



信息科学研究所
Institute of Information Science



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

当视觉显著性检测遇上多源数据类型

—— 理论、模型与方法

汇报人：丛润民

2020年6月

汇报提纲

1 研究背景及意义

2 立体显著性检测

3 协同显著性检测

4 视频显著性检测

5 遥感显著性检测

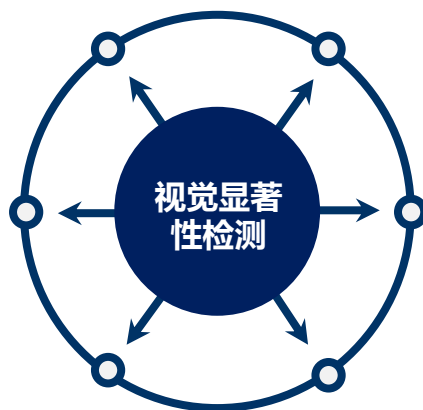
6 研究总结与展望

研究背景及意义

- 2017年7月国务院印发了《**新一代人工智能发展规划**》，并提出了六个方面的重点任务。其中，建立新一代人工智能基础理论体系的主旨版块要求突破应用基础理论瓶颈，进一步加强大数据智能、跨媒体感知计算、自主协同与决策、类脑智能计算理论等基础理论研究，并**着重开展跨学科探索性研究，推动人工智能与神经科学、认知科学等相关基础学科的交叉融合**。
- 人类的视觉系统可在大范围、复杂的场景中快速定位出最吸引注意的内容。研究人员也**希望计算机可以模拟人类的视觉注意机制**，具备自动定位场景中显著性内容的能力，进而为后续处理提供有效的辅助信息，这样“视觉显著性检测”任务应运而生。

交叉学科研究

视觉显著性检测是一个跨计算机科学、神经学、生物学、心理学的交叉学科方向；



应用前景广阔

研究领域：目标检测、图像编辑、视频浓缩、压缩编码、质量评价等
工程领域：智能拍照、推荐系统、自动驾驶、智慧医疗、对地监测等

研究背景及意义

多源数据驱动的视觉显著性检测



汇报提纲

1 研究背景及意义

2 立体显著性检测

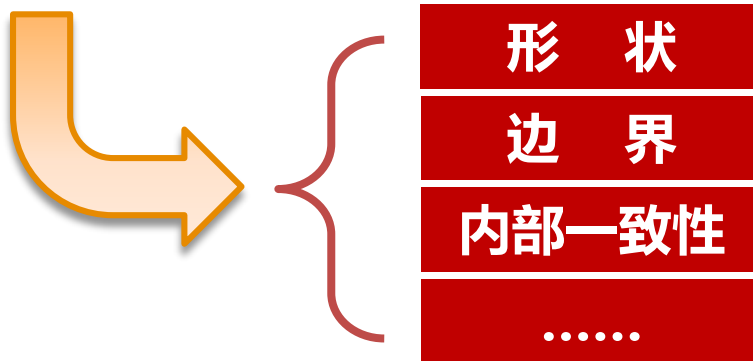
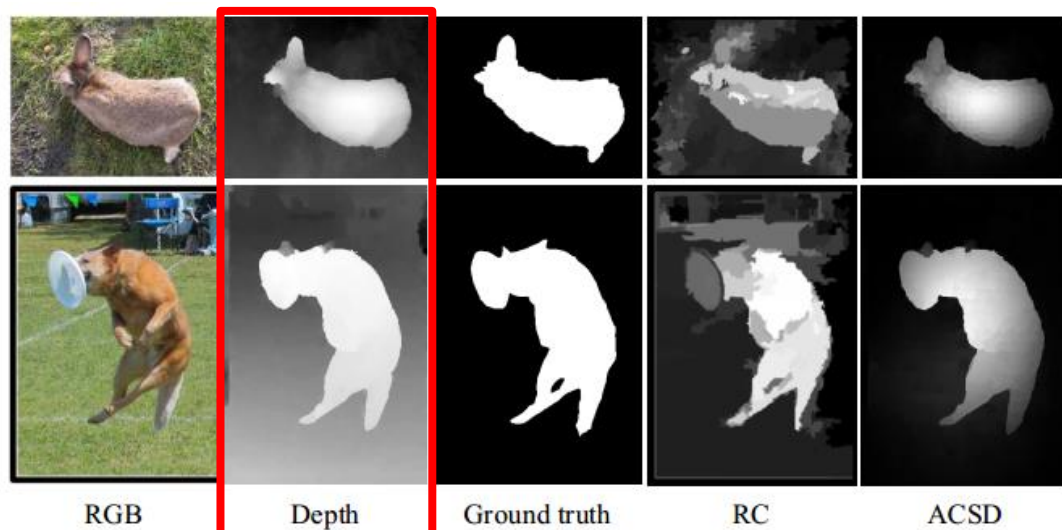
3 协同显著性检测

4 视频显著性检测

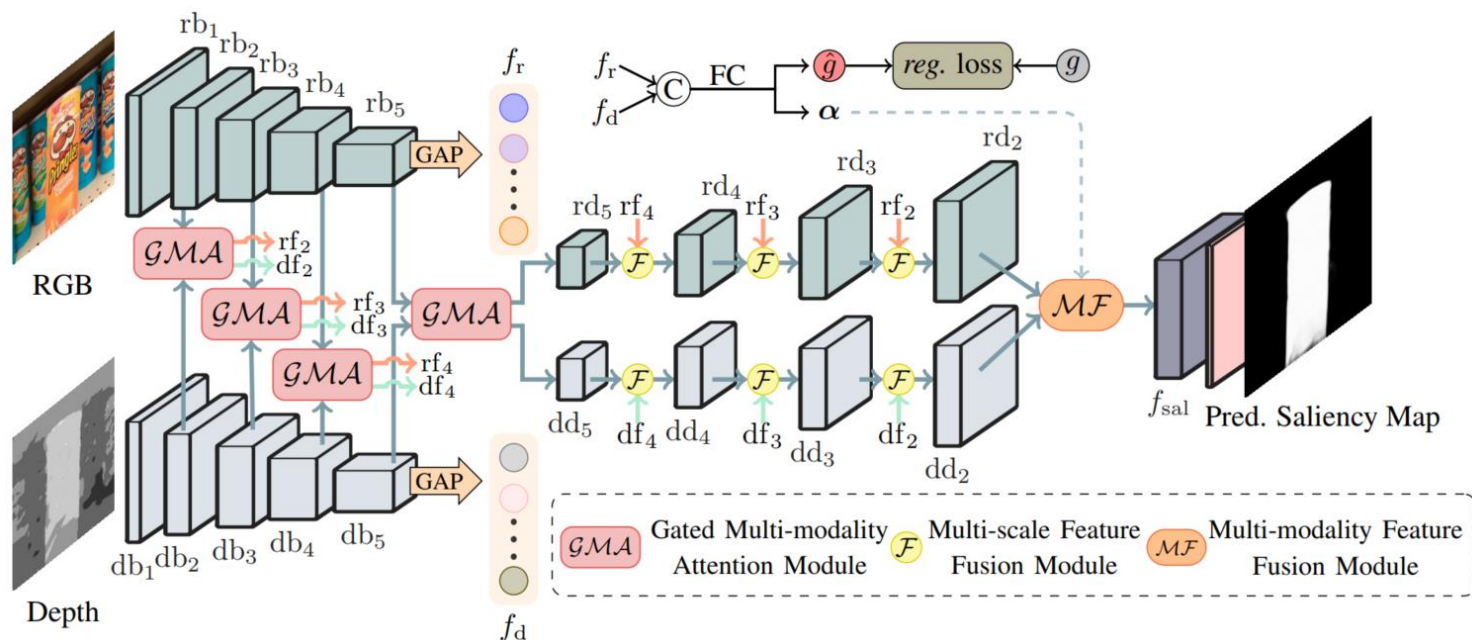
5 遥感显著性检测

6 研究总结与展望

立体显著性检测



方法1：深度置信感知的门控注意力网络



- 在不增加训练标签成本的前提下，设计了一种显著性检测任务导向的**深度置信感知模块**，通过网络训练自动感知深度图置信度响应，降低低质深度图对模型的影响；
- **多模态门控注意力模块**集成了跨模态互补信息，其中空间注意力机制用于降低信息冗余，门函数控制器用于调控跨模态信息融合比例；
- 无需任何预处理和后处理操作，八个数据集上均**性能最佳**。

方法1：深度置信感知的门控注意力网络

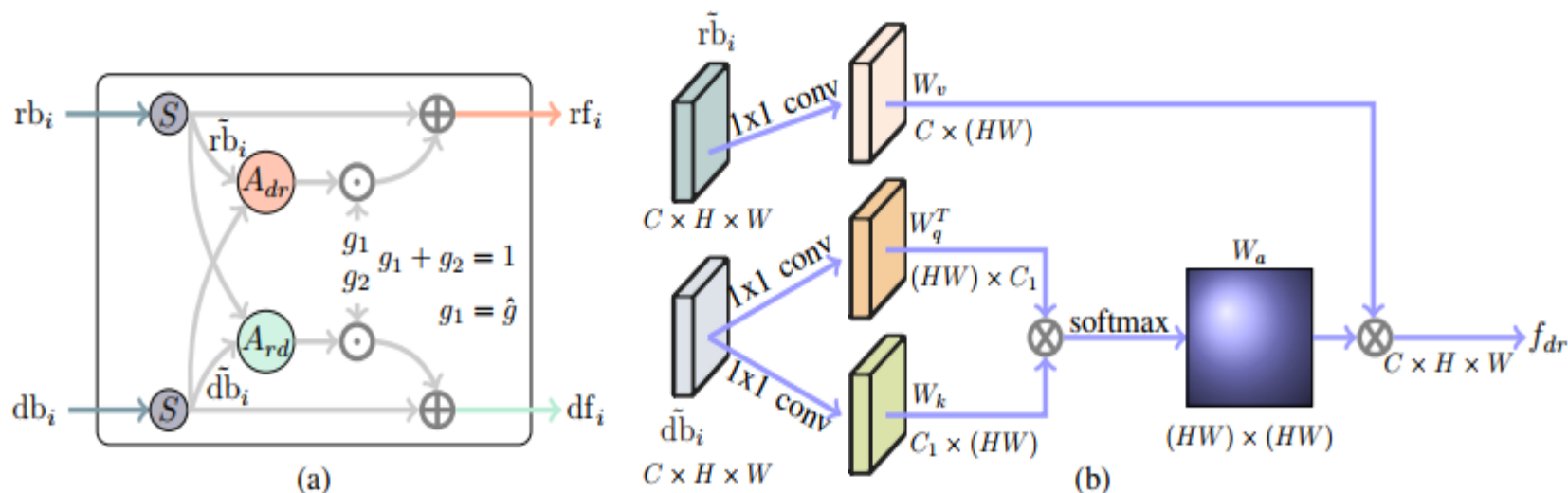
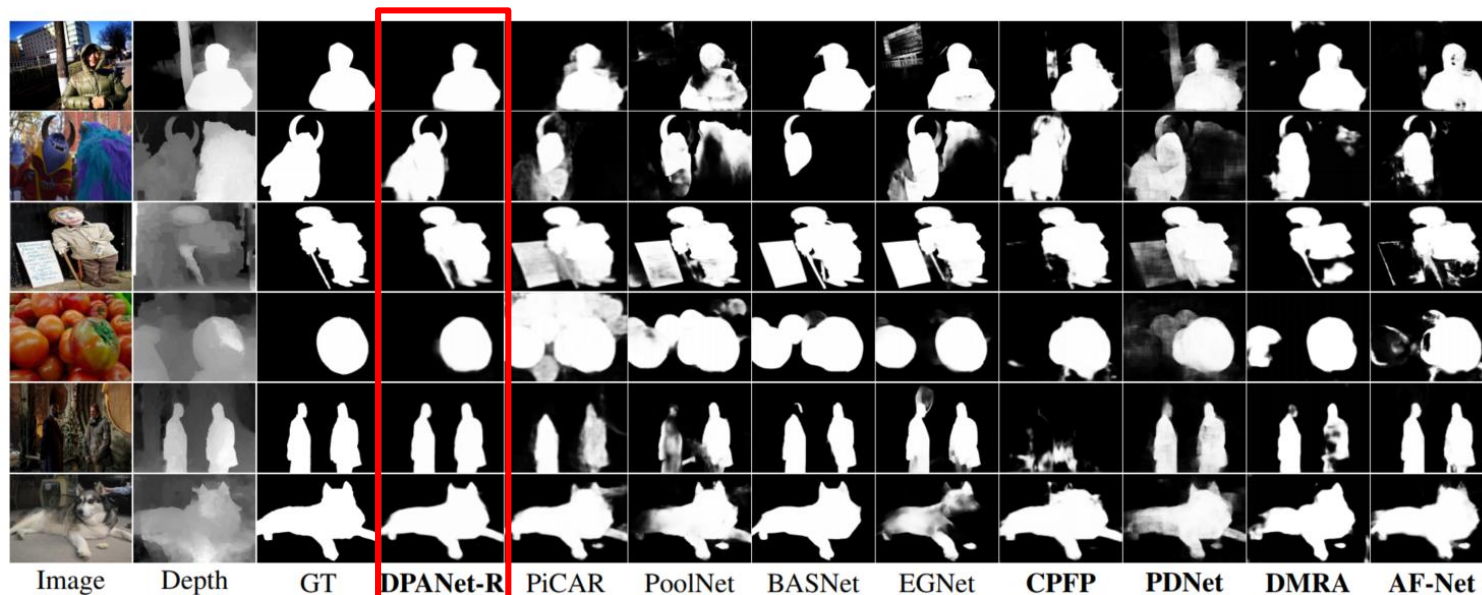


Fig. 3. Illustration of GMA module. (a) shows the constructure of GMA module, and (b) represents the operation A_{dr} . The operation A_{rd} is symmetrical to the A_{dr} (exchange the position of rb_i and db_i). For conciseness, we just show the A_{dr} .

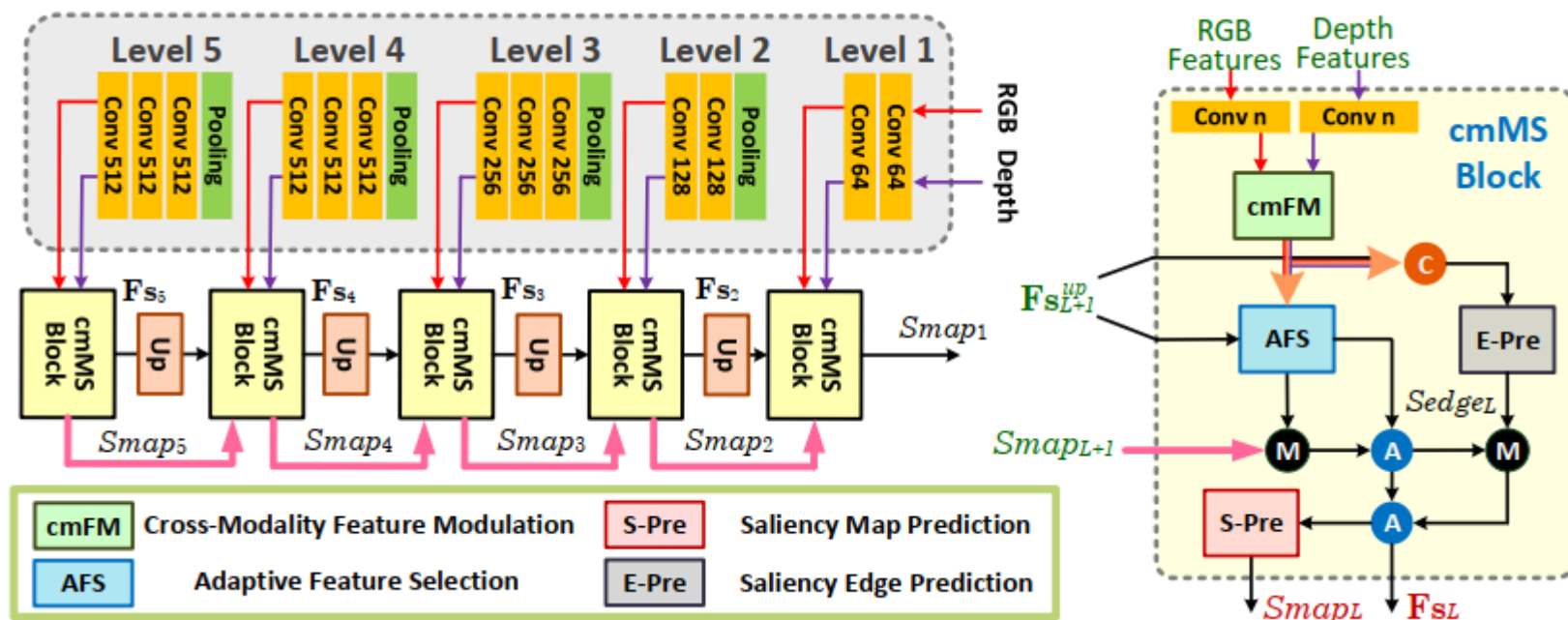
- 在不增加训练标签成本的前提下，设计了一种显著性检测任务导向的深度置信感知模块，通过网络训练自动感知深度图置信度响应，降低低质深度图对模型的影响；
- **多模态门控注意力模块**集成了跨模态互补信息，其中空间注意力机制用于降低信息冗余，门函数控制器用于调控跨模态信息融合比例；
- 无需任何预处理和后处理操作，八个数据集上均性能最佳。

方法1：深度置信感知的门控注意力网络



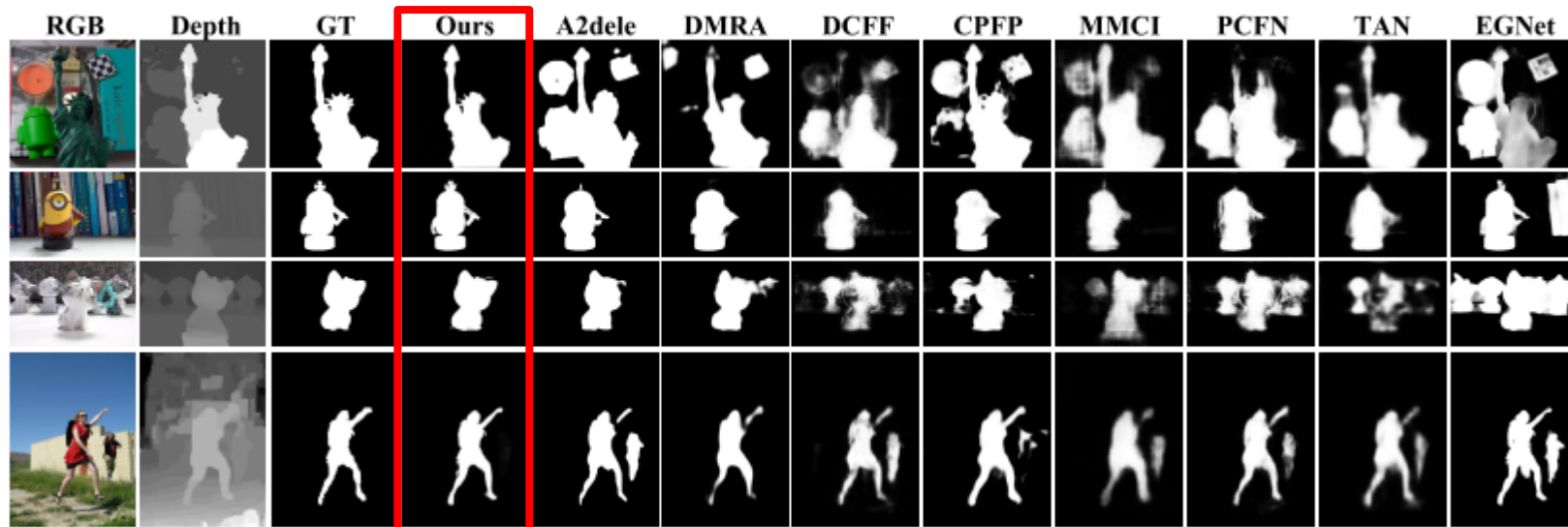
Method	Backbone	RGBD135			SSD			LFSD			NJUD-test		
		maxF ↑	S _m ↑	MAE ↓	maxF ↑	S _m ↑	MAE ↓	maxF ↑	S _m ↑	MAE ↓	maxF ↑	S _m ↑	MAE ↓
DPANet-R (ours)	ResNet-50	0.933	0.922	0.023	0.895	0.877	0.046	0.880	0.862	0.074	0.931	0.922	0.035
DPANet-V (ours)	VGG-16	0.931	0.917	0.024	0.869	0.872	0.052	0.844	0.839	0.086	0.925	0.916	0.039
AF-Net (Arxiv19)	VGG-16	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)	VGG-19	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)	VGG-16	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)	VGG-16	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)	VGG-19/16	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)	VGG-16	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)	VGG-16	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)	VGG-16	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)	GoogleNet	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)	ResNet-50	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)	ResNet-34	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)	ResNet-50	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)	VGG-16	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)	ResNet-50	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 ³ Net (IJCAI18)	ResNeXt-101	0.857	0.845	0.045	0.711	0.672	0.144	0.843	0.818	0.089	0.805	0.771	0.105

方法2：跨模态调制与选择网络



- 提出了一种RGB-D SOD方法，重点关注跨模态信息的融合与选择，可以有效地整合跨模态互补性，并自适应地选择与显著性相关的特征；
- 交叉模态特征调制模块通过将相应的深度特征作为先验来增强RGB特征表示，自适应特征选择模块在融合重要的多模态空间特征的同时，步进式的强调通道特征在单一和交叉模态中的重要性；
- 在六个RGB-D SOD基准数据集上，获得了优秀的性能。

方法2：跨模态调制与选择网络



	STEREO Dataset			NLPR-Test Dataset			NJUD-Test Dataset			LFSD Dataset			SSD Dataset			DUT-Test Dataset		
	$F_\beta \uparrow$	MAE \downarrow	$S_m \uparrow$	$F_\beta \uparrow$	MAE \downarrow	$S_m \uparrow$	$F_\beta \uparrow$	MAE \downarrow	$S_m \uparrow$	$F_\beta \uparrow$	MAE \downarrow	$S_m \uparrow$	$F_\beta \uparrow$	MAE \downarrow	$S_m \uparrow$	$F_\beta \uparrow$	MAE \downarrow	$S_m \uparrow$
PoolNet [23]	0.8757	0.0655	0.8359	0.8627	0.0448	0.8573	0.8740	0.0676	0.8600	0.8474	0.0945	0.8217	0.7644	0.1099	0.7491	0.8828	0.0669	0.8392
EGNet [42]	0.8717	0.0671	0.8363	0.8452	0.0504	0.8497	0.8667	0.0704	0.8562	0.8445	0.0871	0.8300	0.7040	0.1351	0.7072	0.8876	0.0641	0.8439
DF [33]	0.6961	0.1738	0.6279	0.6480	0.1079	0.6710	0.6355	0.1987	0.5930	0.8534	0.1424	0.7791	0.7631	0.1511	0.7422	0.7747	0.1455	0.7051
CTMF [15]	0.8265	0.1023	0.8230	0.8407	0.0561	0.8549	0.8572	0.0847	0.8493	0.8147	0.1202	0.7883	0.7550	0.1003	0.7757	0.8417	0.0971	0.8226
PCFN [3]	0.8838	0.0606	0.8722	0.8635	0.0437	0.8592	0.8875	0.0592	0.8768	0.8290	0.1118	0.7919	0.8447	0.0627	0.8427	0.8094	0.0999	0.7878
MMCI [6]	0.8610	0.0796	0.8504	0.8412	0.0591	0.8524	0.8684	0.0789	0.8588	0.8128	0.1318	0.7793	0.8230	0.0820	0.8133	0.8044	0.1125	0.7818
TAN [4]	0.8865	0.0591	0.8701	0.8765	0.0410	0.8736	0.8882	0.0605	0.8785	0.8275	0.1108	0.7935	0.8350	0.0629	0.8393	0.8236	0.0926	0.7948
CPEP [41]	0.8856	0.0537	0.8702	0.8878	0.0359	0.8760	0.7994	0.0794	0.7984	0.8495	0.0881	0.8200	0.8014	0.0818	0.8067	0.7866	0.0995	0.7335
DCFF [5]	0.8867	0.0638	0.8706	0.8779	0.0439	0.8695	0.8910	0.0646	0.8774	0.8220	0.1191	0.7917	0.8388	0.0769	0.8316	0.8141	0.1014	0.7835
DMRA [30]	0.8953	0.0474	0.8778	0.8870	0.0339	0.8646	0.9003	0.0529	0.8804	0.8723	0.0754	0.8391	0.8579	0.0583	0.8569	0.9082	0.0477	0.8637
ASIF-Net [19]	0.8939	0.0493	0.8686	0.9002	0.0298	0.8844	0.9007	0.0471	0.8887	0.8584	0.0896	0.8144	0.8633	0.0562	0.8566	0.8574	0.0725	0.8141
A2dele [31]	0.8997	0.0431	0.8713	0.8976	0.0285	0.8770	0.8939	0.0510	0.8704	0.8577	0.0740	0.8306	0.8248	0.0691	0.8093	0.9145	0.0426	0.8611
ours	0.9084	0.0422	0.8895	0.9137	0.0273	0.8999	0.9149	0.0442	0.9040	0.8882	0.0720	0.8465	0.8650	0.0524	0.8615	0.9328	0.0366	0.8853

汇报提纲

1 研究背景及意义

2 立体显著性检测

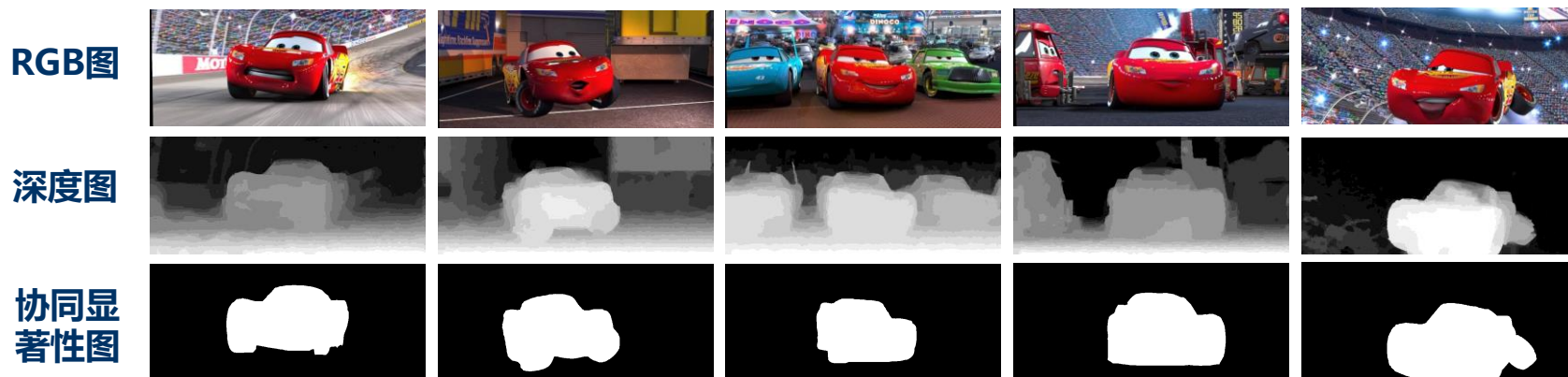
3 协同显著性检测

4 视频显著性检测

5 遥感显著性检测

6 研究总结与展望

协同显著性检测

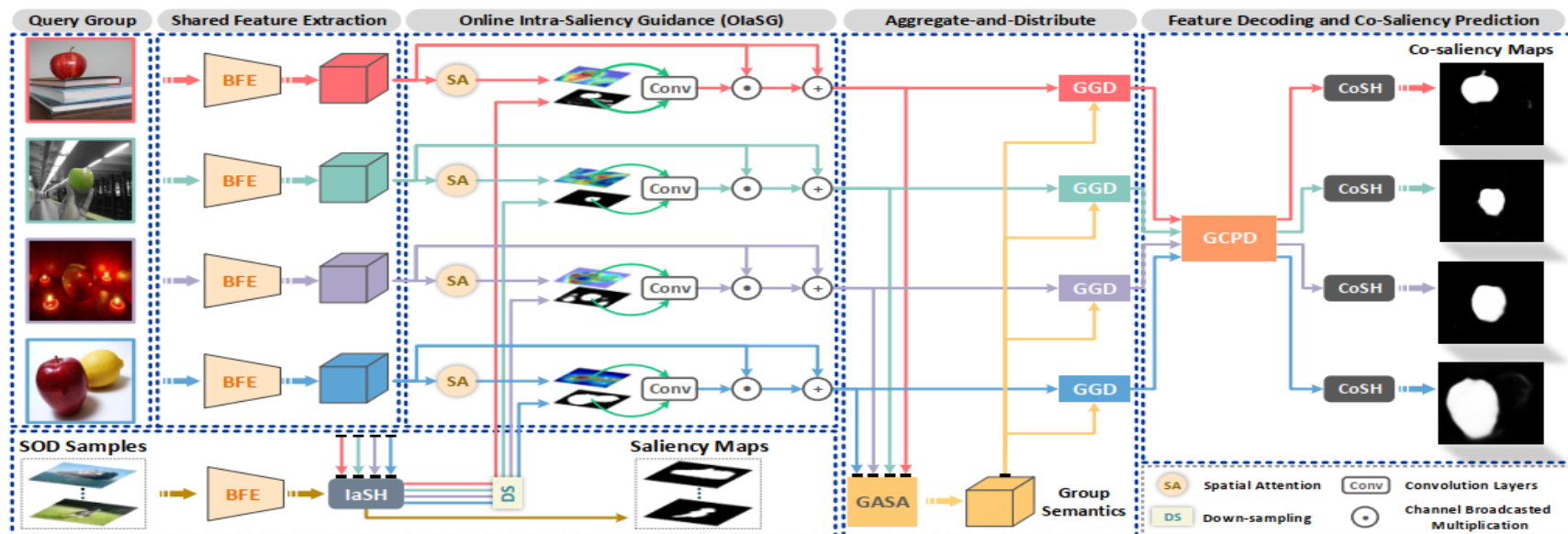


RGB/RGB-D图像协同显著性检测

关键问题

- 如何挖掘图像之间的对应关系来约束显著性目标的共有属性是一个挑战；
- 如何利用深度信息辅助显著性目标检测是一个挑战；

方法：协同集成-分发网络

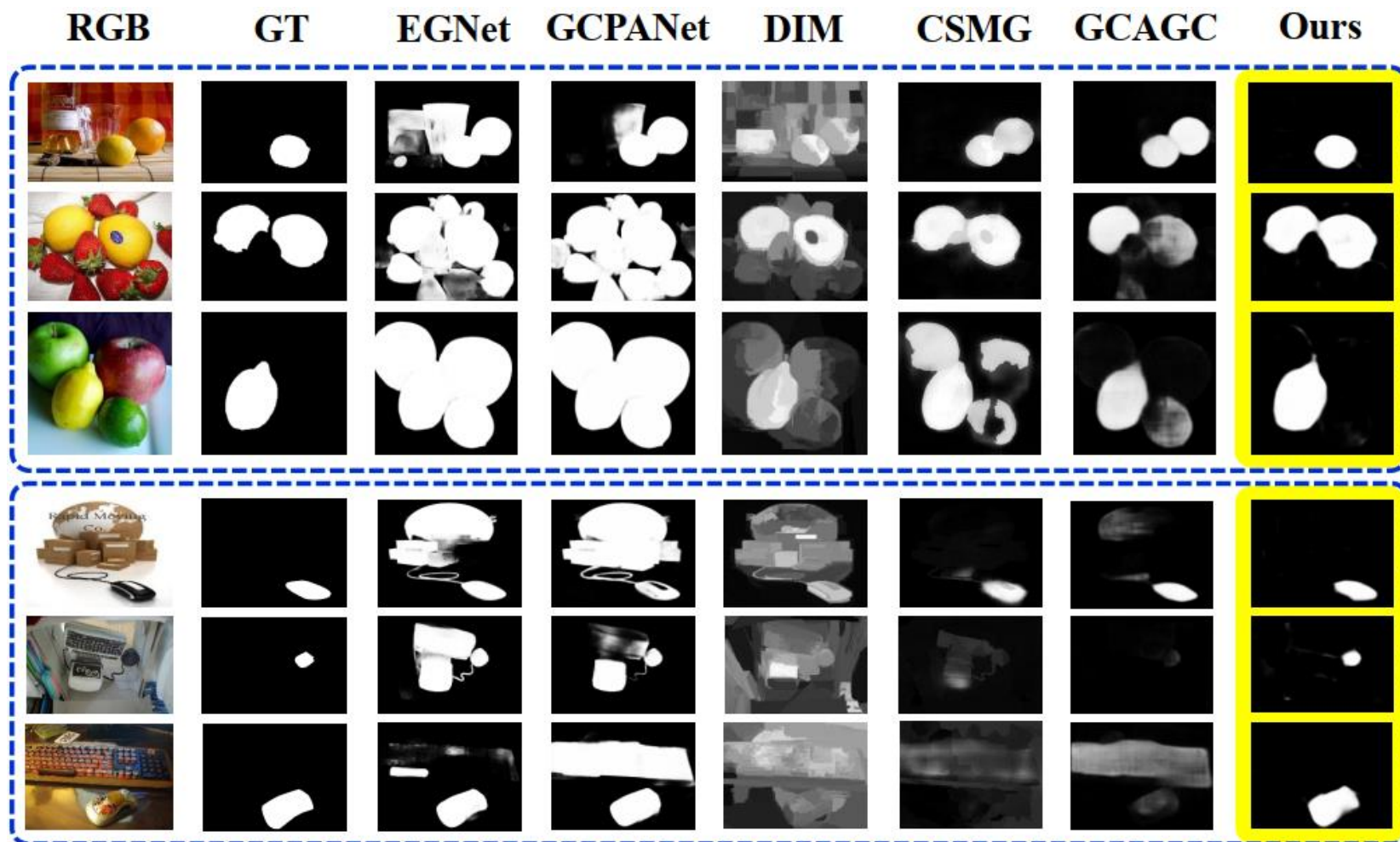


- 设计了一个**在线图内显著性引导模块**提供显著性先验知识，使网络能够动态学习如何结合显著性线索和深度个体特征；
- 提出了一个**双阶段集成-分发结构**来学习群组对应关系和协同显著性线索，其中，第一阶段利用一个**群组注意力语义集成模块**来捕获图间关系并获取长范围语义依赖性，第二阶段提出**门控群组分发模块**以一种动态的方式将群组语义信息分发至每个单一个体；
- 为了在特征解码过程中保持图间关系，提出了一种**群组一致性保持的解码器结构**，可以生成保持群组一致性的全分辨率协同显著性图。

方法：协同集成-分发网络

	Cosal2015 Dataset			CoSOD3k Dataset			MSRC Dataset			iCoseg Dataset		
	$F_\beta \uparrow$	MAE \downarrow	$S_m \uparrow$	$F_\beta \uparrow$	MAE \downarrow	$S_m \uparrow$	$F_\beta \uparrow$	MAE \downarrow	$S_m \uparrow$	$F_\beta \uparrow$	MAE \downarrow	$S_m \uparrow$
CPD [38]	0.8228	0.0976	0.8168	0.7661	0.1068	0.7788	0.8250	0.1714	0.7184	0.8768	0.0579	0.8565
EGNet [46]	0.8281	0.0987	0.8206	0.7692	0.1061	0.7844	0.8101	0.1848	0.7056	0.8880	0.0601	0.8694
GCPANet [5]	0.8557	0.0813	0.8504	0.7808	0.1035	0.7954	0.8133	0.1487	0.7575	0.8924	0.0468	0.8811
UMLF [14]	0.7298	0.2691	0.6649	0.6895	0.2774	0.6414	0.8605	0.1815	0.8007	0.7623	0.2389	0.6828
CODW [42]	0.7252	0.2741	0.6501	–	–	–	0.8020	0.2645	0.7152	0.8271	0.1782	0.7510
DIM [19]	0.6363	0.3126	0.5943	0.5603	0.3267	0.5615	0.7419	0.3101	0.6579	0.8273	0.1739	0.7594
GoNet [17]	0.7818	0.1593	0.7543	–	–	–	0.8598	0.1779	0.7981	0.8653	0.1182	0.8221
CSMG [43]	0.8340	0.1309	0.7757	0.7641	0.1478	0.7272	0.8609	0.1892	0.7257	0.8660	0.1050	0.8122
RCGS [32]	0.8245	0.1004	0.7958	–	–	–	0.7692	0.2134	0.6717	0.8005	0.0976	0.7860
GCAGC [44]	0.8666	0.0791	0.8433	0.8066	0.0916	0.7983	0.7903	0.2072	0.6768	0.8823	0.0773	0.8606
CoADNet-R	0.8771	0.0609	0.8672	0.8204	0.0643	0.8402	0.8710	0.1094	0.8269	0.8997	0.0411	0.8863
CoADNet-DR	0.8874	0.0599	0.8705	0.8308	0.0652	0.8416	0.8618	0.1323	0.8103	0.9225	0.0438	0.8942

方法：协同集成-分发网络



汇报提纲

1 研究背景及意义

2 立体显著性检测

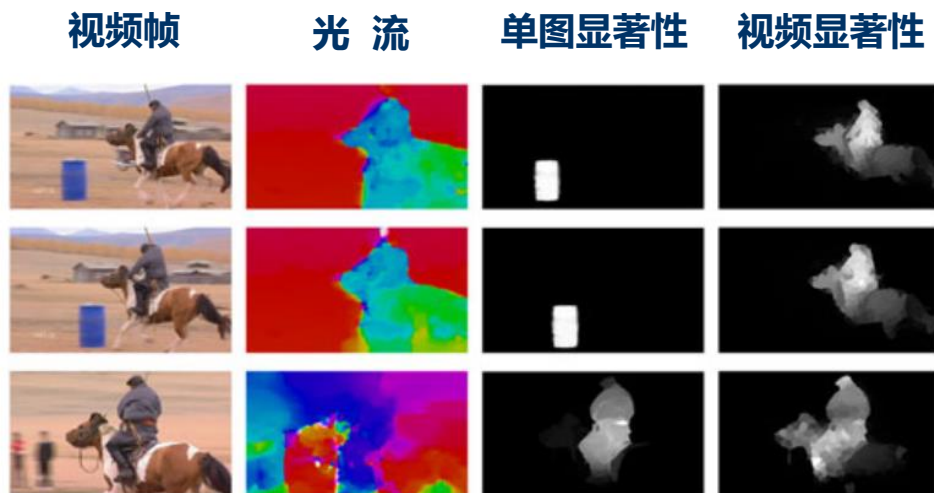
3 协同显著性检测

4 视频显著性检测

5 遥感显著性检测

6 研究总结与展望

视频显著性检测



定 义

连续检测视频序列中**反复出现**的**与运动相关**的显著性目标。

“连续” 强调空时一致性；

“反复出现” 强调时序关系；

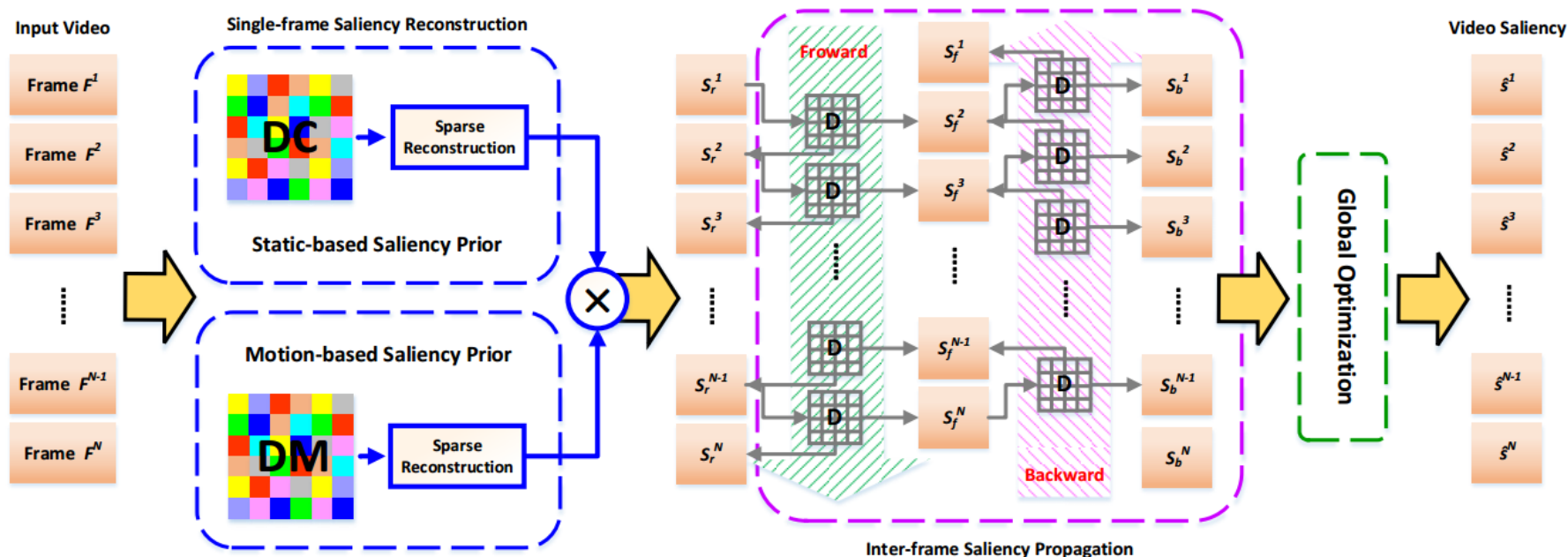
“与运动相关” 强调运动属性；

“显著性目标” 强调单帧显著性；

关键问题

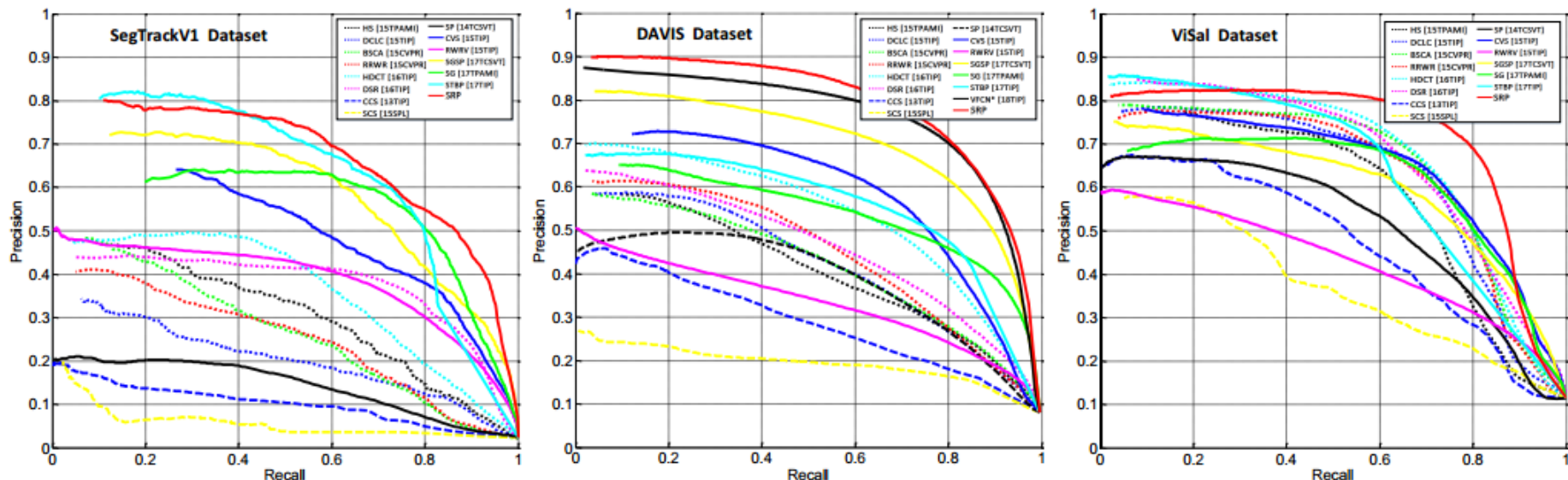
- 运动信息在视频显著性目标检测任务中起到至关重要的作用；
- 帧间关系用于捕捉整个视频序列中显著性目标的共有属性；
- 空时一致性约束视频中的显著性目标的连续性和一致性；

方法：稀疏重建与传播

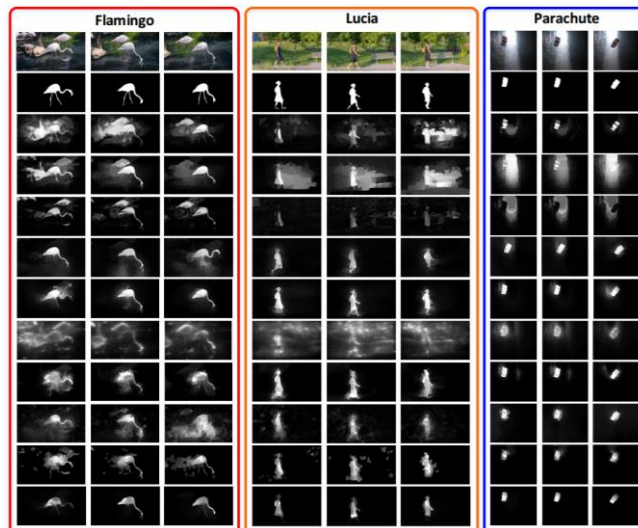


- **基于稀疏的显著性重建：**充分利用静态先验和运动先验生成单帧显著性图。运动先验定义为**运动紧致性线索**和**运动唯一性线索**；
- **基于稀疏的显著性传播：**通过**前向-后向传播策略**捕获时间空间中的对应关系并生成帧间显著性图；
- 为了改善整个视频中显著对象的全局一致性，提出了包含数据项、时空平滑项、空间互斥项和全局一致项的**全局优化模型**。

方法：稀疏重建与传播



	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



汇报提纲

1 研究背景及意义

2 立体显著性检测

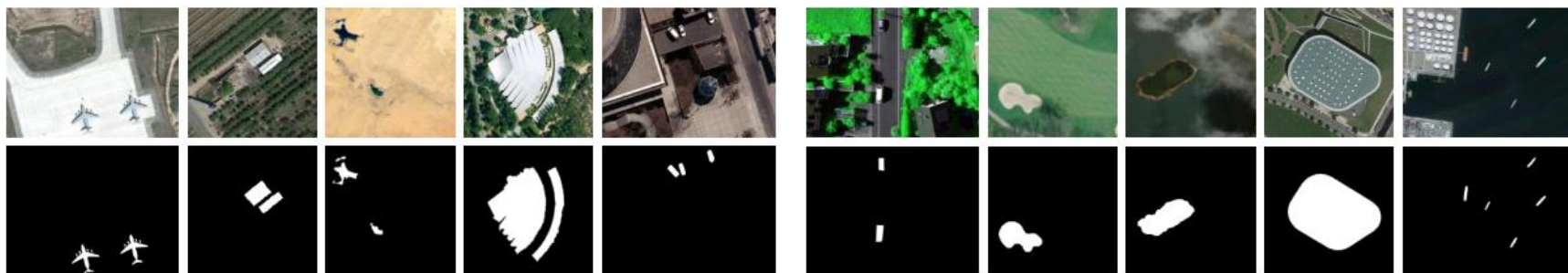
3 协同显著性检测

4 视频显著性检测

5 遥感显著性检测

6 研究总结与展望

遥感显著性检测



光学遥感图像显著性目标检测

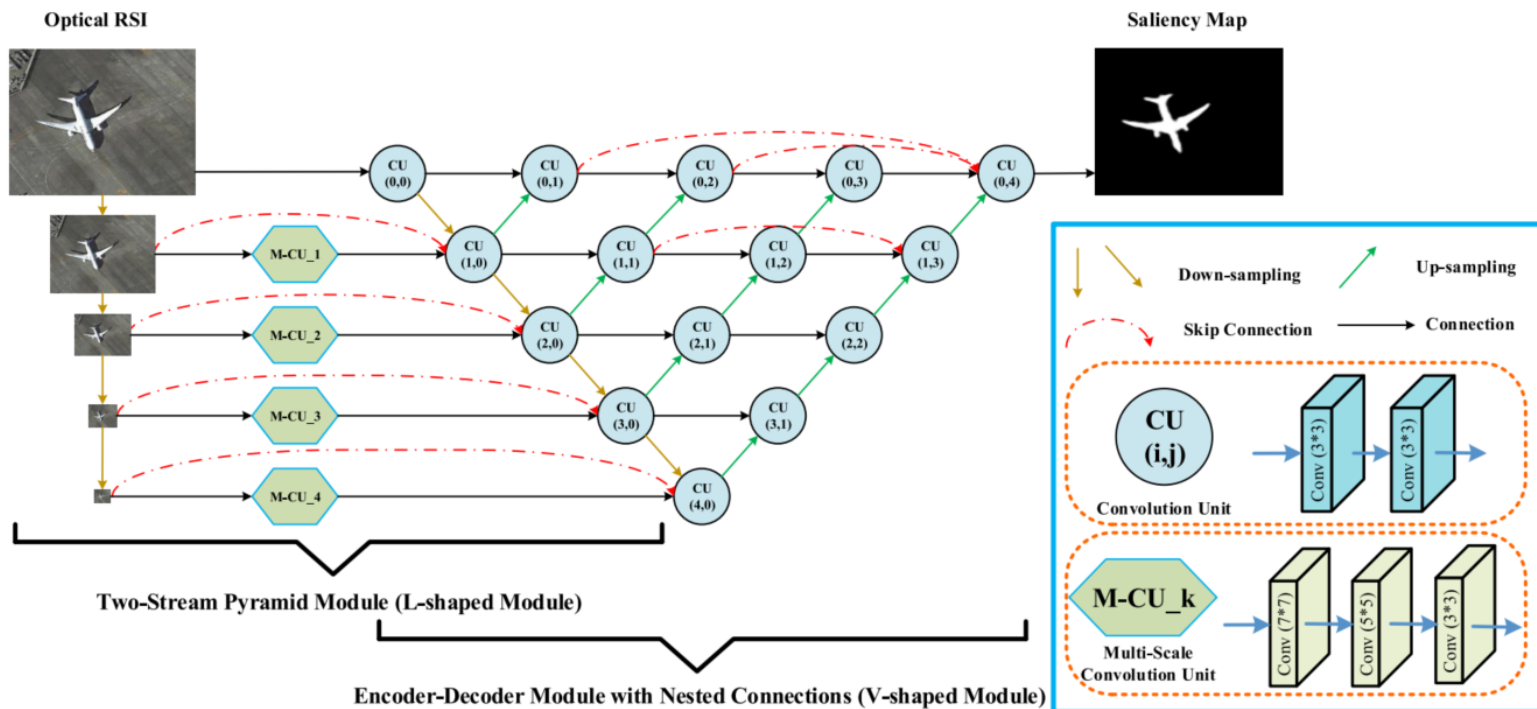
关键问题1

- 光学遥感图像中显著性目标类型多样、尺寸变化不一，这都极大地增加了任务的难度；

关键问题2

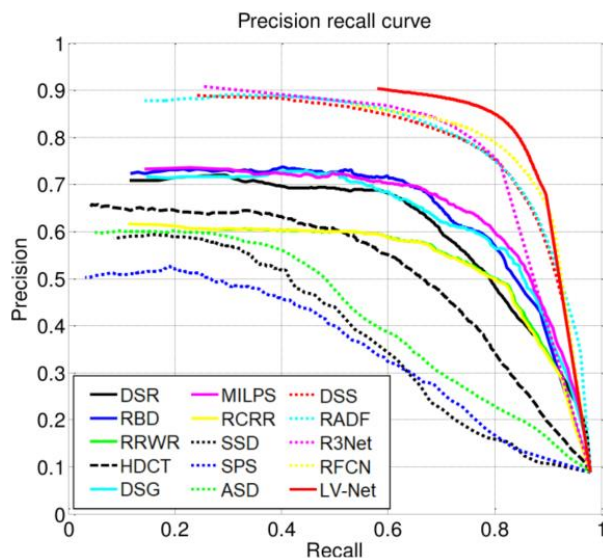
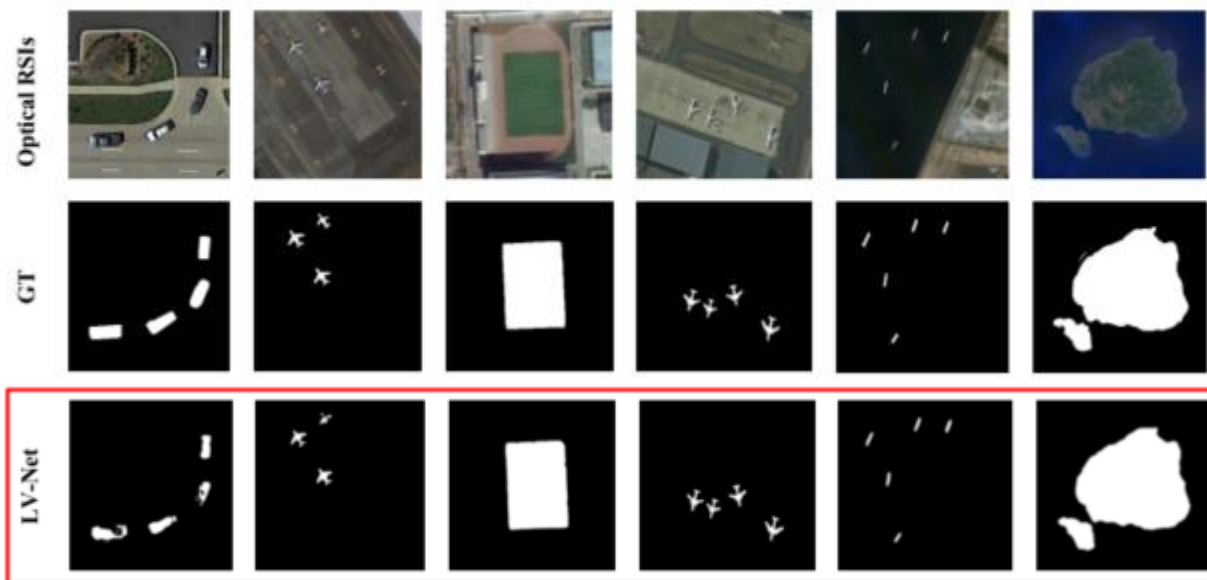
- 光学遥感图像中场景范围广、背景复杂、存在因成像模式导致的阴影噪声、云层遮挡等问题；

方法：双流金字塔嵌套网络



- **首个**基于深度学习的端到端的光学遥感图像显著性目标检测算法；
- **双流金字塔模块**学习一组互补特征，以解决显著性目标的尺度变化问题并捕获局部细节，**具有嵌套连接的编码器-解码器模块**自动确定有判别力特征，以抑制杂乱的背景并突出显著性目标；
- 构造了**首个**用于显著性目标检测的具有挑战性的**光学遥感数据集**。

方法：双流金字塔嵌套网络



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

汇报提纲

1 研究背景及意义

2 立体显著性检测

3 协同显著性检测

4 视频显著性检测

5 遥感显著性检测

6 研究总结与展望

研究总结与展望

研究总结

不同类型数据之间存在**共性和个性**、**区别与联系**，不同数据模式带来了新问题、新挑战，探索多源数据驱动的视觉注意力感知与计算任务期望在提供个性化解决方案的同时推出更为通用的数据建模方法，打通数据壁垒，实现相关技术模块化，进而构建**泛化能力更强**、**鲁棒性更好**的通用模型。

未来展望

- 新方法的尝试：小样本学习、弱/自监督学习、跨域学习、零样本学习等；
- 新数据的扩展：光场、RGB+X、360°图像/视频、点云、航空遥感等；
- 新思路的探索：实例级显著性目标检测与分割、显著性增强与修正等；

感谢



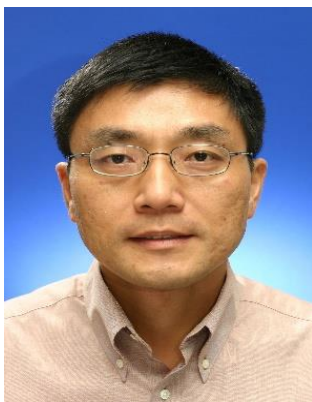
赵耀教授
@BJTU



黄庆明教授
@UCAS



雷建军教授
@TJU



Prof. Weisi Lin
@NTU



Prof. Sam Kwong
@CityU



Prof. Fatih Porikli
@ANU

附录

- ① **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.
- ② **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, DOI: 10.1109/TCYB.2019.2932005, 2019.
- ③ **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, vol. 28, no. 10, pp. 4819-4831, 2019.
- ④ **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.
- ⑤ **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.

附录

- ⑥ **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, vol. 29, no. 10, pp. 2941-2959, 2019.
- ⑦ **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.
- ⑧ 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, vol. 57, no. 11, pp. 9156-9166, 2019. (**# co-first and corresponding author**)
- ⑨ Chongyi Li, **Runmin Cong#**, Sam Kwong, Junhui Hou, Huazhu Fu, Guopu Zhu, Dingwen Zhang, and Qingming Huang, ASIF-Net: Attention steered interweave fusion network for RGBD salient object detection, IEEE Transactions on Cybernetics, 2020. In Press. (**# co-first and corresponding author**)

附录

- ⑩ **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, ISBN 978-3-030-28602-6, Dec. 2019.
(**Book Chapter**)
- ⑪ 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)
- ⑫ **丛润民**, 雷建军, 付华柱, 王文冠, 黄庆明, 牛力杰, 视频显著性检测研究进展, 软件学报, 29(8): 2527–2544, 2018. (CCF A, EI, **第十五届北京青年优秀科技论文奖**)

-
- Zuyao Chen, **Runmin Cong**, Qianqian Xu, and Qingming Huang, DPANet: Depth potentiality-aware gated Attention network for RGB-D salient object detection, under review, 2020.
 - Qijian Zhang, **Runmin Cong**, Junhui Hou, Chongyi Li, and Yao Zhao, CoADNet: Collaborative aggregation-and-distribution networks for co-salient object detection, under review, 2020.



信息科学研究所
Institute of Information Science



谢谢!

邮箱: rmcong@bjtu.edu.cn

主页: <https://rmcong.github.io/>