



中国图象图形学学会  
优博论坛

# 光学遥感图像 显著性目标检测初探







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

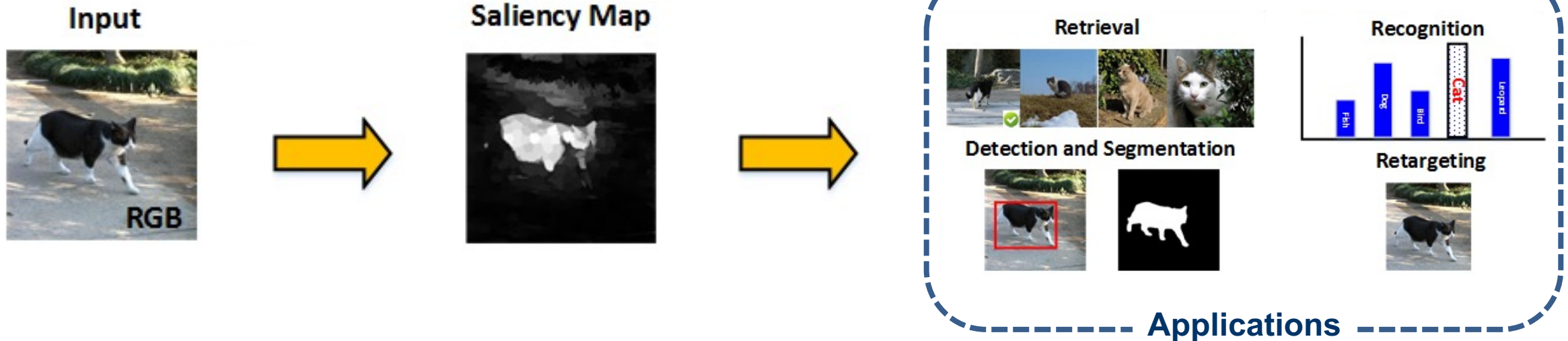
2023-03-26



# ▶ Outline

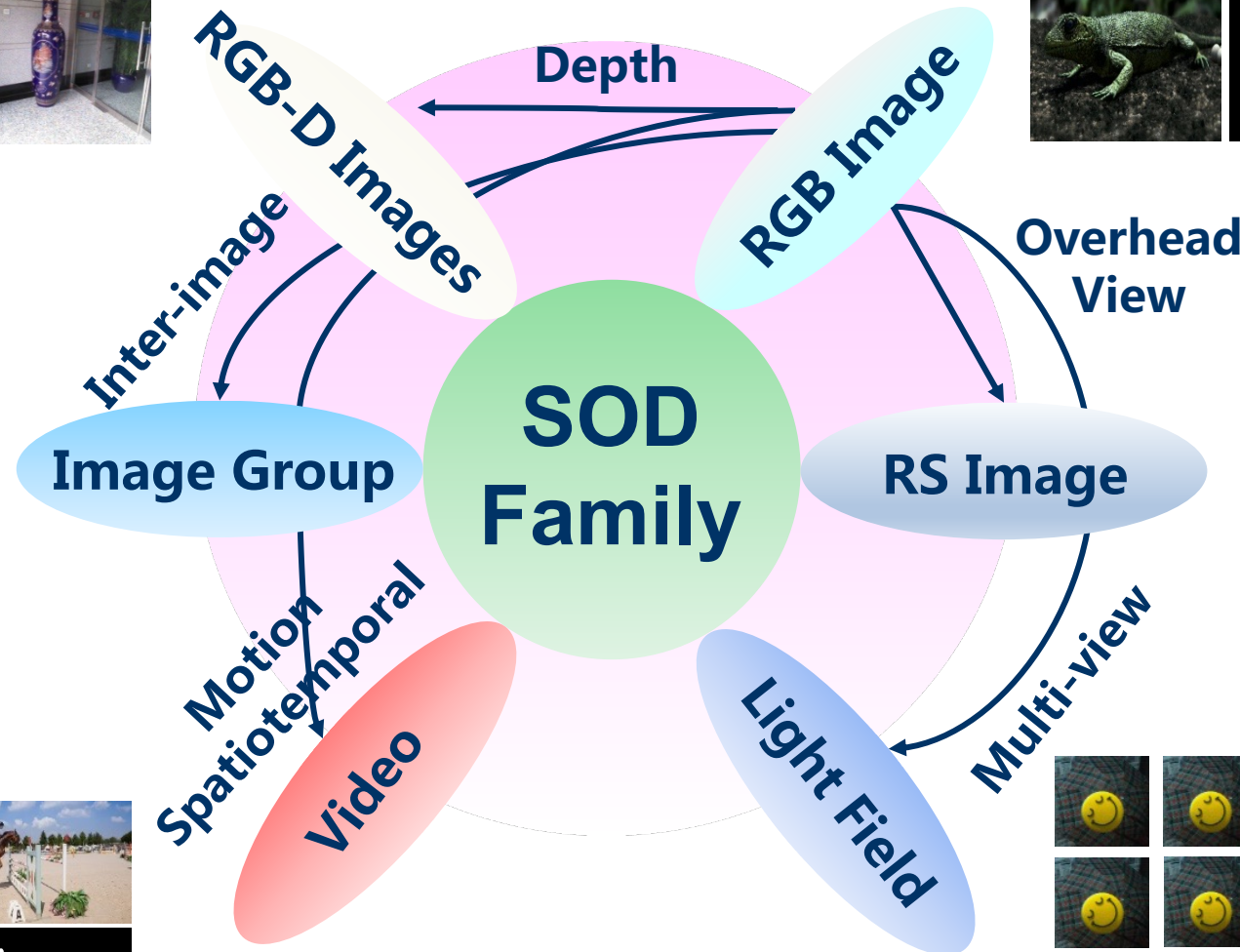
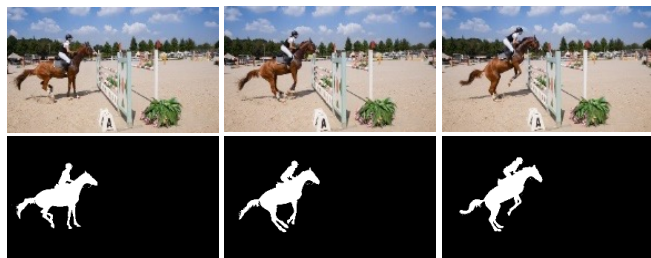
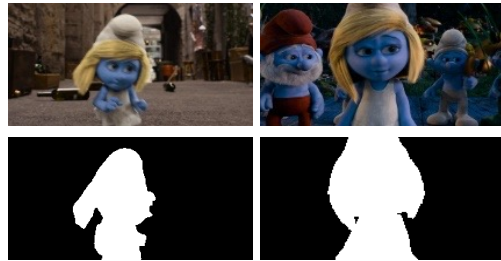
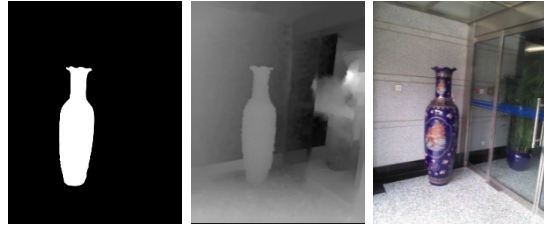
- Introduction
- Technical Methods
  - **TGRS 2019** — — Nested network with two-stream pyramid for salient object detection in optical remote sensing images  Highly Cited Paper
  - **TIP 2021** — — Dense attention fluid network for salient object detection in optical remote sensing images  Hot Paper  Highly Cited Paper
  - **TGRS 2022** — — RRNet: Relational reasoning network with parallel multi-scale attention for salient object detection in optical remote sensing images  Highly Cited Paper
- Future Work

# Introduction

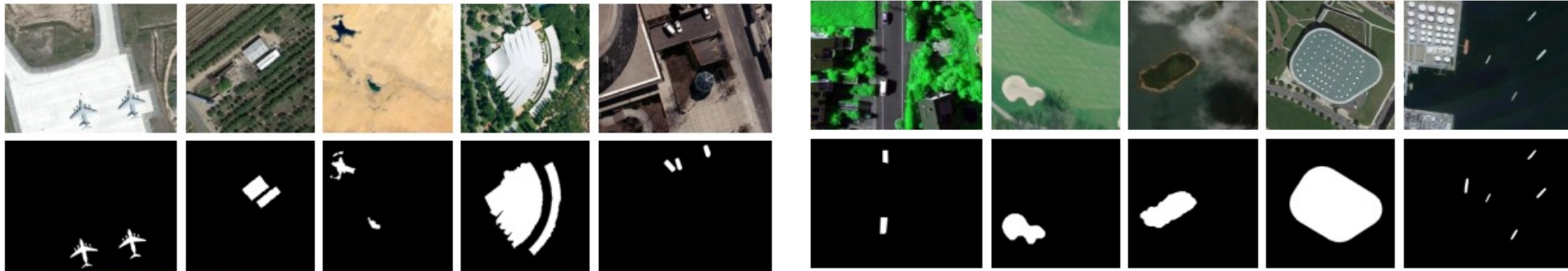


Simulating the human visual attention mechanism, salient object detection aims at detecting the salient regions automatically, which has been applied in image/video segmentation, image/video retrieval, image retargeting, video coding, quality assessment, action recognition, and video summarization.

# Introduction



## 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.

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

Chongyi Li, Runmin Cong\*, Junhui Hou, Sanyi Zhang, Yue Qian, Sam Kwong  
*IEEE Transaction on Geoscience and Remote Sensing, 2019*

[https://li-chongyi.github.io/proj\\_optical\\_saliency.html](https://li-chongyi.github.io/proj_optical_saliency.html)

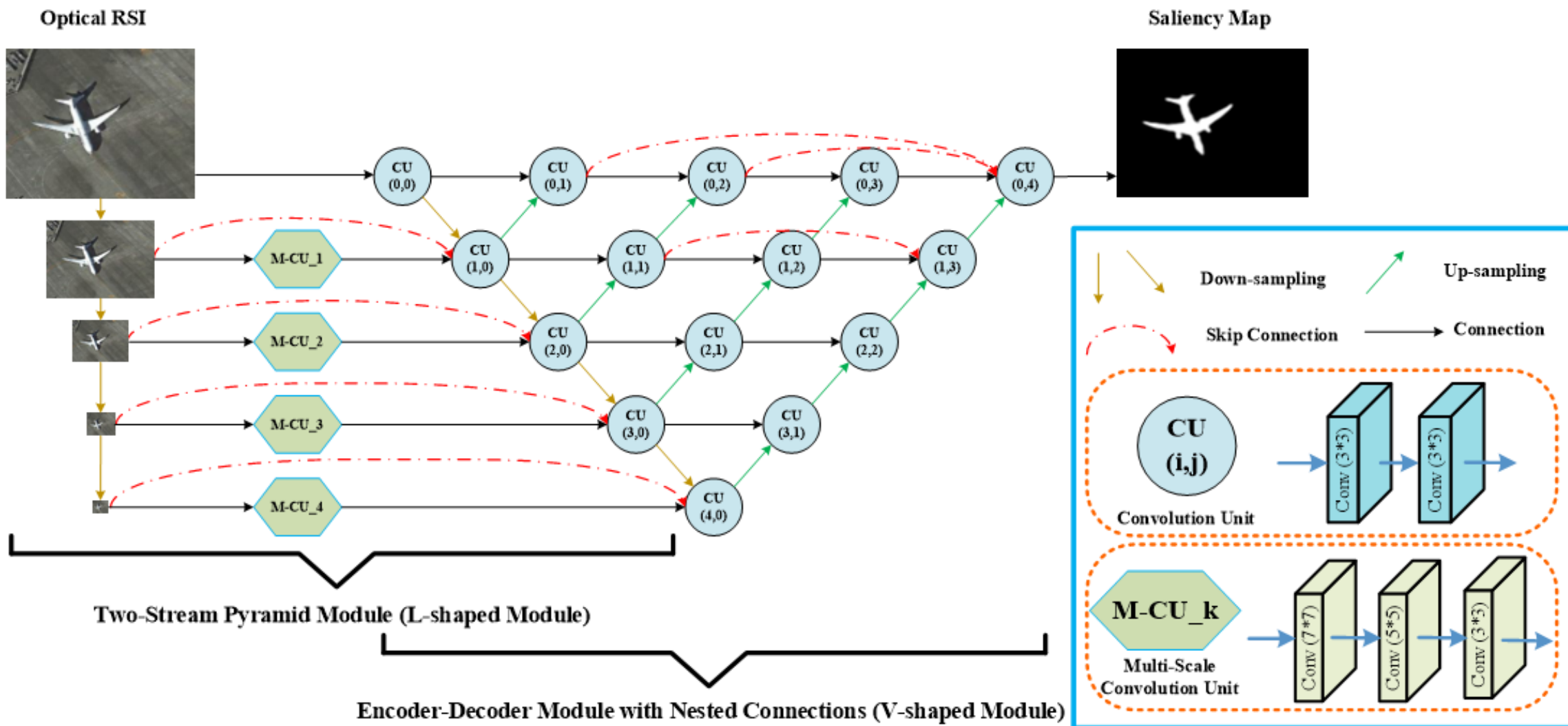
# Contributions

---



- a) 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.
- b) 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.
- c) 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.

# Our Method





# Two-Stream Pyramid Module

---

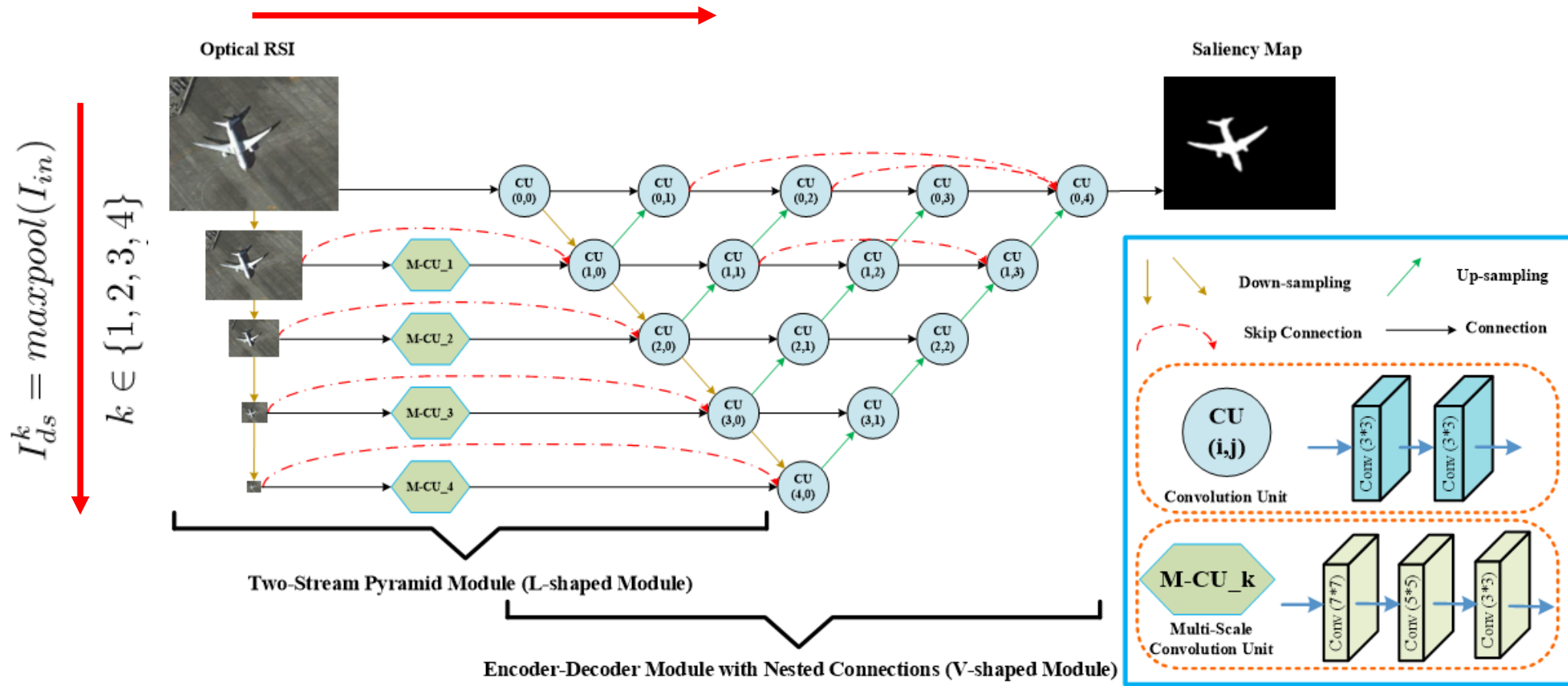


- The type and scale of the objects in the optical RSI are variable and diverse, including some small scaled airplanes or large bodies of water. To **deal with the scale variability of image patterns**, we design an input pyramid structure and pass scaled versions through our network.
- First, we progressively down-sample the input optical RSI for input pyramid generation. Then, we extract the multi-scale feature representations of each down-sampled input through a multi-scale convolution unit, and finally form a multi-scale feature pyramid.
- **The input pyramid preserves original detail features of input images, and the feature pyramid provides abstract semantic features.** We concatenate multi-resolution input versions and multi-scale features at different levels to form the two-stream pyramid and obtain complementary features.

# Two-Stream Pyramid Module



$$F_{7 \times 7}^k = \sigma(\mathbf{W}_{7 \times 7}^k * I_{ds}^k + \mathbf{b}_{7 \times 7}^k), \quad F_{5 \times 5}^k = \sigma(\mathbf{W}_{5 \times 5}^k * F_{7 \times 7}^k + \mathbf{b}_{5 \times 5}^k), \quad F_{3 \times 3}^k = \sigma(\mathbf{W}_{3 \times 3}^k * F_{5 \times 5}^k + \mathbf{b}_{3 \times 3}^k),$$



# Encoder-Decoder Module with Nested Connections

---



- The complementary features hierarchically extracted by the two-stream pyramid structure are passed to an encoder-decoder module, which **gradually integrates encoder detail features and decoder semantic features with nested connections**.
- At the end, the salient regions of an input optical RSI are predicted by the integrated features in a deeply supervised manner, where the encoder and decoder pathways are connected through a series of nested connections.
- The **nested connections would automatically select more discriminative saliency features** by the supervised learning, so that it could facilitate the fusion of encoder-decoder features and remit the interferences of cluttered and noisy backgrounds.

# Encoder-Decoder Module with Nested Connections

---



- To accurately capture the salient objects with exact boundaries, some encoder-decoder network architectures) usually concatenate encoder detail features and decoder semantical features through the brute-force skip connections (e.g., U-Net). Unfortunately, we found that the brute-force skip connections can degrade the quality of saliency prediction because the cluttered and noisy encoder features can also be passed through the prediction layer, especially for optical RSIs with complicated backgrounds. The 'bad' features seriously affect the accuracy of the saliency prediction. Therefore, **we use the nested connections to gradually filter out the 'bad' distractive features and make salient objects stand out by task-driven learning.**

# Loss Function

---



$$L = -(y \log(z) + (1 - y) \log(1 - z))$$

- we found that this loss function does not always work ( $L \rightarrow \infty$ ) when the predicted score  $z$  is 0 or 1. It is possible for optical RSI when there is no salient object. Thus, we rewrite the sigmoid cross-entropy loss as:

$$L = -(y \log(F_{clip}(z)) + (1 - y) \log(1 - F_{clip}(z)))$$

- where  $F_{clip}$  is a function that returns a tensor of the same type and shape as input with its values clipped to  $\rho$  and  $\mu$ . Specifically, any values less than  $\rho$  are set to  $\rho$ , while any values greater than  $\mu$  are set to  $\mu$ .

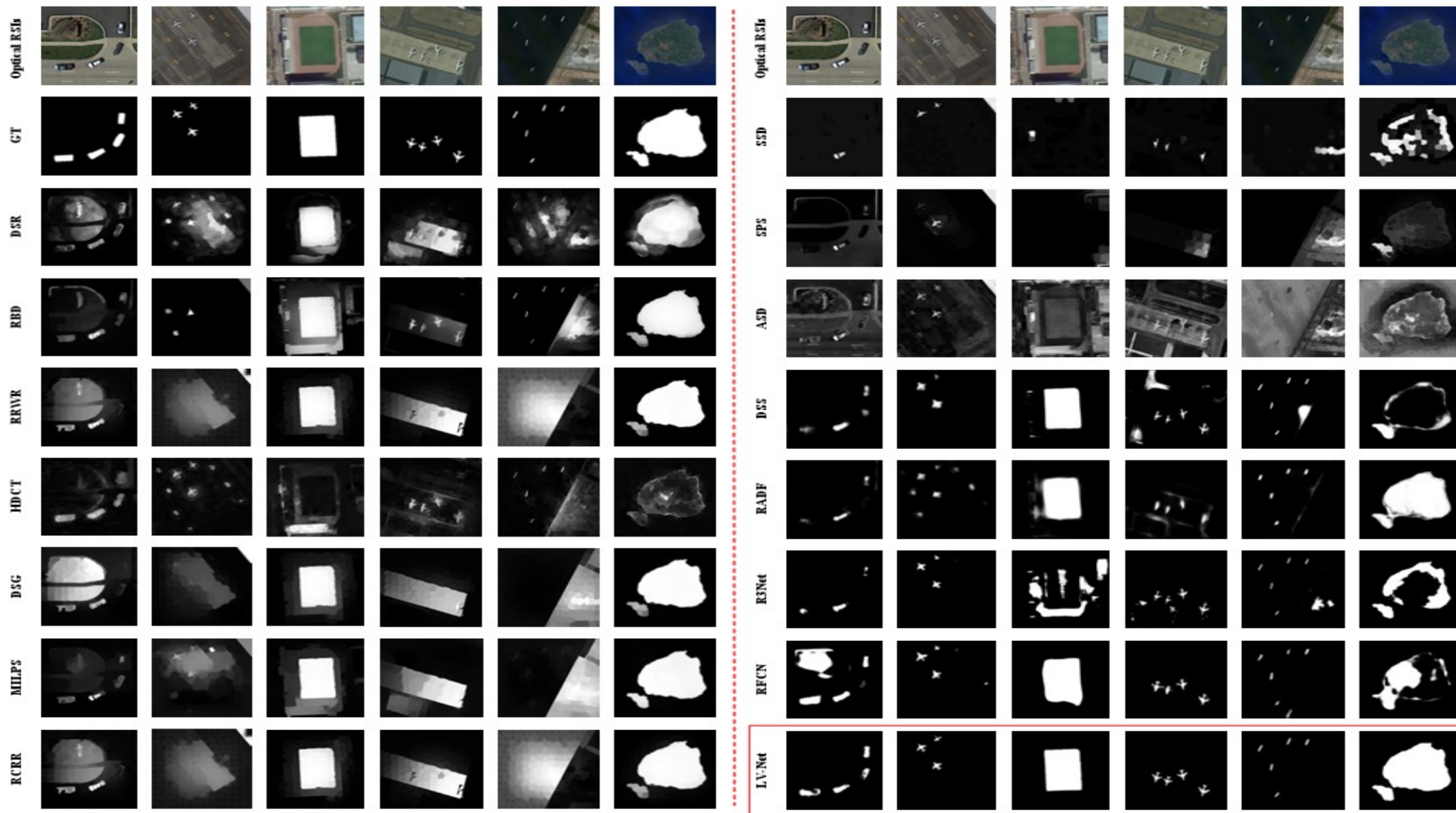
# ORSSD Dataset

---



- 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  $1264 \times 987$ ,  $800 \times 600$ , and  $256 \times 256$ , 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://li-chongyi.github.io/proj\\_optical\\_saliency.html](https://li-chongyi.github.io/proj_optical_saliency.html).

# Experiments



# Experiments

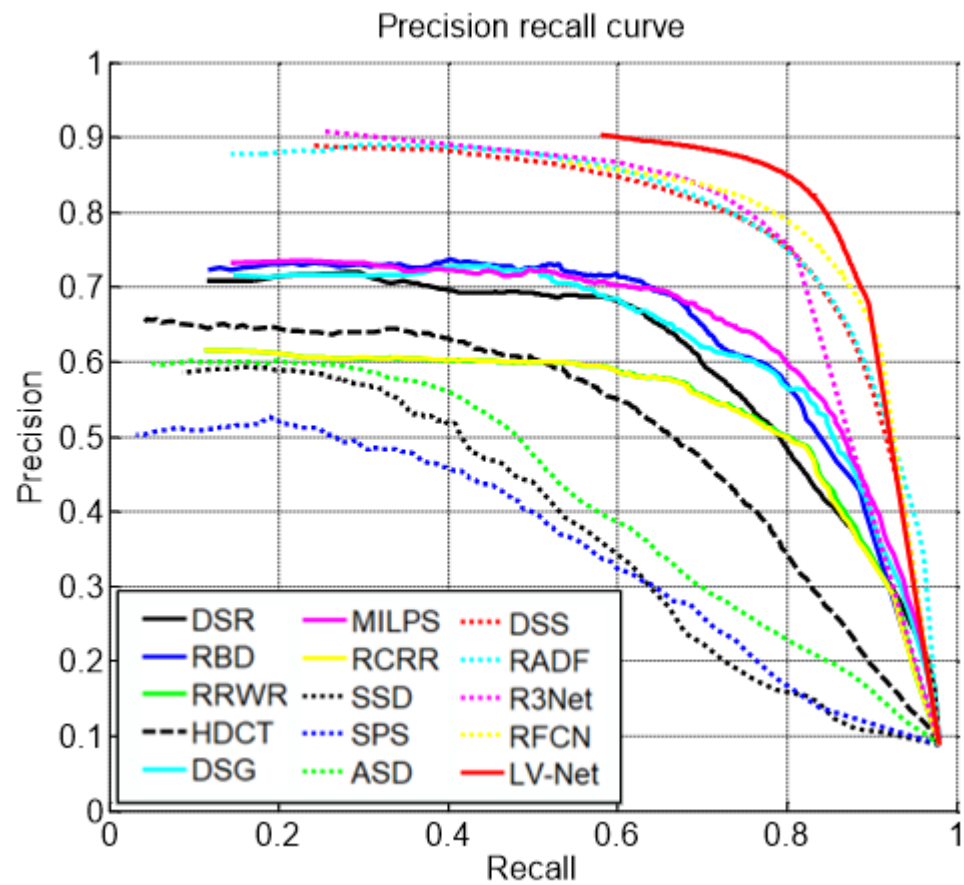


TABLE II  
QUANTITATIVE COMPARISONS WITH DIFFERENT METHODS ON THE TESTING SUBSET OF ORSSD DATASET.

Method	Precision	Recall	$F_{\beta}$	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	<b>0.8672</b>	<b>0.7653</b>	<b>0.8414</b>	<b>0.0207</b>	<b>0.8815</b>



# Conclusion

---



- In this paper, **we proposed the LV-Net for salient object detection in optical RSIs**. Benefiting from both the two-stream pyramid module and the nested connections, the proposed LV-Net can accurately locate the salient objects with diverse scales and effectively suppress the cluttered backgrounds.
- Moreover, **we constructed an optical RSI dataset for salient object detection with pixel-wise annotation**.
- Experiments demonstrate the proposed method **significantly outperforms** the state-of-the-art methods both qualitatively and quantitatively. The module analysis and parameter discussion demonstrate the effectiveness of each designed component and the parameter settings in the proposed LV-Net



Hot Paper



Highly Cited Paper

# Dense Attention Fluid Network for Salient Object Detection in Optical Remote Sensing Images

Qijian Zhang, Runmin Cong\*, Chongyi Li, Ming-Ming Cheng,  
Yuming Fang, Xiaochun Cao, and Yao Zhao

*IEEE Transactions on Image Processing, 2021*

[https://rmcong.github.io/proj\\_DAFNet.html](https://rmcong.github.io/proj_DAFNet.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**.

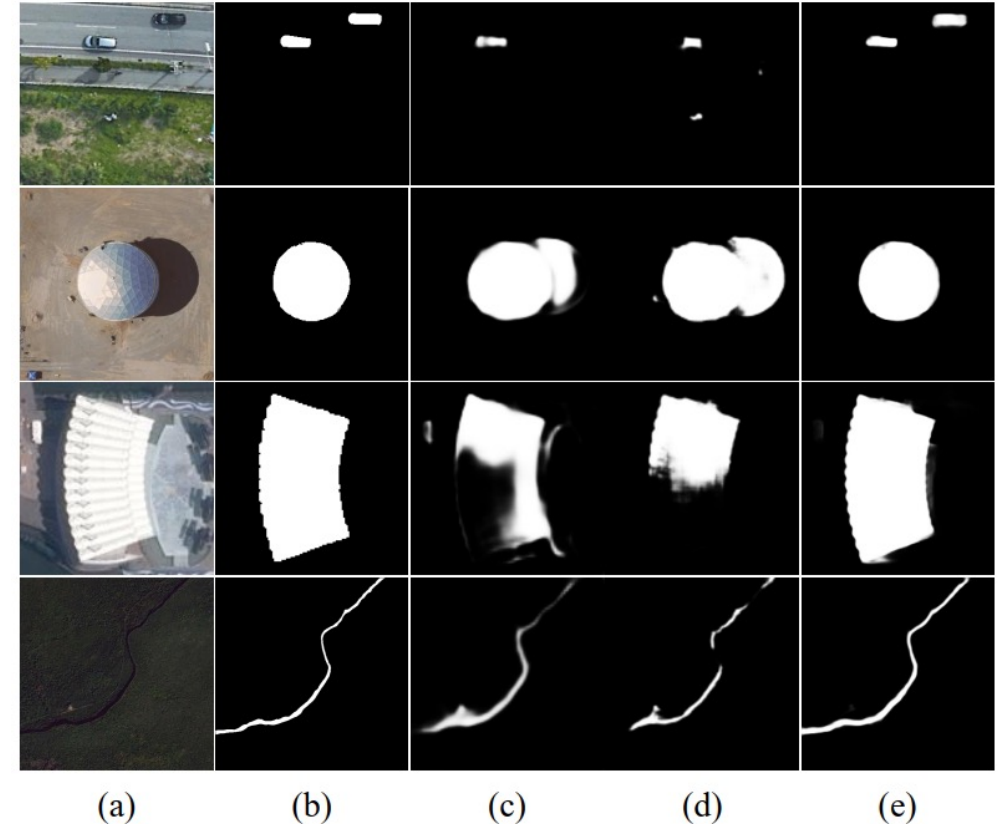


Fig. 1. Visual illustration of SOD results for optical RSIs by applying different methods. (a) Optical RSIs. (b) Ground truth. (c) PFAN [11]. (d) LVNet [6]. (e) Proposed DAFNet.

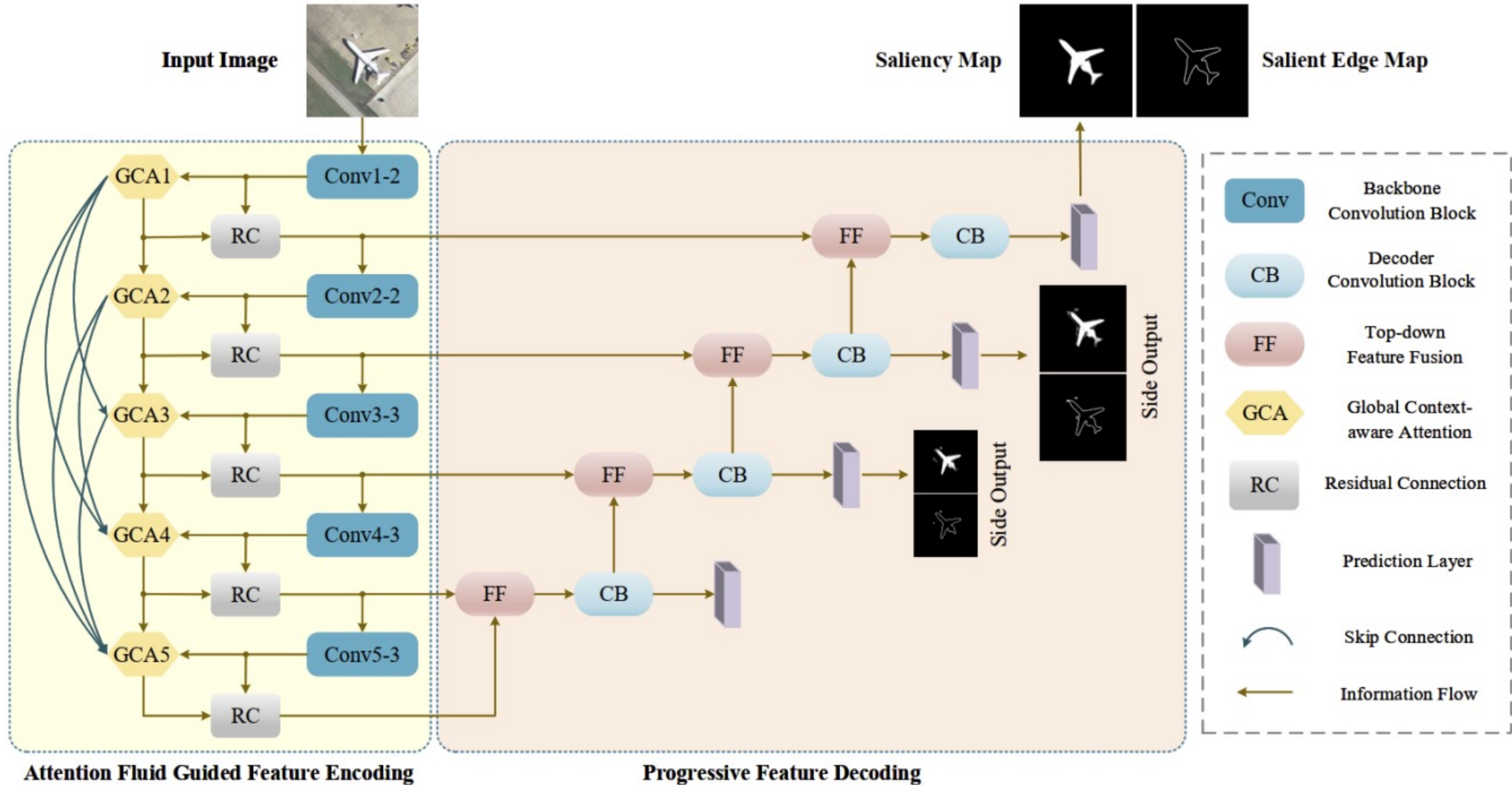
# Contributions

---



- a) An end-to-end Dense Attention Fluid Network (DAFNet) is proposed to achieve SOD in optical RSIs, equipped with a **Dense Attention Fluid (DAF) structure** decoupled from the backbone feature extractor and a **Global Context-aware Attention (GCA) mechanism**.
- b) The DAF structure is designed to **combine the multi-level attention cues**, where shallow-layer attention cues flow into the attention units of deeper layers so that low-level attention cues could **be propagated as guidance information to enhance the high-level attention**.
- c) The GCA mechanism is proposed to **model the global context semantic relationships** by a global feature aggregation module, and **tackle the scale variation** by a cascaded pyramid attention module.
- d) A **large-scale benchmark dataset** including 2, 000 images and corresponding pixel-wise annotations is constructed for SOD in optical RSIs. The proposed DAFNet **consistently outperforms 15 state-of-the-art competitors** in the experiments.

# Our Method



# Attention Fluid Guided Feature Encoding

---



- The attention fluid guided feature encoding consists of:
  - ◆ a feature fluid that generates hierarchical feature representations with stronger discriminative ability by incorporating attention cues mined from the corresponding global context-aware attention modules.
  - ◆ an attention fluid where low-level attention maps flow into deeper layers to guide the generation of high-level attentions .

# Global Context-aware Attention Mechanism



- We investigate a novel **global context-aware attention (GCA) mechanism** that explicitly captures the long-range semantic dependencies among all spatial locations in an attention manner. The GCA module consists of two main functional components:
  - ◆ The **global feature aggregation (GFA)** module consumes raw side features generated from the backbone convolutional block and produces aggregated features that encode global contextual information.
  - ◆ The **cascaded pyramid attention (CPA)** module is used to address the scale variation of objects in optical RSIs, which takes the aggregated features from GFA as input and produces a progressively refined attention map under a cascaded pyramid framework.

# Global Context-aware Attention Mechanism



## ● Global Feature Aggregation

- The GFA module aims to **achieve feature alignment and mutual reinforcement between saliency patterns** by aggregating global semantic relationships among pixel pairs, which is beneficial to generate intact and uniform saliency map.
- Aggregated feature map  $F^S$  with global contextual dependencies:

$$F^S = f^S + \delta \cdot (f^S \odot G^S)$$

- Refined feature map  $F_g^S$  with more compact channel information:

$$F_g^S = F^S \odot \Gamma^S$$



# Global Context-aware Attention Mechanism

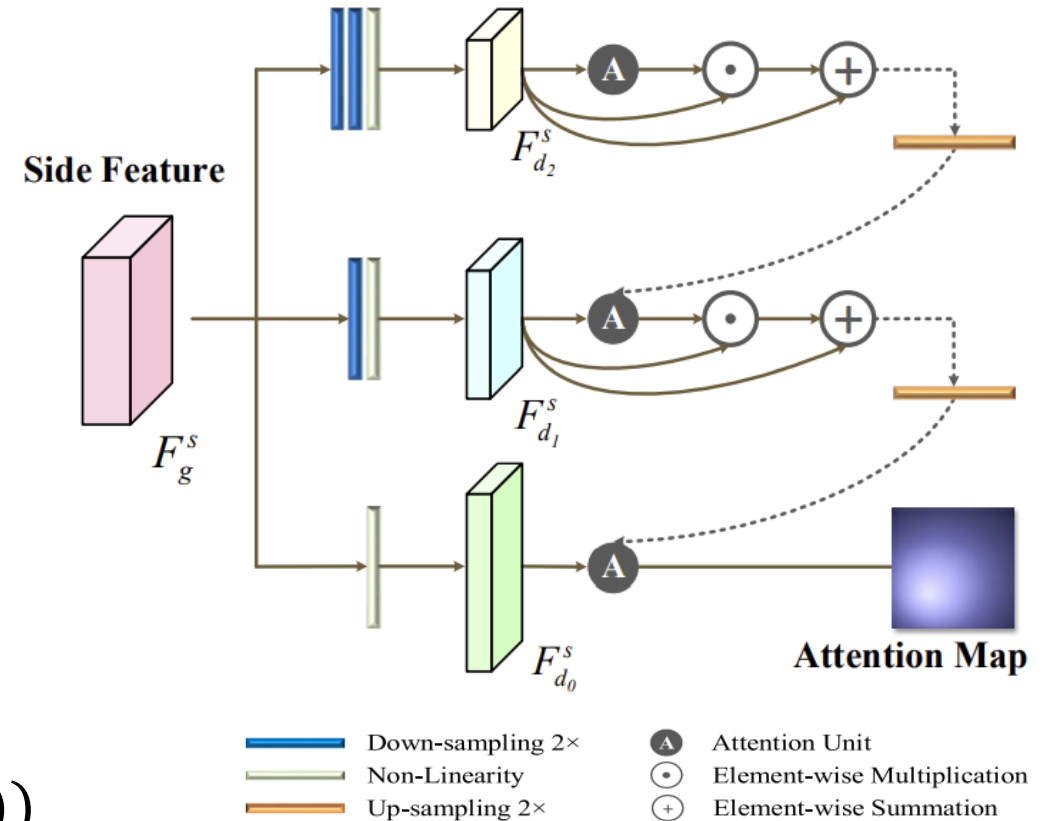


## ● Cascaded Pyramid Attention

- We design a **cascaded pyramid attention** to progressively refine both features and attentive cues **from coarse to fine**.
- The CPA module produces a full-resolution attention map  $\hat{A}^s$  at the original feature scale, which can be formulated as:

$$\hat{A}^s = Att(\text{concat}(F_{d_0}^s, (F_{d_1}^s \odot \boxed{A_{d_1}^s} + F_{d_1}^s) \uparrow))$$

$$A^s = Att(F_g^s) = \sigma \left( \text{conv} \left( \text{concat} \left( \text{avepool}(F_g^s), \text{maxpool}(F_g^s) \right); \hat{\theta} \right) \right)$$



# Dense Attention Fluid Structure



- Each GCA module consumes a raw side feature map  $f^s$ , and produces an attention map  $\hat{A}^s$ .
- First, we build **sequential connections among the attention maps** generated from hierarchical feature representations. Moreover, considering **the hierarchical attention interaction** among different levels, we add **feed-forward skip connections** to form the attention fluid. Formally, the above updating process is denoted as:

$$\hat{A}^s \leftarrow \sigma(\text{conv}(\text{concat}((\hat{A}^1) \downarrow, \dots, (\hat{A}^{s-1}) \downarrow, \hat{A}^s)))$$

- With the updated attention map, the final feature map at the  $s^{\text{th}}$  convolution stage  $F_c^s$  can be generated via the residual connection:

$$F_c^s = \text{concat}(F_{d_0}^s, (F_{out_1}^s) \uparrow) \odot (\hat{A}^s + O^s)$$

# Progressive Feature Decoding

---



- Each decoding stage consists of three procedures.
- First, we employ **top-down feature fusion (FF)** to align the spatial resolution and number of channels between adjacent side feature maps via up-sampling and  $1 \times 1$  convolution, and then perform pointwise summation.
- Second, a **bottleneck convolutional block (CB)** is deployed to further integrate semantic information from fusion features.
- Third, we deploy a **mask prediction layer** and an **edge prediction layer** for the decoded features, and use a Sigmoid layer to map the range of saliency scores into  $[0, 1]$ .
- The final output of our DAFNet is derived from the predicted saliency map at the top decoding level.

# Loss Function



- To accelerate network convergence and yield more robust saliency feature representations, we formulate a hierarchical optimization objective by applying deep supervisions to the side outputs at different convolution stages. We further introduce edge supervisions to capture fine-grained saliency patterns and enhance the depiction of object contours.

$$\ell = \sum_{s=1}^3 (\omega_m^s \cdot \ell_m^s + \omega_e^s \cdot \ell_e^s)$$

class-balanced binary  
cross-entropy loss  
function for saliency  
supervision

class-balanced binary  
cross-entropy loss  
function for salient edge  
supervision

# EORSSD Dataset

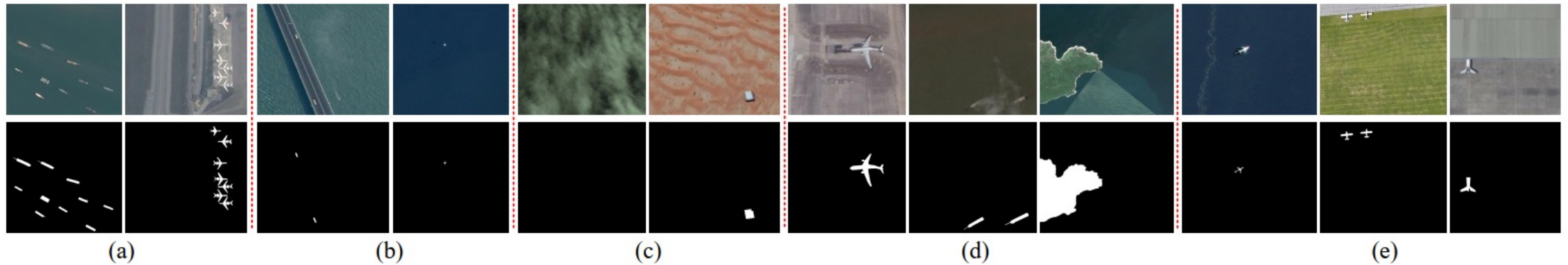


Fig. 4. Visualization of the more challenging EORSSD dataset. The first row shows the optical RSI, and the second row exhibits the corresponding ground truth. (a) Challenge in the number of salient objects. (b) Challenge in small salient objects. (c) Challenge in new scenarios. (d) Challenge in interferences from imaging. (e) Challenge in specific circumstances.

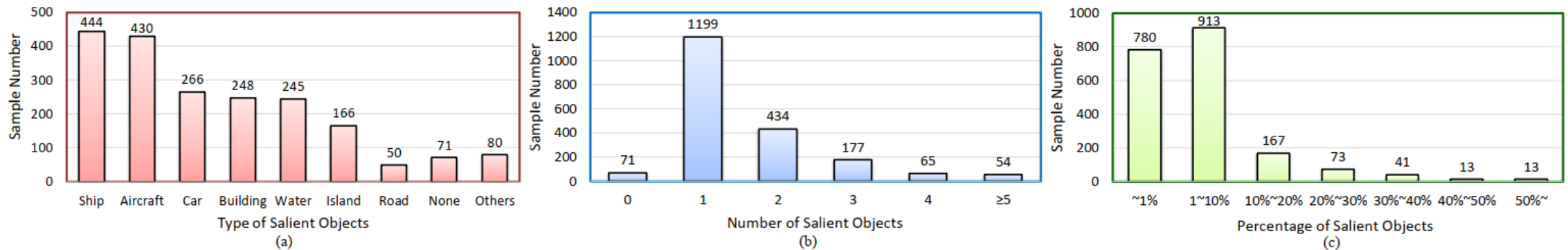


Fig. 5. Statistical analysis of EORSSD dataset. (a) Type analysis of salient object. (b) Number analysis of salient object. (c) Size analysis of salient object.

Download: <https://github.com/rmcong/EORSSD-dataset>

# Experiments

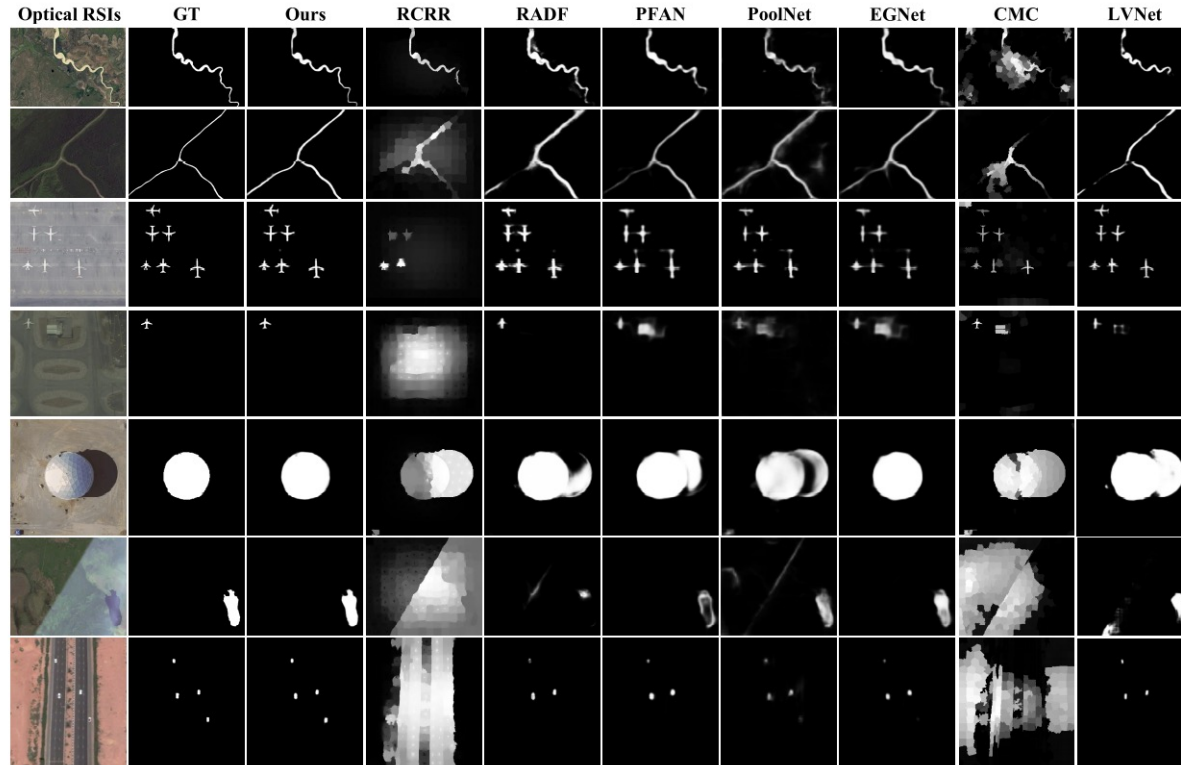


TABLE V

QUANTITATIVE EVALUATION OF ABLATION STUDIES ON THE TESTING SUBSET OF EORSSD DATASET

Baseline	GFA	CPA	DAF	$F_\beta$	MAE	$S_m$
✓				0.8391	0.0125	0.8432
✓	✓			0.8504	0.0098	0.8661
✓	✓	✓		0.8742	0.0083	0.8760
✓	✓	✓	✓	0.8922	0.0060	0.9167

	ORSSD Dataset			EORSSD Dataset		
	$F_\beta \uparrow$	MAE $\downarrow$	$S_m \uparrow$	$F_\beta \uparrow$	MAE $\downarrow$	$S_m \uparrow$
DSG [26]	0.6630	0.1041	0.7195	0.5837	0.1246	0.6428
RRWR [25]	0.5950	0.1324	0.6835	0.4495	0.1677	0.5997
HDCT [22]	0.5775	0.1309	0.6197	0.5992	0.1087	0.5976
SMD [23]	0.7075	0.0715	0.7640	0.6468	0.0770	0.7112
RCRR [24]	0.5944	0.1277	0.6849	0.4495	0.1644	0.6013
DSS [28]	0.7838	0.0363	0.8262	0.7158	0.0186	0.7874
R3Net [27]	0.7998	0.0399	0.8141	0.7709	0.0171	0.8193
RADF [29]	0.7881	0.0382	0.8259	0.7810	0.0168	0.8189
PFAN [11]	0.8344	0.0543	0.8613	0.7740	0.0159	0.8361
PoolNet [39]	0.7911	0.0358	0.8403	0.7812	0.0209	0.8218
EGNet [16]	<b>0.8438</b>	0.0216	0.8721	<b>0.8060</b>	<b>0.0109</b>	0.8602
CMC [46]	0.4214	0.1267	0.6033	0.3663	0.1057	0.5800
VOS [45]	0.4168	0.2151	0.5366	0.3599	0.2096	0.5083
SMFF [41]	0.4864	0.1854	0.5312	0.5738	0.1434	0.5405
LVNet [6]	0.8414	<b>0.0207</b>	<b>0.8815</b>	0.8051	0.0145	<b>0.8645</b>
DAFNet-V	<b>0.9174</b>	<b>0.0125</b>	<b>0.9191</b>	<b>0.8922</b>	<b>0.0060</b>	<b>0.9167</b>
DAFNet-R	<b>0.9235</b>	<b>0.0106</b>	<b>0.9188</b>	<b>0.9060</b>	<b>0.0053</b>	<b>0.9185</b>

# Conclusion

---



- This paper focuses on salient object detection in optical remote sensing images and proposes an end-to-end encoder-decoder framework dubbed as DAFNet, in which attention mechanism is incorporated to guide the feature learning.
- Benefiting from the attention fluid structure, our DAFNet learns to **integrate low-level attention cues into the generation of high-level attention maps in deeper layers**. Moreover, we **investigate the global context-aware attention mechanism to encode long-range pixel dependencies** and **explicitly exploit global contextual information**. In addition, we construct **a new large-scale optical RSI benchmark dataset** for SOD with pixel-wise saliency annotations.
- Extensive experiments and ablation studies demonstrate the effectiveness of the proposed DAFNet architecture.

# 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](https://rmcong.github.io/proj_RRNet.html)



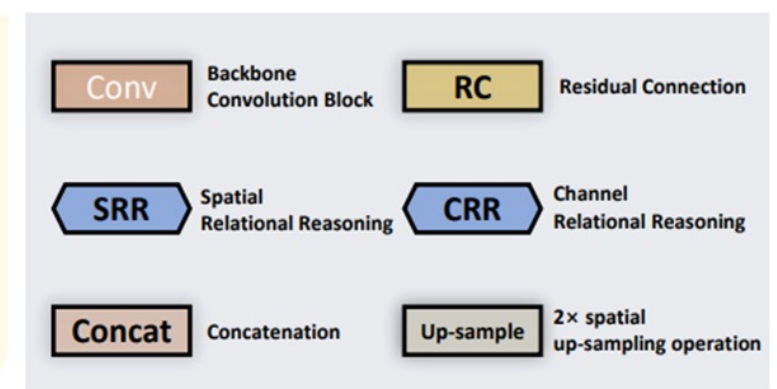
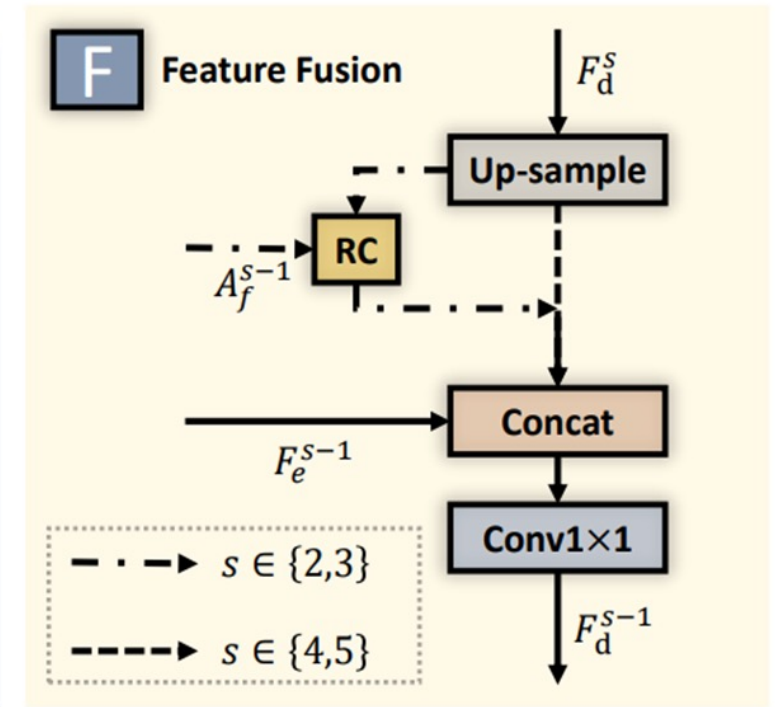
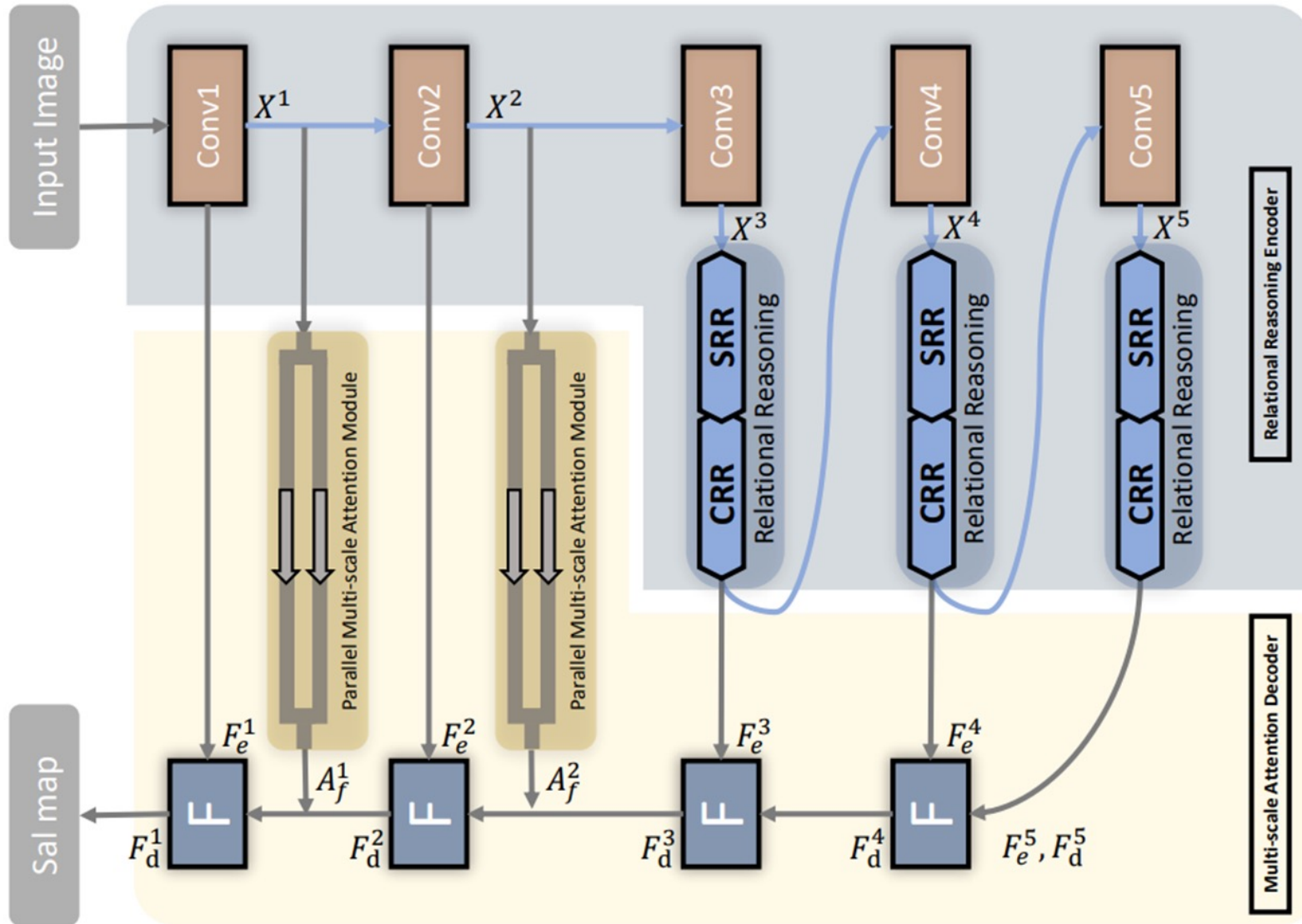
# Contributions

---

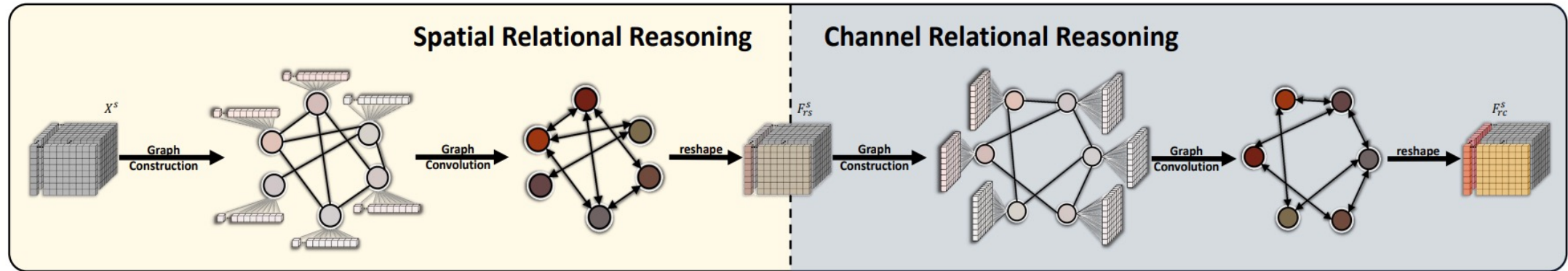


- a) We propose a novel **end-to-end** relational reasoning network with parallel multi-scale attention (RRNet) for SOD in optical RSIs, which consists of a **relational reasoning encoder** and a **multi-scale attention decoder**.
- b) We design a **relational reasoning module** in the high-level layers of the encoder stage to model the semantic 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.

# Our Method



# Relational Reasoning Encoder



## Graph Construction

$$\tilde{\Lambda}(G^s) = \text{diag}(\text{conv}_{1 \times 1}(\text{avepool}(G^s))).$$

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

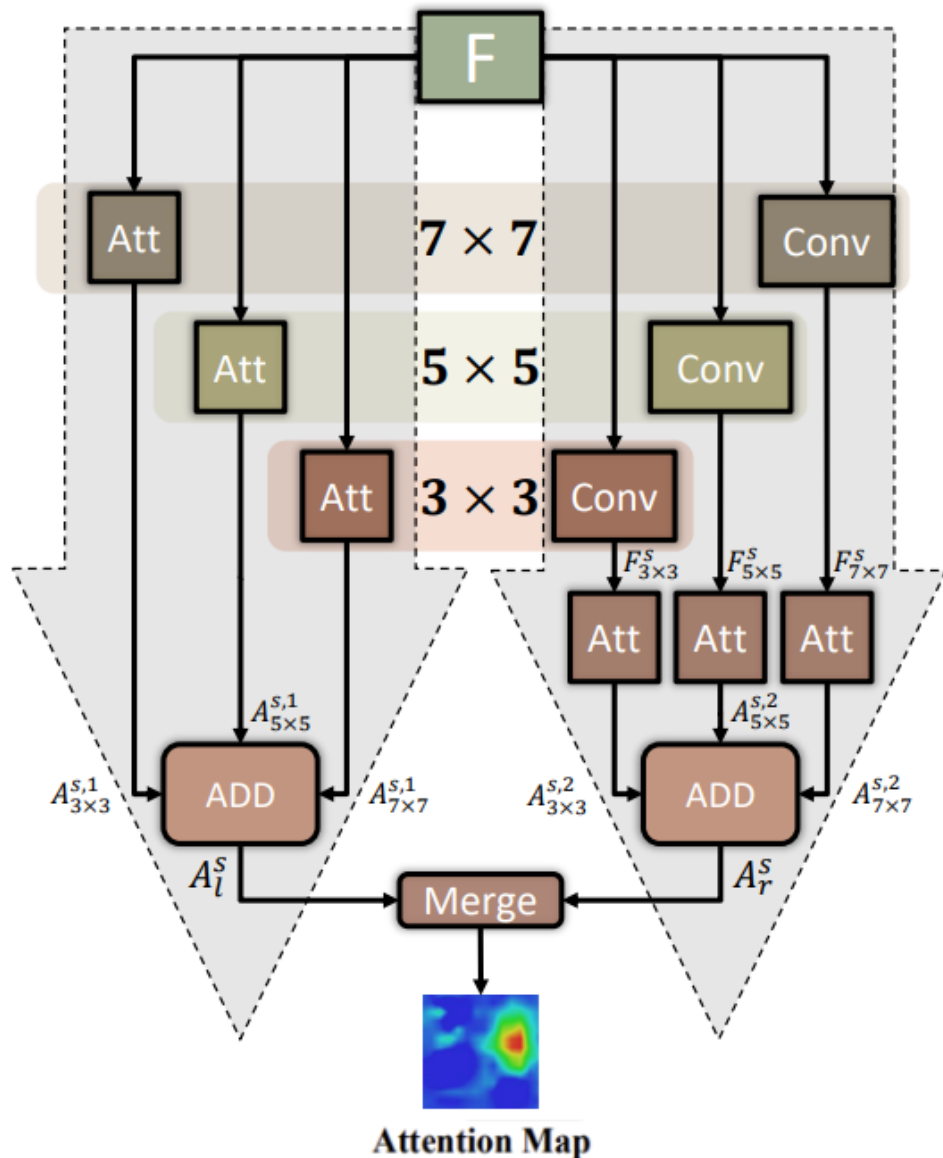
## Graph Convolution

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

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

We design a relational reasoning module in the **high-level layers** of the encoder stage to model the semantic 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

$$A_{3 \times 3}^{s,l} = \sigma(\text{conv}_{3 \times 3}(\Gamma^s; \hat{\theta}_{3 \times 3}))$$

$$A_{5 \times 5}^{s,l} = \sigma(\text{conv}_{5 \times 5}(\Gamma^s; \hat{\theta}_{5 \times 5}))$$

$$A_{7 \times 7}^{s,l} = \sigma(\text{conv}_{7 \times 7}(\Gamma^s; \hat{\theta}_{7 \times 7}))$$

$$A_l^s = \frac{1}{3}(A_{3 \times 3}^{s,l} \oplus A_{5 \times 5}^{s,l} \oplus A_{7 \times 7}^{s,l})$$

## Right Branch

$$F_{3 \times 3}^s = \sigma(\text{conv}_{3 \times 3}(X^s; \hat{\omega}_{3 \times 3})),$$

$$F_{5 \times 5}^s = \sigma(\text{conv}_{5 \times 5}(X^s; \hat{\omega}_{5 \times 5})),$$

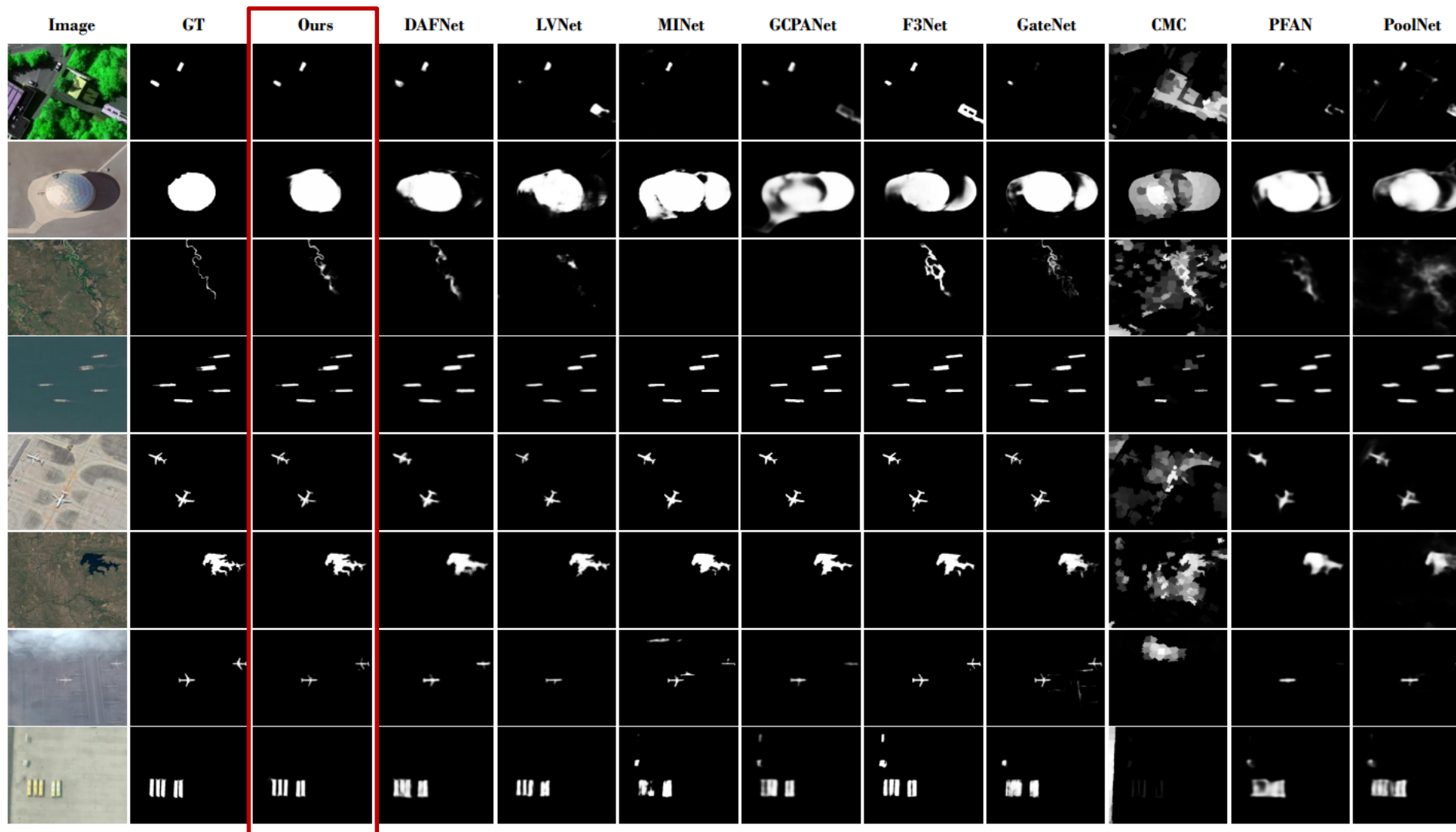
$$F_{7 \times 7}^s = \sigma(\text{conv}_{7 \times 7}(X^s; \hat{\omega}_{7 \times 7})),$$

$$A_r^s = \frac{1}{3}(A_{3 \times 3}^{s,r} \oplus A_{5 \times 5}^{s,r} \oplus A_{7 \times 7}^{s,r}).$$

## Fusion

$$A_f^s = \sigma(\text{conv}_{1 \times 1}(\text{concat}(A_l^s, A_r^s)))$$

# Experiments



# Experiments



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

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

$$S = \alpha * S_o + (1 - \alpha) * S_r,$$

	ORSSD Dataset				EORSSD Dataset			
	$F_{\beta}$	$E_m$	MAE	$S_m$	$F_{\beta}$	$E_m$	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	<b>.0053</b>	.9185
Ours	<b>.9203</b>	<b>.9808</b>	<b>.0103</b>	<b>.9282</b>	<b>.9119</b>	<b>.9720</b>	.0076	<b>.9230</b>

TABLE II  
ABLATION ANALYSIS ON THE EORSSD DATASET.

Baseline	PMA	SRR	CRR	$F_{\beta}$	$E_m$	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
✓	✓	✓	✓	<b>0.9119</b>	<b>0.9720</b>	<b>0.0076</b>	<b>0.9230</b>

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

Modules		$F_{\beta}$	$E_m$	MAE	$S_m$
<b>full model</b>		<b>0.9119</b>	<b>0.9720</b>	<b>0.0076</b>	<b>0.9230</b>
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

# Conclusion

---



- A novel end-to-end SOD model for optical RSIs is presented, named RRNet, which is capable of **reasoning semantic information** and **restoring detail information**.
- The relational reasoning in the spatial space and channel space is designed to model the **relationship** between different salient objects or different parts of the salient object, which can effectively **suppress background interference** and force the **generation of complete salient objects**.
- We propose a parallel multi-scale attention module that utilizes **attention mechanism** to improve the detection accuracy and **restore the details** of different scale objects through multi-scale design.
- Experimental evaluations over two datasets indicate that our method **outperforms** the state-of-the-art salient object detectors.

# Future work



1

Extending the existing saliency detection datasets in optical remote sensing image to include a wider range, more diverse, and challenging scene types.

2

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

3

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





**THANKS  
FOR WATCHING**