Dual-stream Network Based on Global Guidance for Salient Object Detection

ID: 1411

Shuyong Gao[1], Qianyu Guo[1], Wei Zhang[1], Wenqiang Zhang[1] , Zhongwei Ji[2]
[1]Department of Computer Science and Technology, Fudan University, Shanghai, China
[2]ZTE Corporation, Shanghai, China

ICASSP 2021 TORONTO Canada June 6-11, 2021 Metro Toronto Convention Centre

Objectives

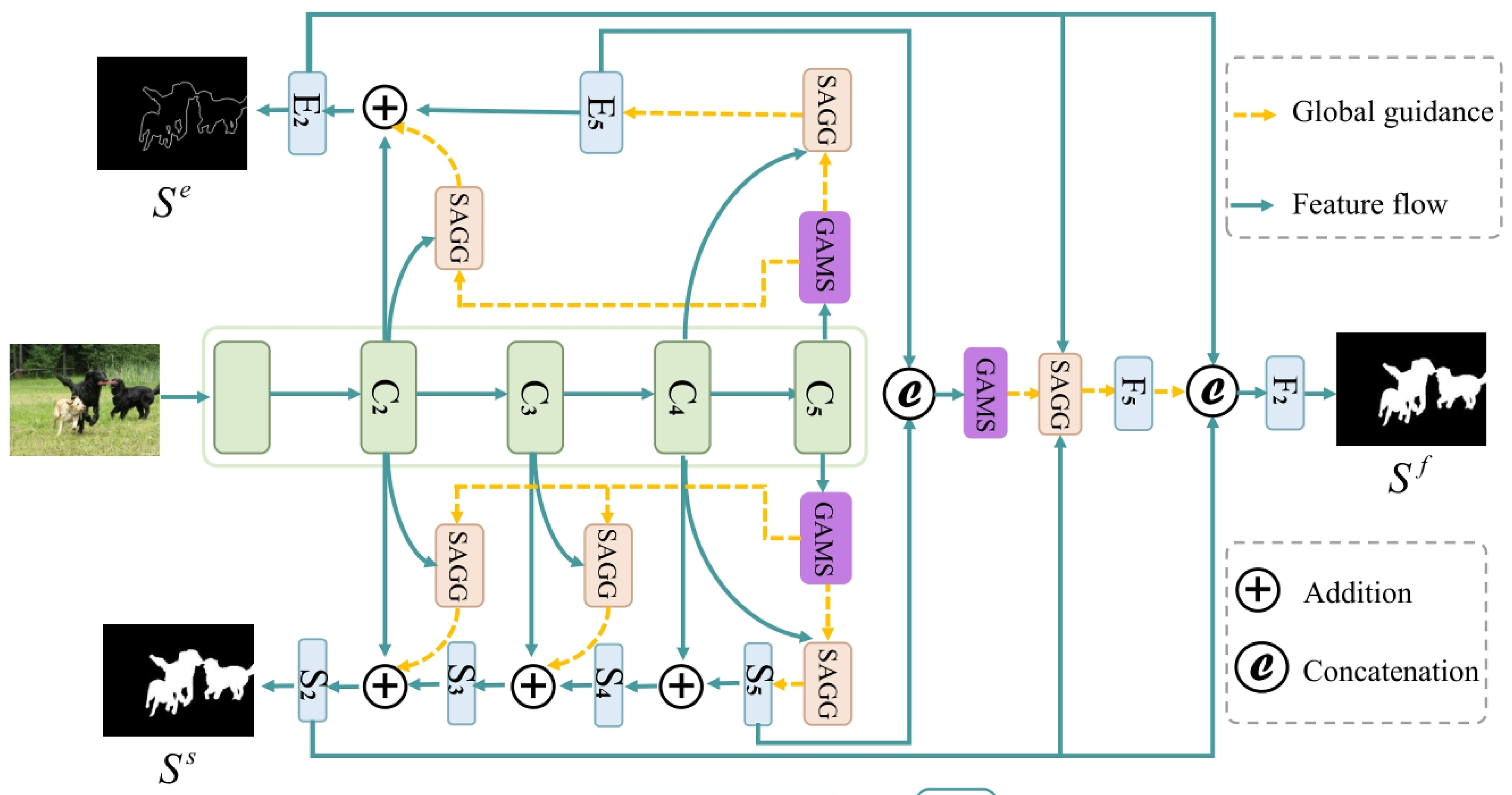
Salient object detection (SOD) aims to detect the visually distinguishable regions of the images or videos to imitate human attention mechanism , which can be widely used in various fields of images processing.

Modivation

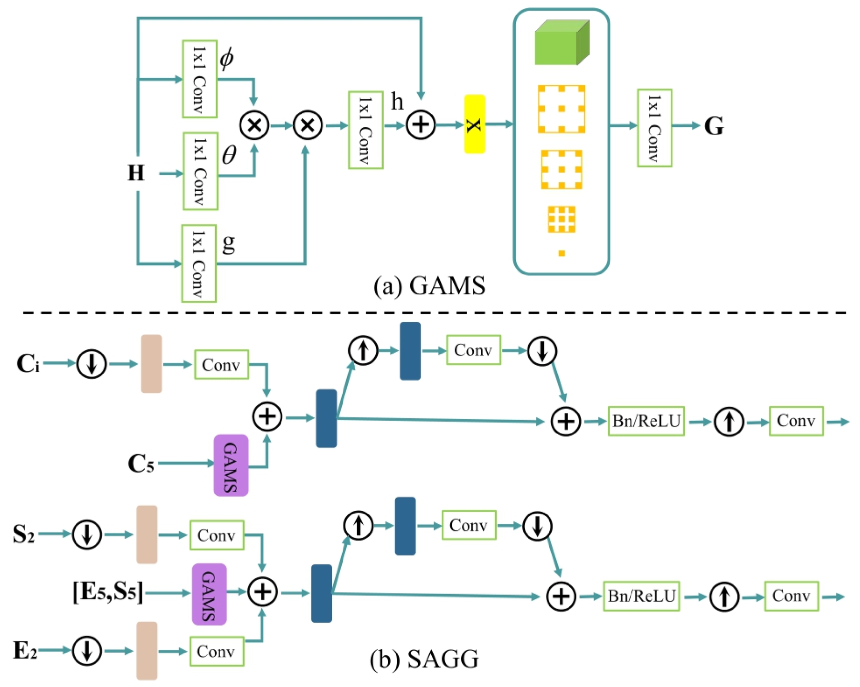
Some methods use high-level features to provide global guidance for some layers of the network. However, there remain several problems: (1) the global guidance has not been fully mined, which leads to its limited capacity; (2) the semantic gap between global guidance and low-level features is ignored, and simple merging methods will cause feature aliasing. So we propose GAMS to mine global guidance and SAGG to integrate the global guidance into each decoding layer.

Method

Network Architecture:



Module Architecture:



Experiments

Compare with SOTA methods:

Method	ECSSD F_β^{max}	ECSSD mF	ECSSD S_m	ECSSD MAE	DUT-OMRON F_β^{max}	DUT-OMRON mF	DUT-OMRON S_m	DUT-OMRON MAE	PASCAL-S F_β^{max}	PASCAL-S mF	PASCAL-S S_m	PASCAL-S MAE	HKU-IS F_β^{max}	HKU-IS mF	HKU-IS S_m	HKU-IS MAE
R^3Net_{18}[13]	0.9248	0.9027	0.9028	0.0555	0.7882	0.7532	0.8172	0.0711	0.8374	0.8155	0.8102	0.1026	0.9096	0.8807	0.8918	0.0478
RAS_{18}[14]	0.9211	0.9006	0.8929	0.0564	0.7863	0.7623	0.8141	0.0617	0.8291	0.8125	0.799	0.1013	0.9128	0.8877	0.8874	0.0454
$DGRL_{18}$[15]	0.9224	0.9130	0.9027	0.0407	0.7742	0.7659	0.8059	0.0618	0.8486	0.8355	0.8358	0.0721	0.9103	0.8998	0.8946	0.0356
PiCANet-R_{18}[16]	0.9349	0.9023	0.9168	0.0464	0.8029	0.7630	0.8301	0.0653	0.8573	0.8228	0.8537	0.0756	0.9185	0.8808	0.9041	0.0433
BASNet-R_{19}[9]	0.9425	0.9274	0.9163	0.037	0.8053	0.7906	0.8362	0.0565	0.8539	0.8344	0.838	0.0758	0.9284	0.9113	0.909	0.0322
PoolNet-R_{19}[7]	0.9415	0.9197	0.9173	0.0417	0.8058	0.7822	0.8322	0.0561	0.8648	0.8480	0.8518	0.0716	0.9305	0.907	0.907	0.033
MLMS-R_{19}[17]	0.9284	0.9007	0.9111	0.0445	0.7741	0.7455	0.809	0.0636	0.8552	0.8254	0.8442	0.0736	0.9207	0.8891	0.9065	0.0387
AFNet-R_{19}[5]	0.9351	0.9157	0.9135	0.0416	0.7972	0.7766	0.8253	0.0574	0.8629	0.8409	0.8494	0.0700	0.9226	0.8998	0.9051	0.0358
EGNet-R_{19}[5]	0.9474	0.9288	0.9245	0.0374	0.8155	0.7942	0.8379	0.0529	0.8653	0.8437	0.8519	0.0740	0.9352	0.9122	0.9178	0.0310
GateNet-R_{20}[6]	0.9454	0.9197	0.9197	0.0401	0.8181	0.7915	0.8374	0.0549	0.875	0.8518	0.857	0.0676	0.9335	0.9097	0.915	0.0331
MINet-R_{20}[18]	0.9475	0.925	0.925	0.0335	0.8099	0.7893	0.8325	0.0555	0.8726	0.852	0.8558	0.0635	0.9349	0.9166	0.9189	0.0285
GCPANet-R_{20}[10]	0.9485	0.9261	0.9267	0.0348	0.8118	0.7879	0.8375	0.0563	0.8752	0.8508	0.864	0.0619	0.9380	0.911	0.9202	0.0309
F^3Net-R_{20}[19]	0.9453	0.9242	0.9242	0.0333	0.8133	0.7944	0.8381	0.0526	0.8776	0.8588	0.86	0.0616	0.9366	0.9187	0.9171	0.028
Ours	0.9498	0.9350	0.9259	0.0335	0.819	0.7993	0.8405	0.0544	0.8841	0.8636	0.8648	0.061	0.9406	0.923	0.9236	0.0276

Visual Comparison:

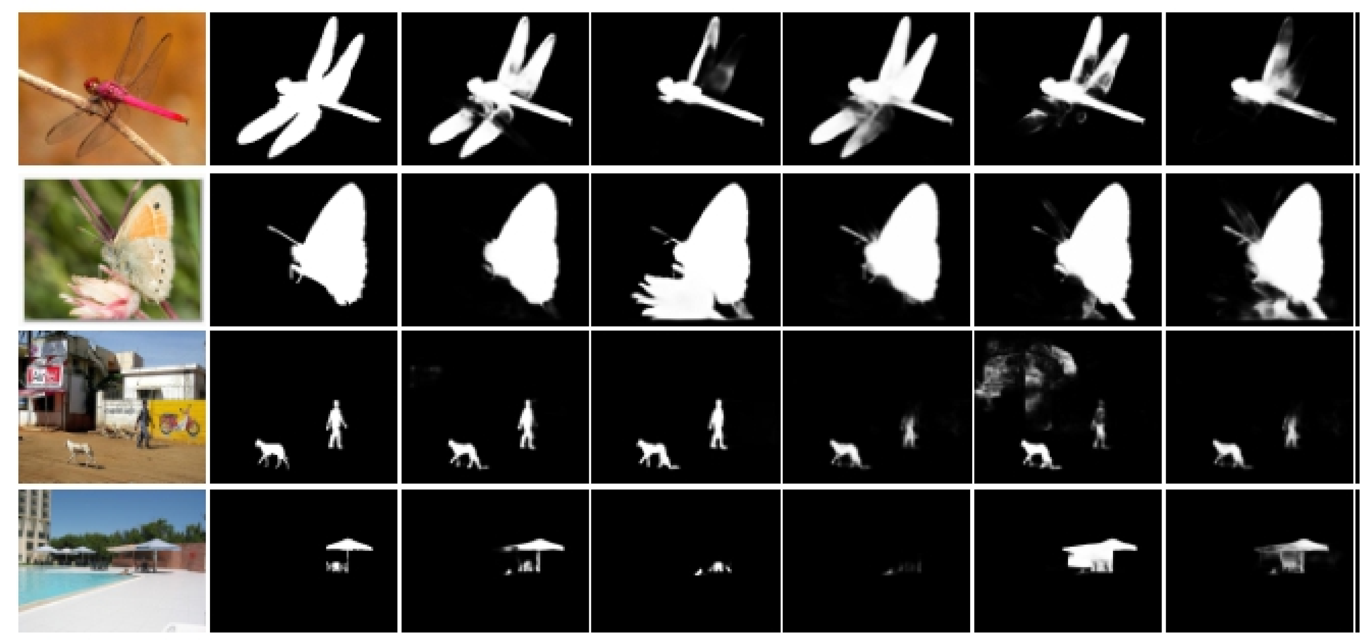


Image Label Ours F³Net GCPANet MINet GateNet

Ablation Studies:

Model	DUTS-test F_β^{max}	DUTS-test mF	DUTS-test S_m	DUTS-test MAE	DUT-OMRON F_β^{max}	DUT-OMRON mF	DUT-OMRON S_m	DUT-OMRON MAE
baseline	0.8824	0.8538	0.8839	0.0387	0.8039	0.7787	0.8286	0.0607
+guide	0.8887	0.8618	0.8871	0.0363	0.8062	0.7832	0.8319	0.0561
guide+nonlocal	0.881	0.8507	0.8832	0.0408	0.8093	0.7783	0.8288	0.0653
guide+ASPP	0.8879	0.8657	0.8893	0.0367	0.8102	0.791	0.8326	0.0586
guide+GAMS	0.8896	0.7938	0.891	0.0355	0.8164	0.7938	0.8374	0.0559
guide+SAGG	0.8918	0.8661	0.8896	0.0362	0.8149	0.7914	0.8355	0.0558
guide+SAGG+GAMS	**0.8955**	**0.873**	**0.8944**	**0.0341**	**0.819**	**0.7993**	**0.8405**	**0.0544**

Conclusion

A dual-stream network based on global guidance for salient object detection is proposed, in which we design the GAMS to provide accurate multi-scale global guidance and SAGG to seamlessly integrate the guidance into each decoding layer of the network.