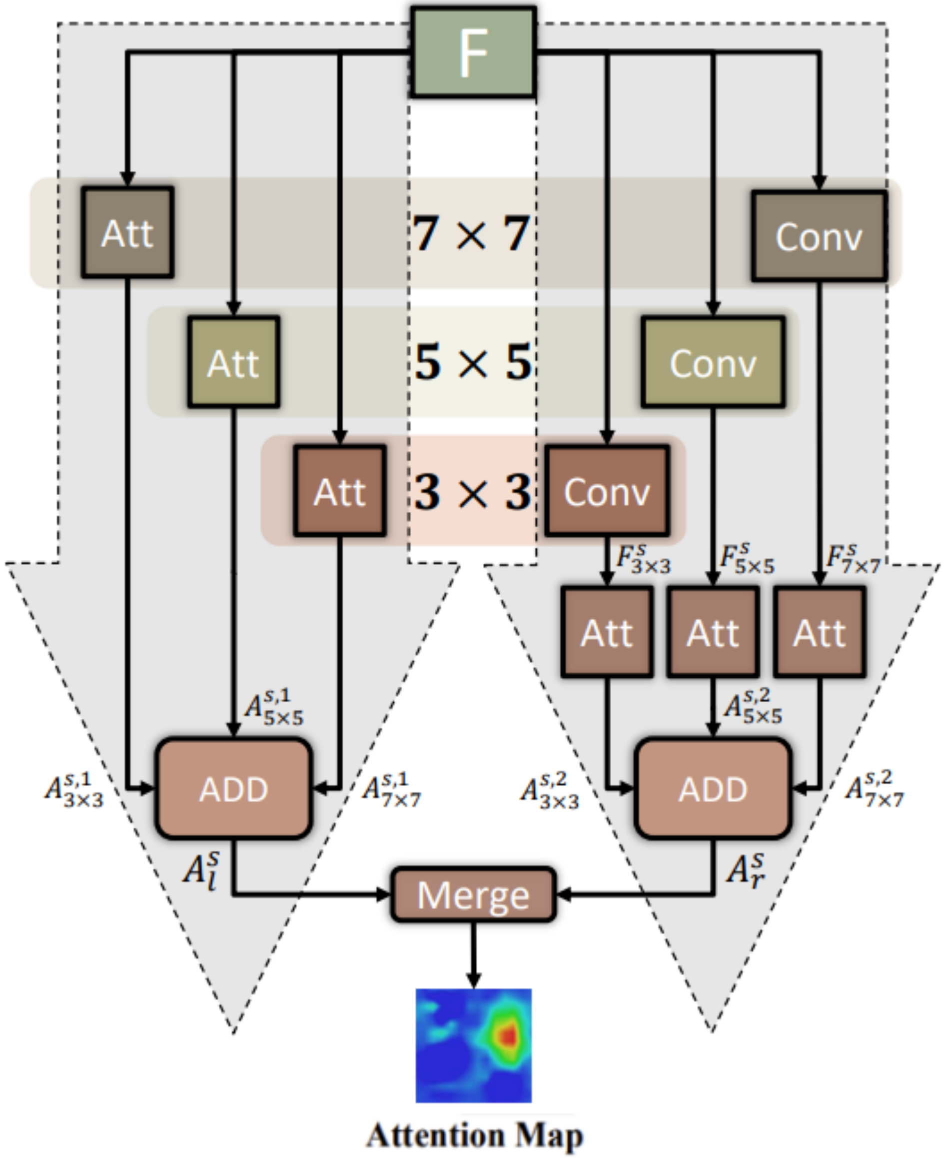
Graph Convolution

$$\tilde{L} = I - \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$$

$$F_r^s = \sigma(\tilde{L} G^s \Theta)$$

We design a relational reasoning module in the **high-level layers** of the encoder stage to model the sematic relations and force the generation of complete salient objects. This is the **first attempt** to introduce relational reasoning in the SOD framework for optical RSIs. Moreover, we innovatively employ relational reasoning **along the spatial and channel dimensions jointly** to obtain more comprehensive semantic relations.

Multi-scale Attention Decoder



We propose a **parallel multi-scale attention** scheme in the **low-level layers** of the decoder stage to recover the detail information in a multi-scale and attention manner. This mechanism can deal with the **object scale variation** issue through the multi-scale design, while effectively recovering the **detail information** with the help of shallower features selected by the parallel attention.

Left Branch

$$A^{s,l}_{3\times3} = \sigma(conv_{3\times3}(\Gamma^s; \hat{\theta}_{3\times3}))$$
$$A^{s,l}_{5\times5} = \sigma(conv_{5\times5}(\Gamma^s; \hat{\theta}_{5\times5}))$$
$$A^{s,l}_{7\times7} = \sigma(conv_{7\times7}(\Gamma^s; \hat{\theta}_{7\times7}))$$

$$A^s_l = \frac{1}{3}(A^{s,l}_{3\times3} \oplus A^{s,l}_{5\times5} \oplus A^{s,l}_{7\times7})$$

Right Branch

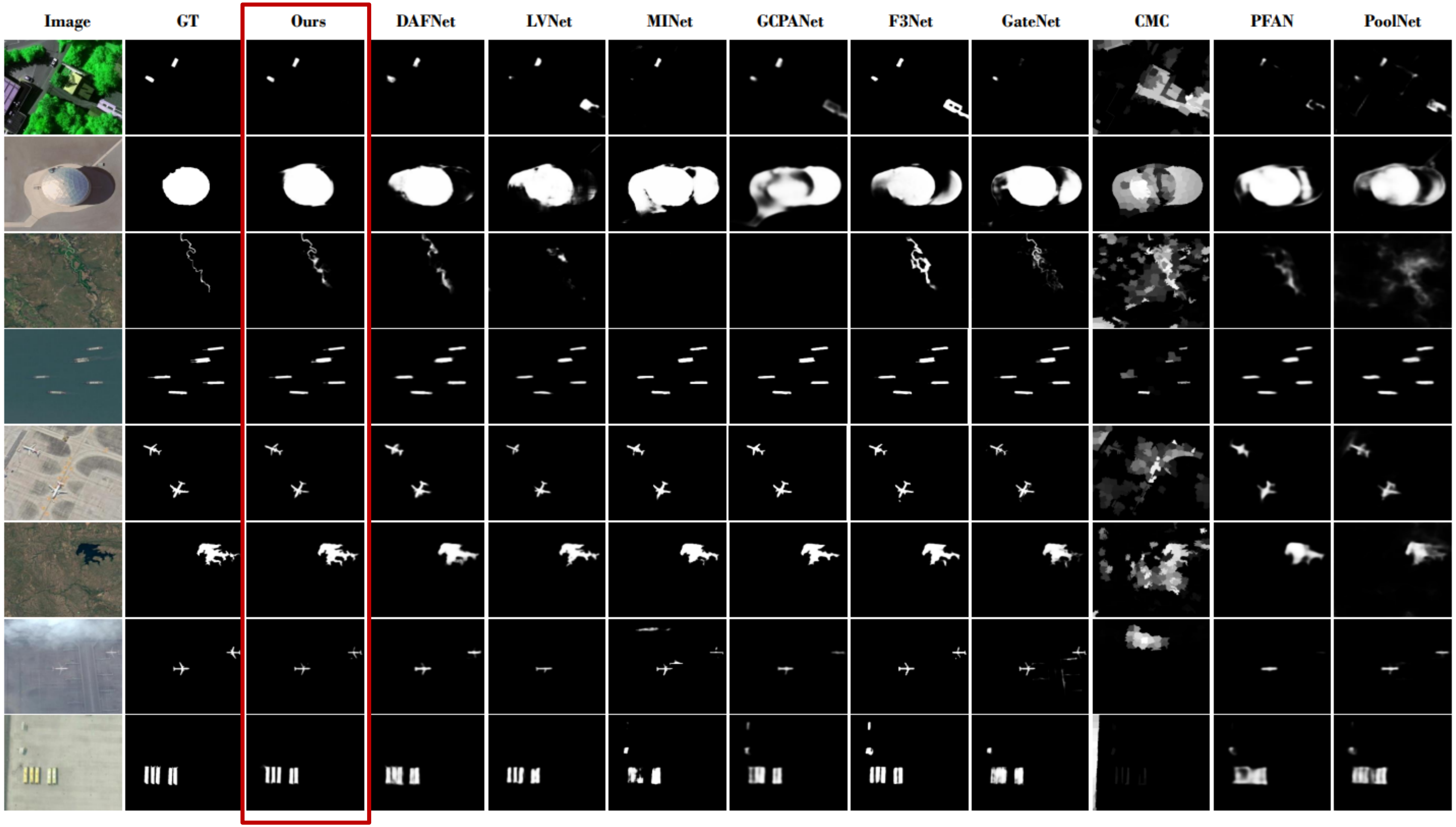
$$F^s_{3\times3} = \sigma(conv_{3\times3}(X^s; \hat{\omega}_{3\times3})),$$
$$F^s_{5\times5} = \sigma(conv_{5\times5}(X^s; \hat{\omega}_{5\times5})),$$
$$F^s_{7\times7} = \sigma(conv_{7\times7}(X^s; \hat{\omega}_{7\times7})),$$

$$A^s_r = \frac{1}{3}(A^{s,r}_{3\times3} \oplus A^{s,r}_{5\times5} \oplus A^{s,r}_{7\times7}).$$

Fusion

$$A^s_f = \sigma(conv_{1\times1}(concat(A^s_l, A^s_r)))$$

Experiments



Experiments

$$F_\beta = \frac{(\beta^2+1)\cdot Precision \cdot Recall}{\beta^2 \cdot Precision + Recall},\qquad MAE = \frac{1}{H\times W}\sum_{y=1}^{H}\sum_{x=1}^{W}|S(x,y)-G(x,y)|,\qquad S = \alpha * S_o + (1-\alpha) * S_r,$$

	ORSSD Dataset				EORSSD Dataset			
	F_β	E_m	MAE	S_m	F_β	E_m	MAE	S_m
R3Net	.7698	.8907	.0409	.8092	.7989	.9547	.0170	.8305
RADF	.7865	.9123	.0386	.8252	.7966	.9227	.0162	.8332
PoolNet	.7911	.9604	.0358	.8403	.8012	.9358	.0209	.8301
PFAN	.8344	.9418	.0543	.8613	.7931	.9334	.0156	.8446
EGNet	.8585	.9727	.0215	.8780	.8310	.9600	.0109	.8692
GateNet	.8794	.9464	.0197	.8853	.8618	.9440	.0131	.8710
F3Net	.8661	.9433	.0215	.8949	.8681	.9487	.0119	.9040
GCPANet	.8833	.9545	.0186	.8865	.8546	.9448	.0123	.8674
MINet	.8751	.9423	.0171	.8865	.8510	.9354	.0104	.8909
SMFF	.4764	.7518	.1897	.5329	.5693	.7892	.1471	.5431
CMC	.4214	.7069	.1267	.6033	.3555	.6785	.1066	.5826
LVNet	.8414	.9342	.0207	.8815	.8213	.9302	.0146	.8642
DAFNet	.9192	.9699	.0105	.9188	.9060	.9684	**.0053**	.9185
Ours	**.9203**	**.9808**	**.0103**	**.9282**	**.9119**	**.9720**	.0076	**.9230**

TABLE II
ABLATION ANALYSIS ON THE EORSSD DATASET.

Baseline	PMA	SRR	CRR	F_β	E_m	MAE	S_m
✓				0.8302	0.9217	0.0148	0.8695
✓	✓			0.8819	0.9523	0.0105	0.9021
✓	✓	✓		0.8947	0.9582	0.0091	0.9156
✓	✓	✓	✓	**0.9119**	**0.9720**	**0.0076**	**0.9230**

TABLE III
FURTHER VALIDATION OF RR AND PMA ON THE EORSSD DATASET.

Modules		F_β	E_m	MAE	S_m
full model		**0.9119**	**0.9720**	**0.0076**	**0.9230**
RR	w/Non-local	0.9102	0.9691	0.0093	0.9225
PMA	w/o PMA(r)	0.9100	0.9707	0.0079	0.9227
	w/o PMA(l)	0.9037	0.9544	0.0089	0.9094

Contributions

a) We propose a novel **end-to-end** relational reasoning network with parallel multi-scale attention (RRNet) for SOD in optical RSIs, which consists of a **relational reasoning encoder** and a **multi-scale attention decoder**.

b) We design a **relational reasoning module** in the high-level layers of the encoder stage to model the sematic relations and force the generation of complete salient objects. This is the **first attempt** to introduce relational reasoning in the SOD framework for optical RSIs. Moreover, we innovatively employ relational reasoning **along the spatial and channel dimensions jointly** to obtain more comprehensive semantic relations.

c) We propose a **parallel multi-scale attention scheme** in the low-level layers of the decoder stage to **recover the detail information** in a multi-scale and attention manner. This mechanism can deal with the **object scale variation** issue through the multi-scale design, while effectively recovering the detail information with the help of shallower features selected by the parallel attention.

Future work

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

2. Extending the saliency detection task in different data sources, such as light filed image, RGB-D video, and underwater image.

3. New ideas and solutions in saliency detection task, such as instance-level saliency detection and segmentation, saliency improvement and refinement.

为天下储人才　　为国家图富强

机器智能与系统控制教育部重点实验室

Key Laboratory of Machine Intelligence & System Control, Ministry of Education

视觉感知与智能系统实验室成立于2013年9月，依托控制科学与技术国家A类学科，致力致力于智能系统感知、决策与应用领域的研究，团队包括国家级特聘教授1人、国家四青人才3人、泰山学者6人、中国科协青托1人。目前承担国家、省部级各类科研经费3000余万元，获得国内外学术奖励10余次。

张 伟

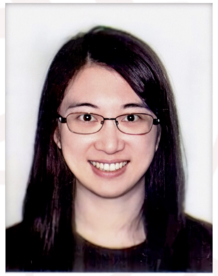
长江学者特聘教授

控制学院副院长

宋然 教授/博导
青年拔尖人才

元辉 教授/博导
国家优青

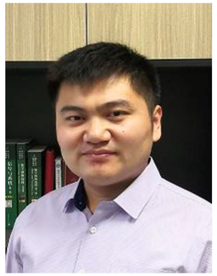
丛润民 教授/博导
中国科协青托



贾潇 教授/博导
青年泰山

张敬林 教授/博导
青年泰山

李腾 教授/博导
山东省优青



李帅 教授/博导
齐鲁青年学者

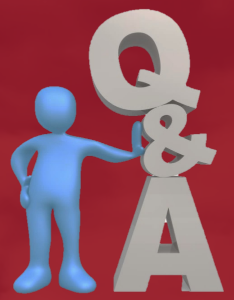
李晓磊 副教授/硕导
加州伯克利分校 博后

李振华 副教授/硕导
卡尔加里大学 博后

鲁威志 副教授/硕导
法国国立科学院 博士

程吉禹 副教授/硕导
香港中文大学 博士

Pourya 副教授/硕导
上海交通大学 博士



THANKS FOR WATCHING