16TH EUROPEAN CONFERENCE ON

COMPUTER VISION

WWW.ECCV2020.EU

Suppress and Balance: A Simple Gated Network for Salient Object Detection

Xiaoqi Zhao[1], Youwei Pang[1], Lihe Zhang[1], Huchuan Lu[1,2] and Lei Zhang[3,4]*

[1]Dalian University of Technology, China

[2]Peng Cheng Laboratory

[3]The Hong Kong Polytechnic University, China

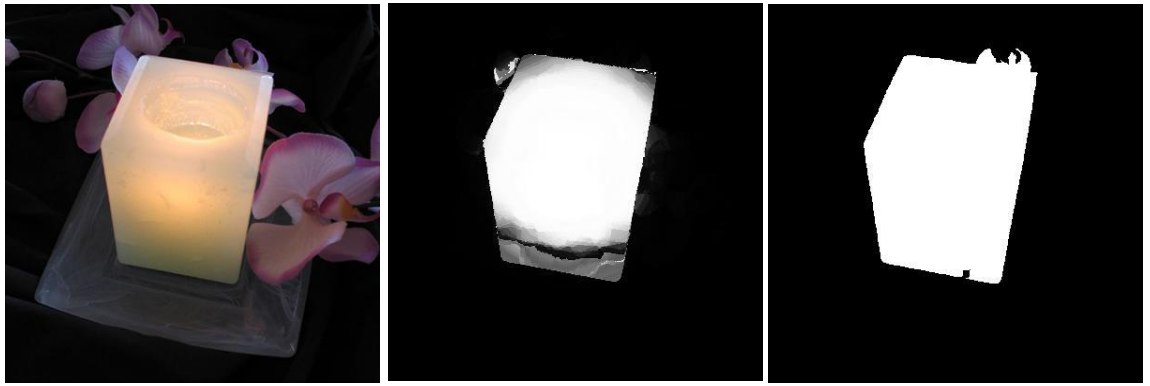
[4]DAMO Academy, Alibaba Group

IIAU-LAB

Introduction

What is Salient Object Detection (SOD)?

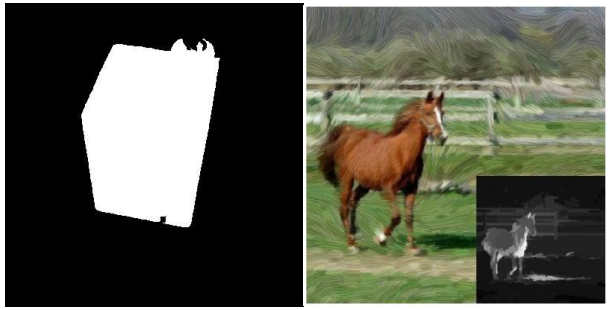
- Identify the visually distinctive regions or objects.
- Accurate segmentation.
- SOD can segment any kind of objects.



Introduction

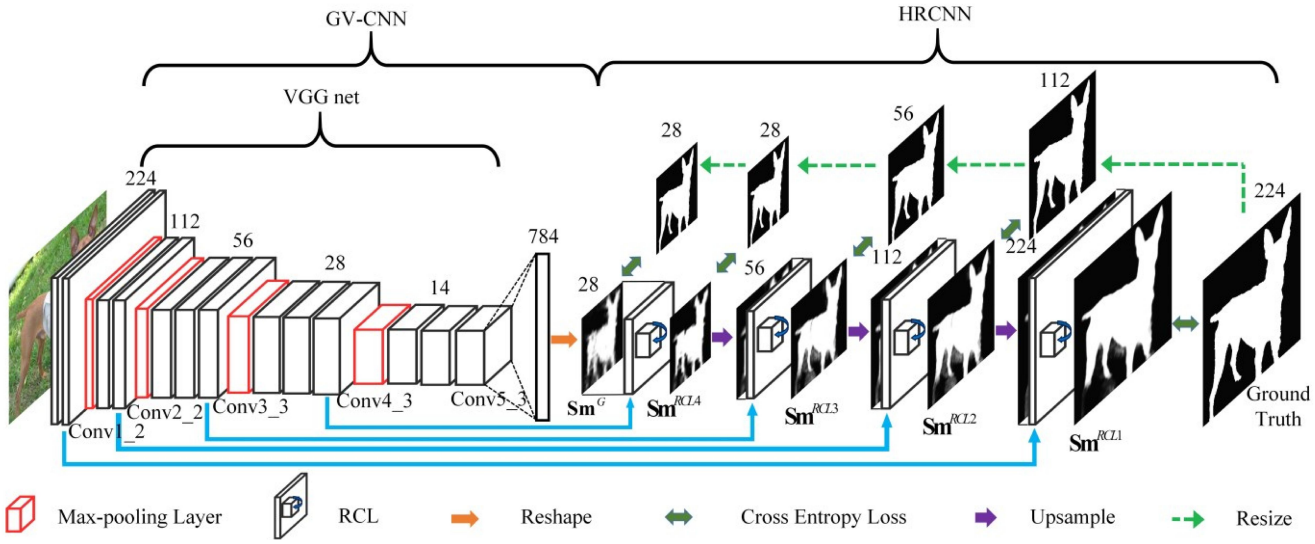
Applications

Portrait Segmention、Non-Photorealistic Rendering、FotoMix、Bokeh...

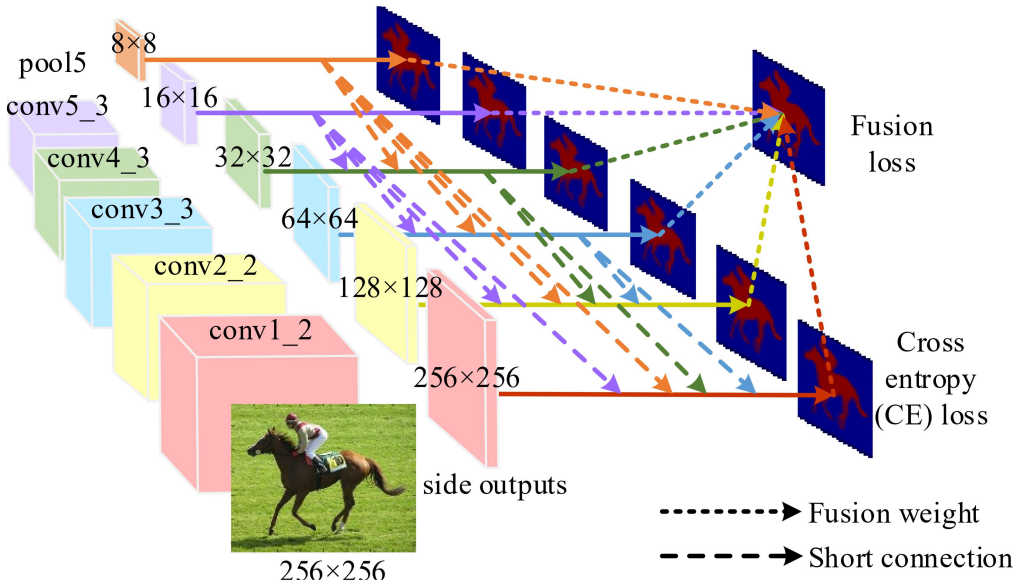


Existing Works

DHSNet(Feature Pyramid Network(FPN)+RCL+Deep Supervision)



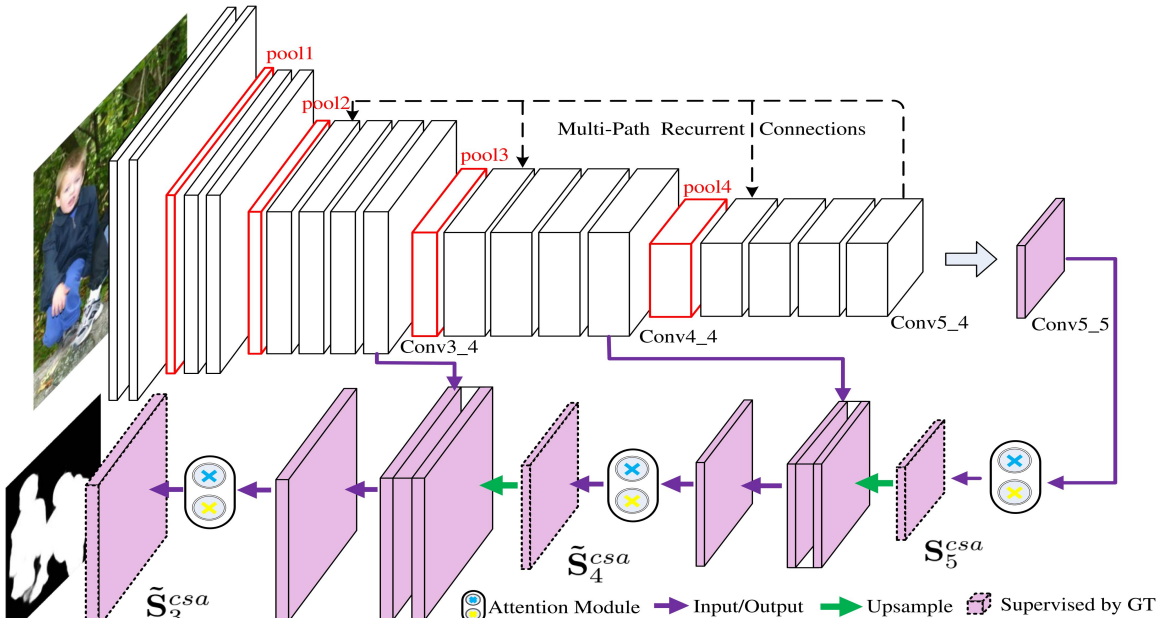
DSS(FPN+Short Connections+Supervision)



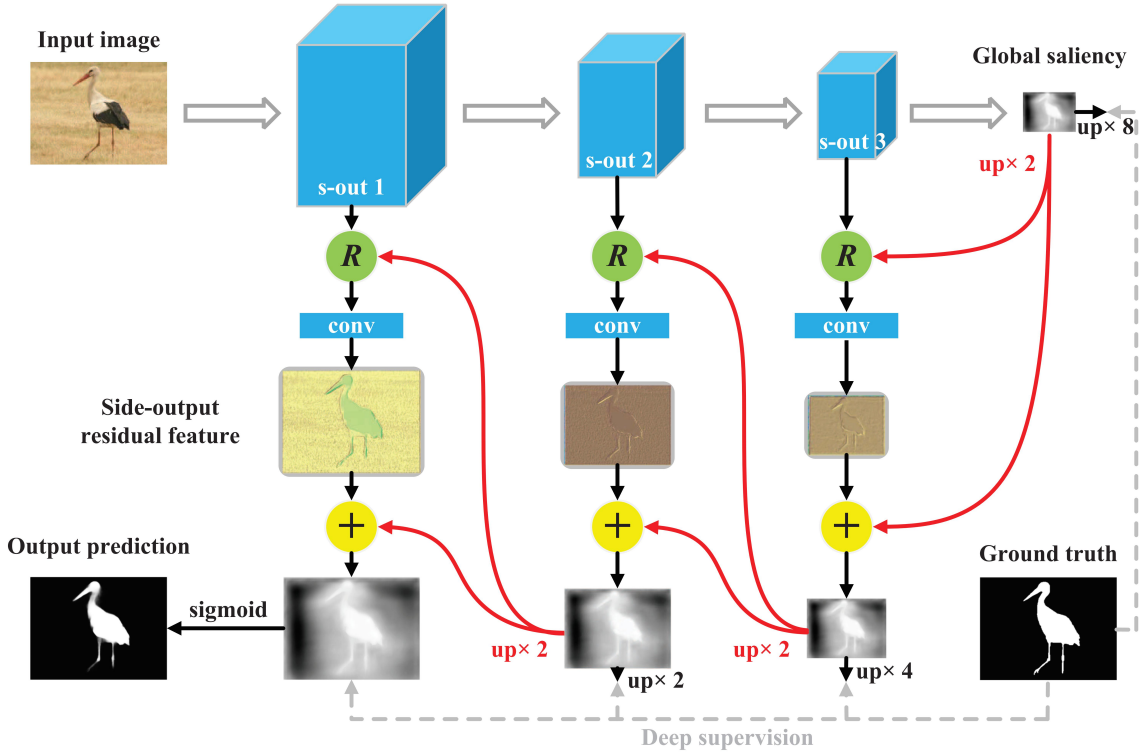
Hou et al., Deeply Supervised Salient Object Detection with Short Connections. CVPR 2017.
Nian Liu, Junwei Han. DHSNet: Deep Hierarchical Saliency Network for Salient Object Detection. CVPR 2016.

Existing Works

PAGRN(FPN+Spatial&Channel attention+Recurrent)



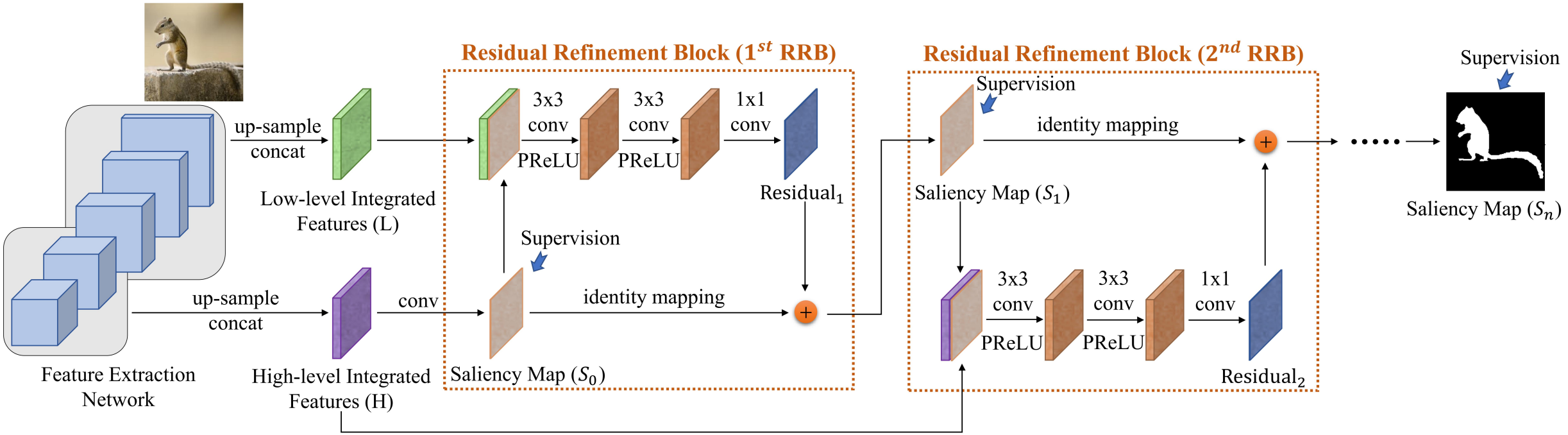
RAS(FPN+Reverse Attention+Residual learning+Deep Supervision)



Chen et al., Reverse Attention for Salient Object Detection. ECCV 2018.
Zhang et al., Progressive Attention Guided Recurrent Network for Salient Object Detection. CVPR 2018.

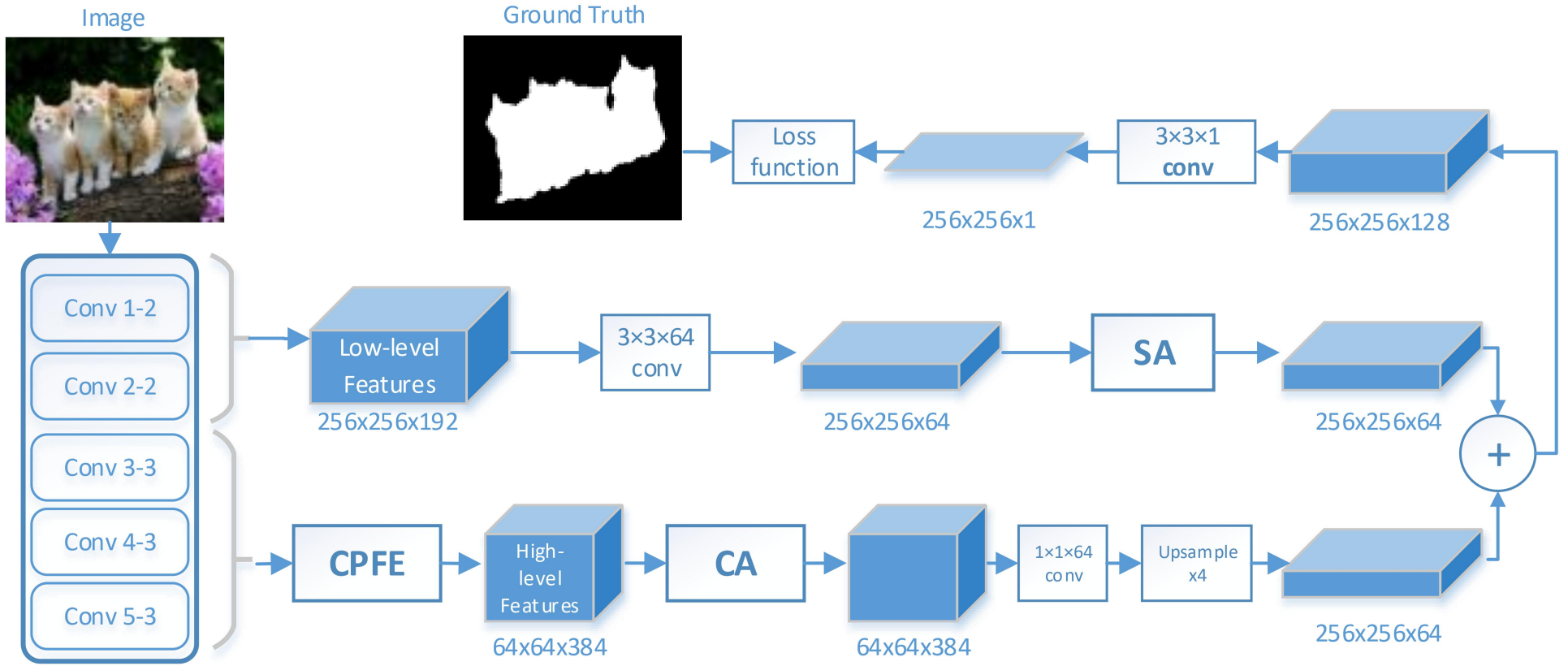
Existing Works

R³Net(Parallel Aggregration+Residual refinement+recurrent)



Existing Works

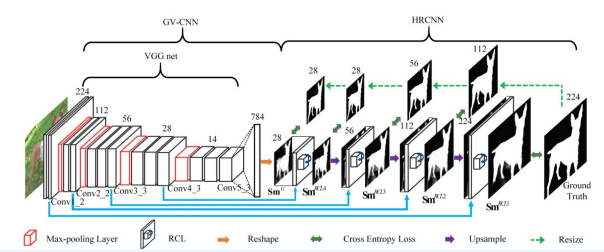
PFA(Parallel Aggregation+ ASPP-like+C&S-attention)

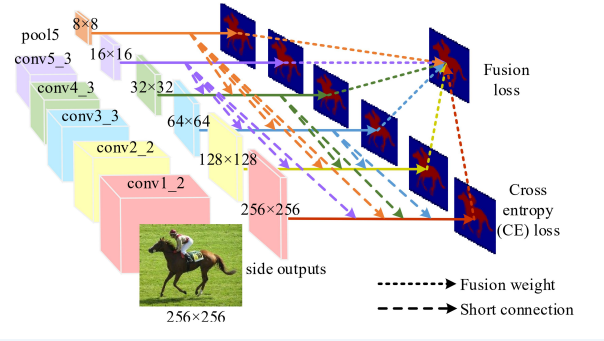


Ting Zhao, Xiangqian Wu. Pyramid Feature Attention Network for Saliency detection. CVPR2019.

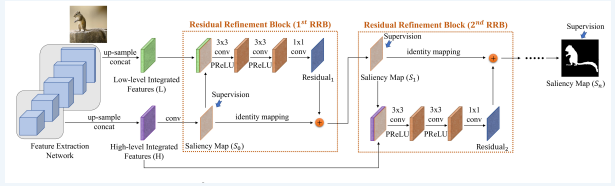
Existing Works

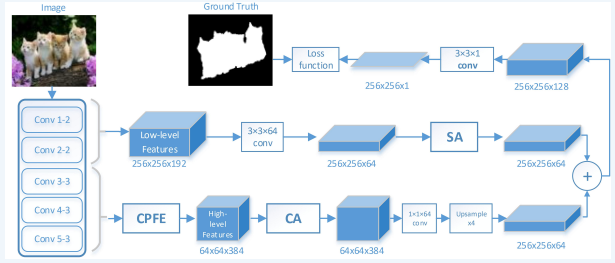
Progressive



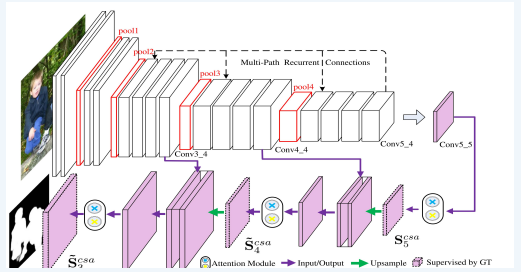


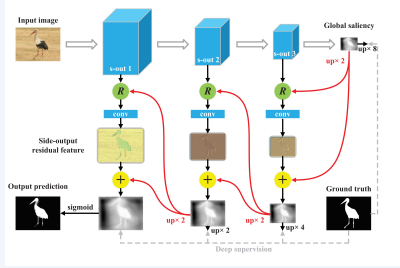
Parallel



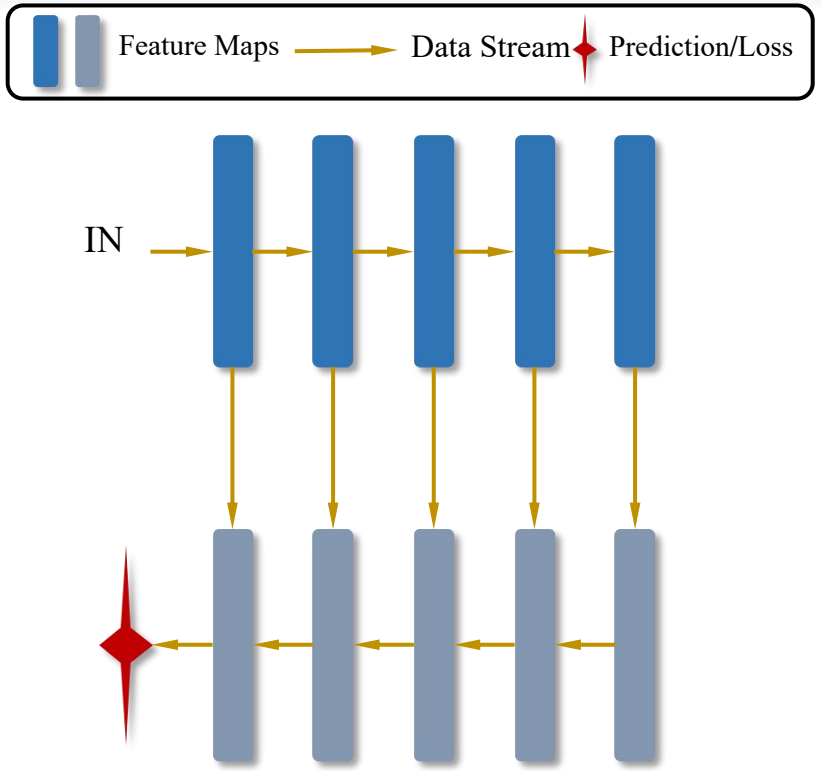


Attention





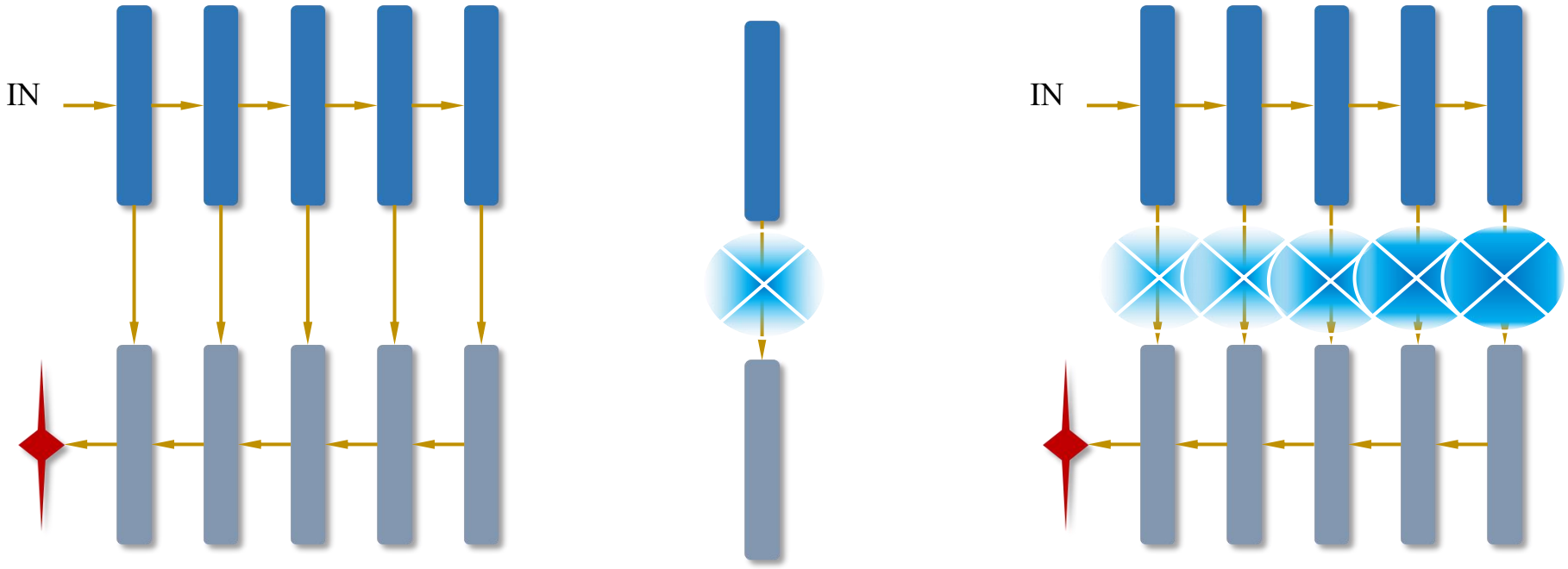
Drawbacks



DHSNet/DSS/Amulet/PAGRN/PASE/...

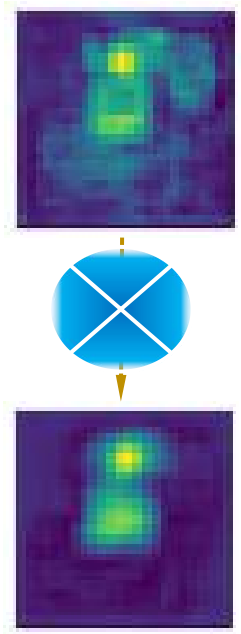
Encoder->Decoder:
Suppress and Balance

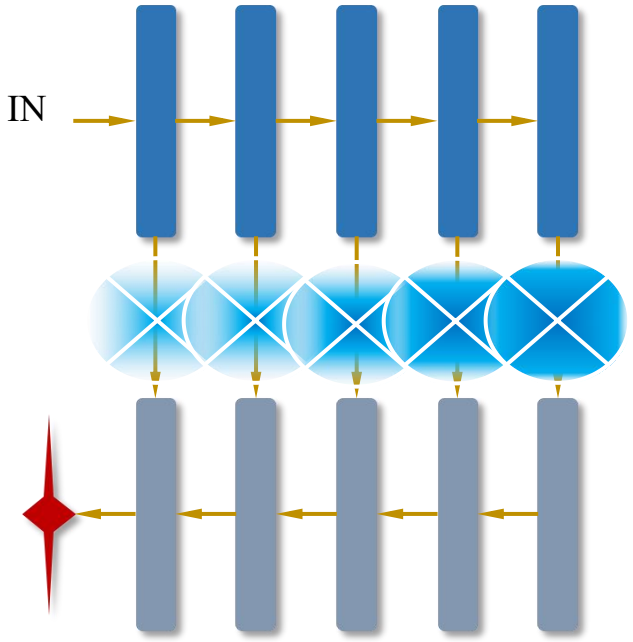
Solution

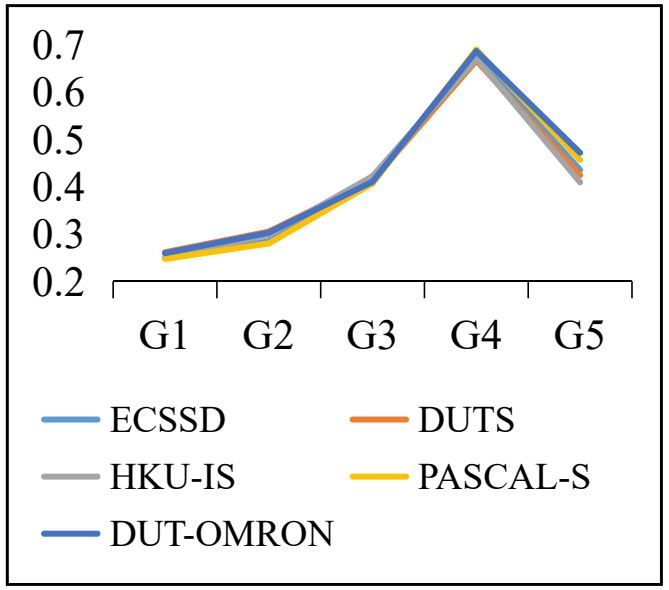


Our Works

Gated FPN

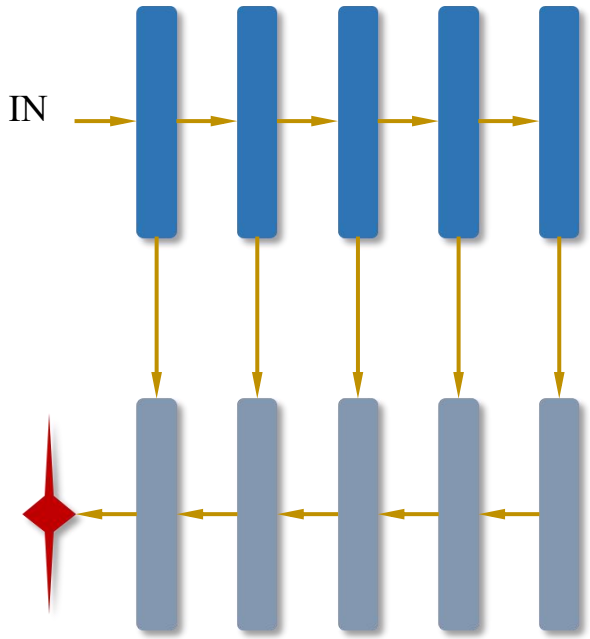




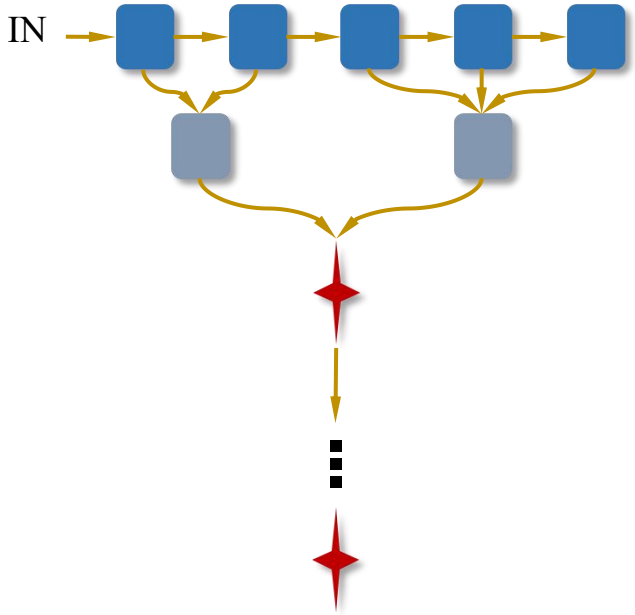


FPN Branch

Drawbacks



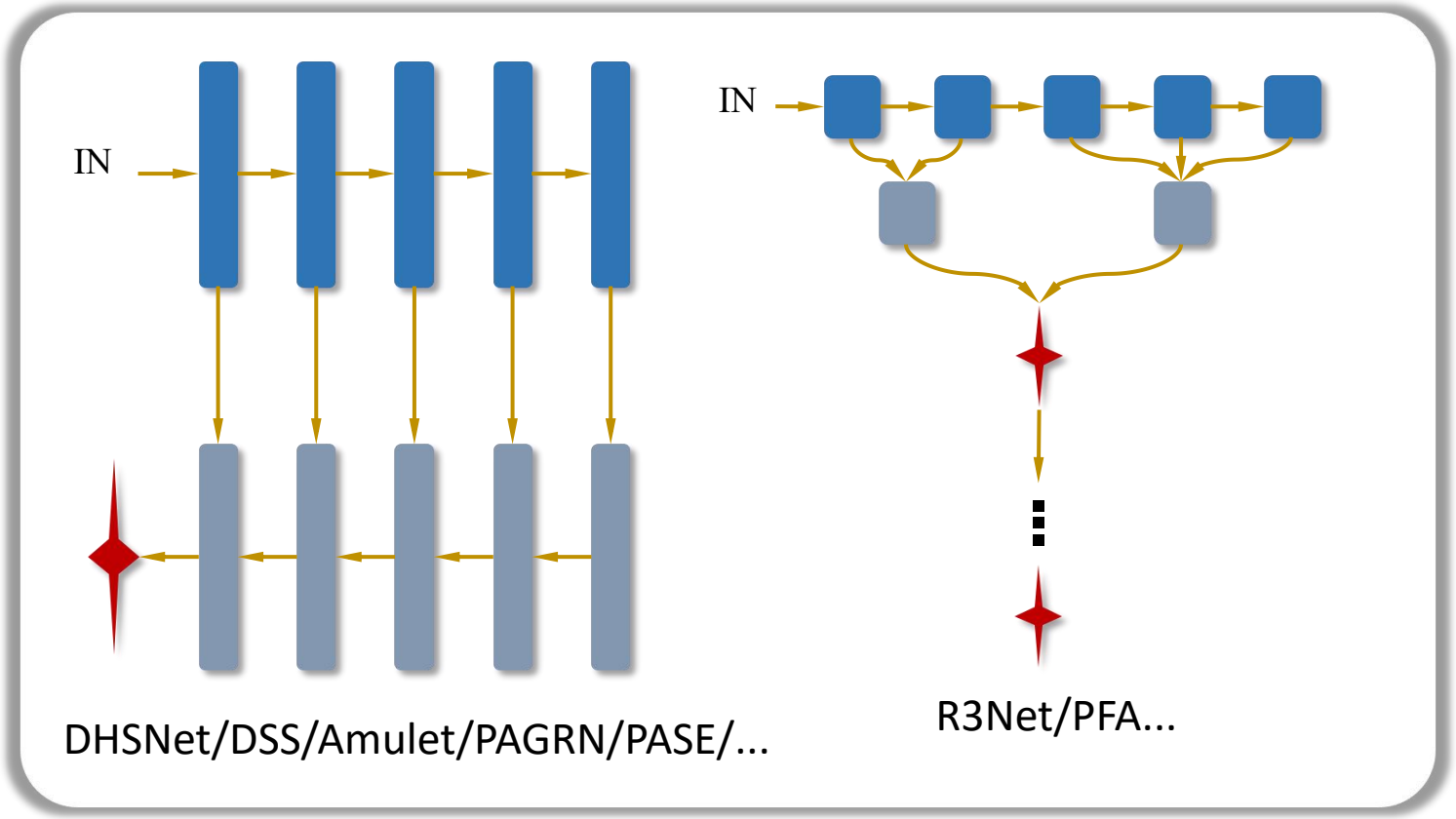
DHSNet/DSS/Amulet/PAGRN/PASE/...

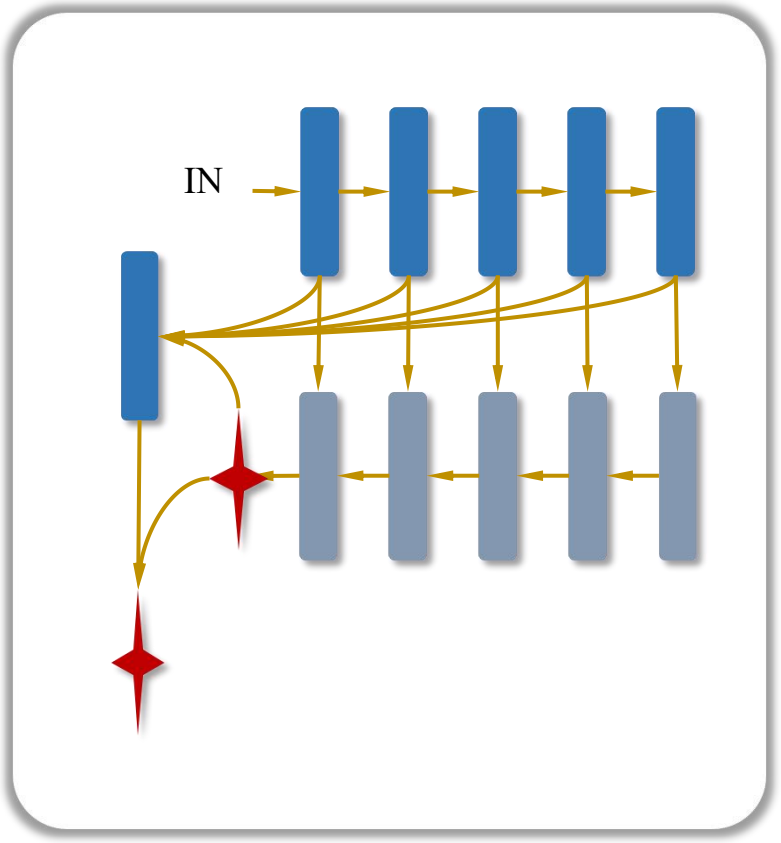


R3Net/PFA...

Only these two basic structures?

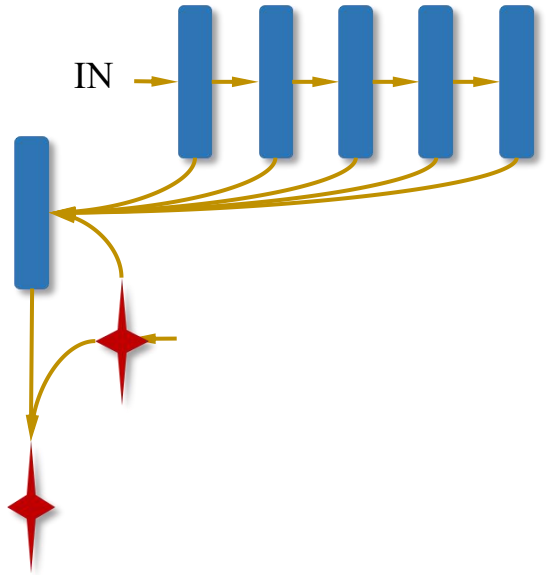
Solution

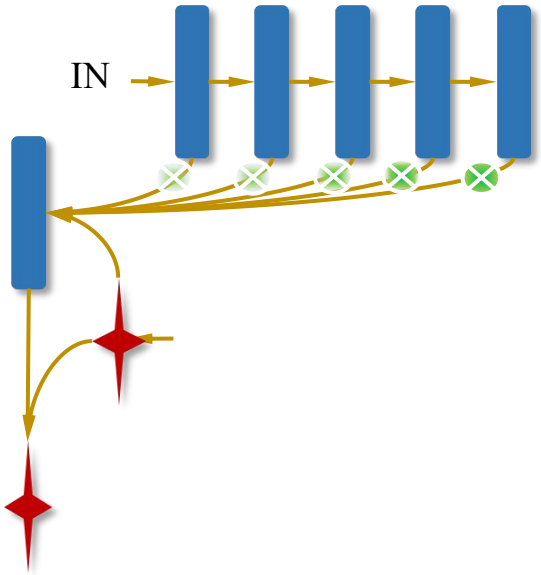


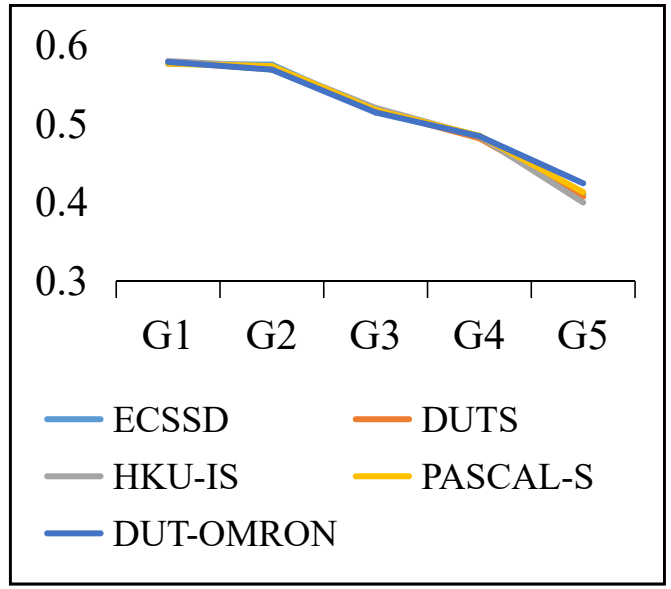


Our Works

Gated Parallel Branch



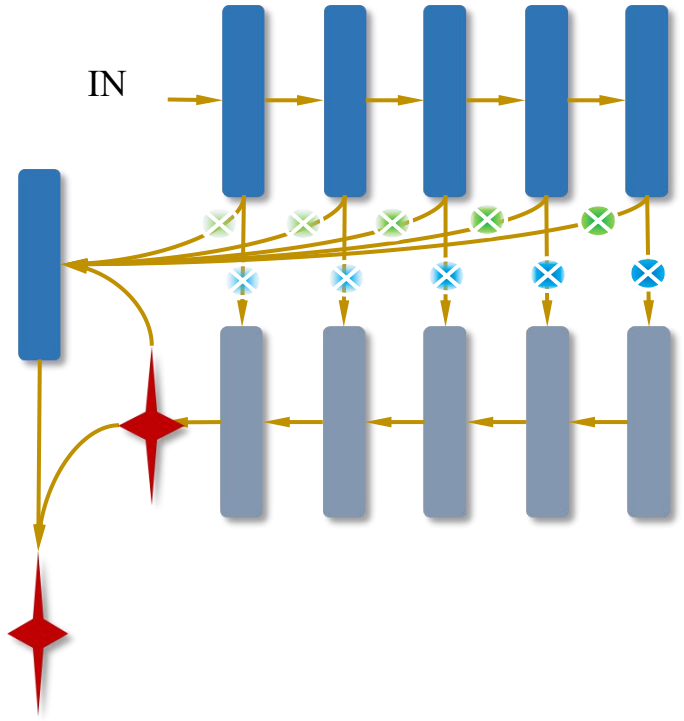


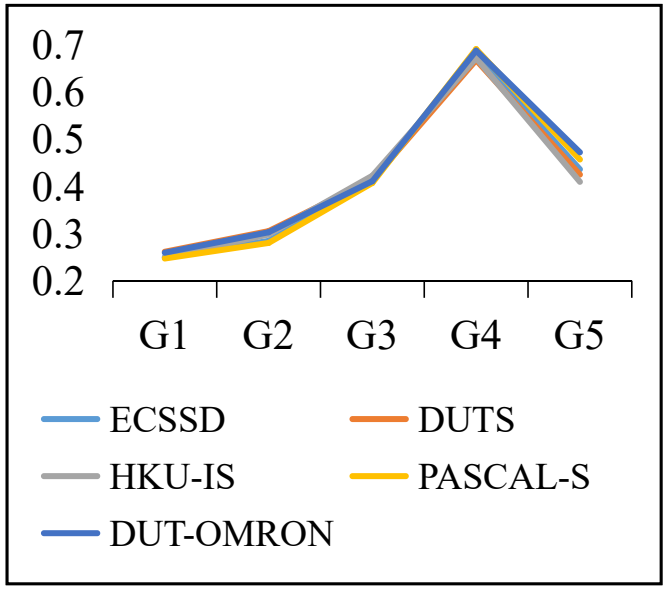


Parallel Branch

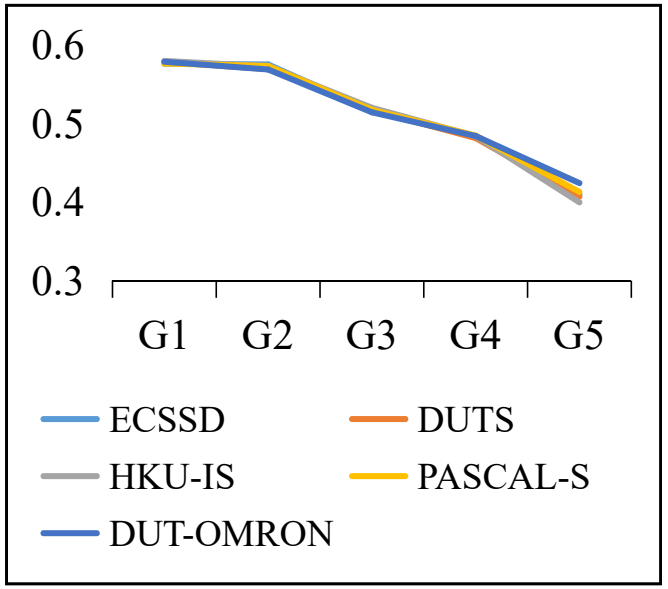
Our Works

Gated Dual Branch



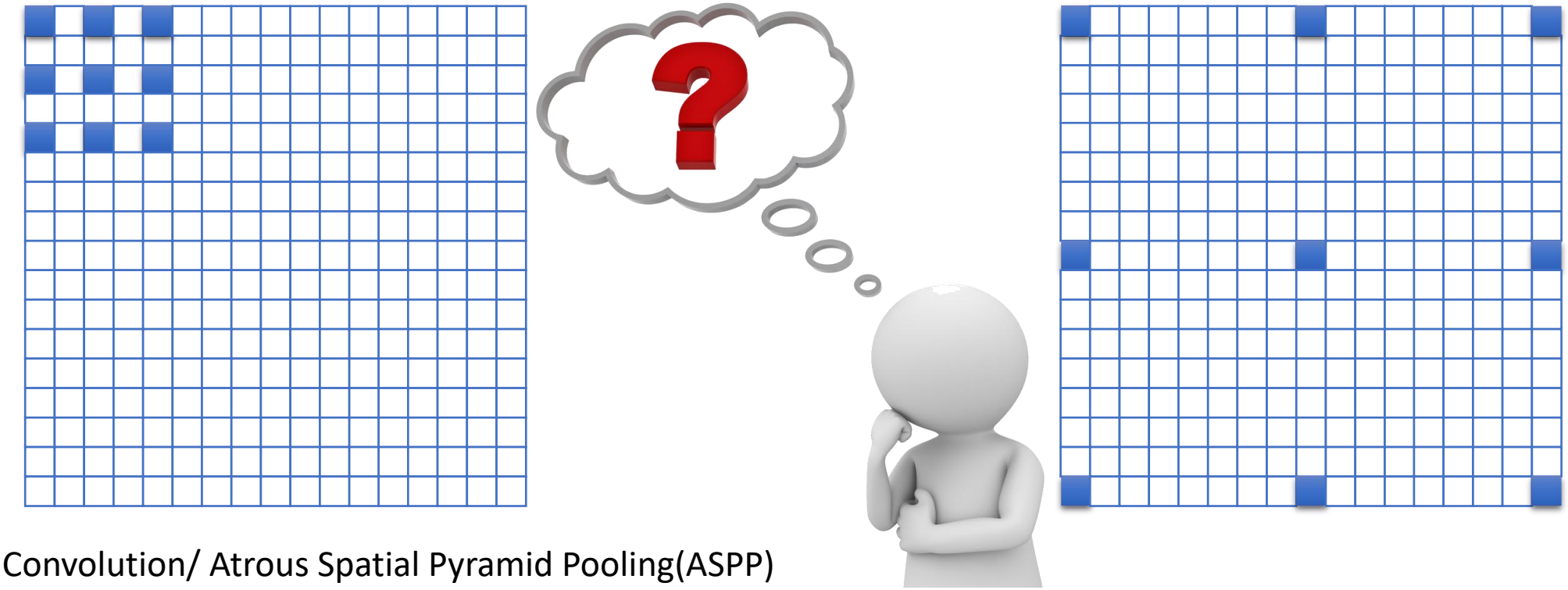


FPN Branch



Parallel Branch

Drawbacks

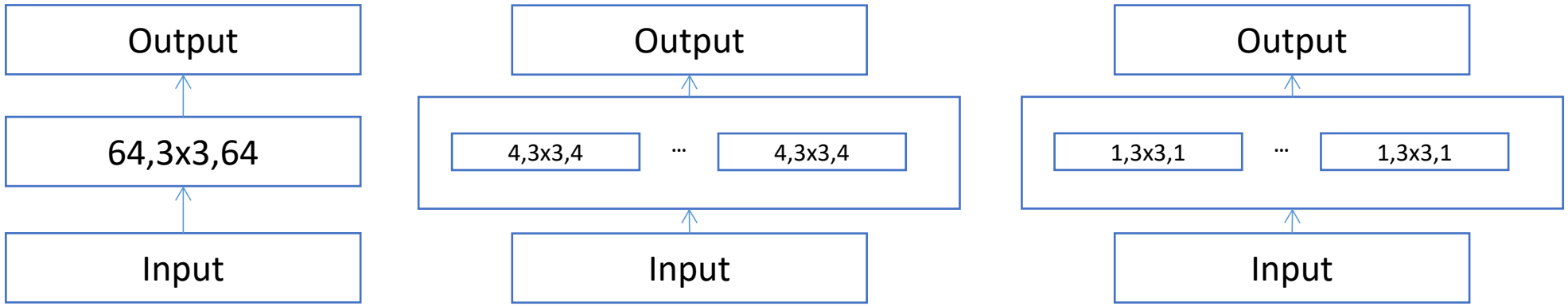


Atrous Convolution/ Atrous Spatial Pyramid Pooling(ASPP)

Large dilation rate, Insert too many zeros.

Solution

Channel Dimension



(a) vanilla convolution

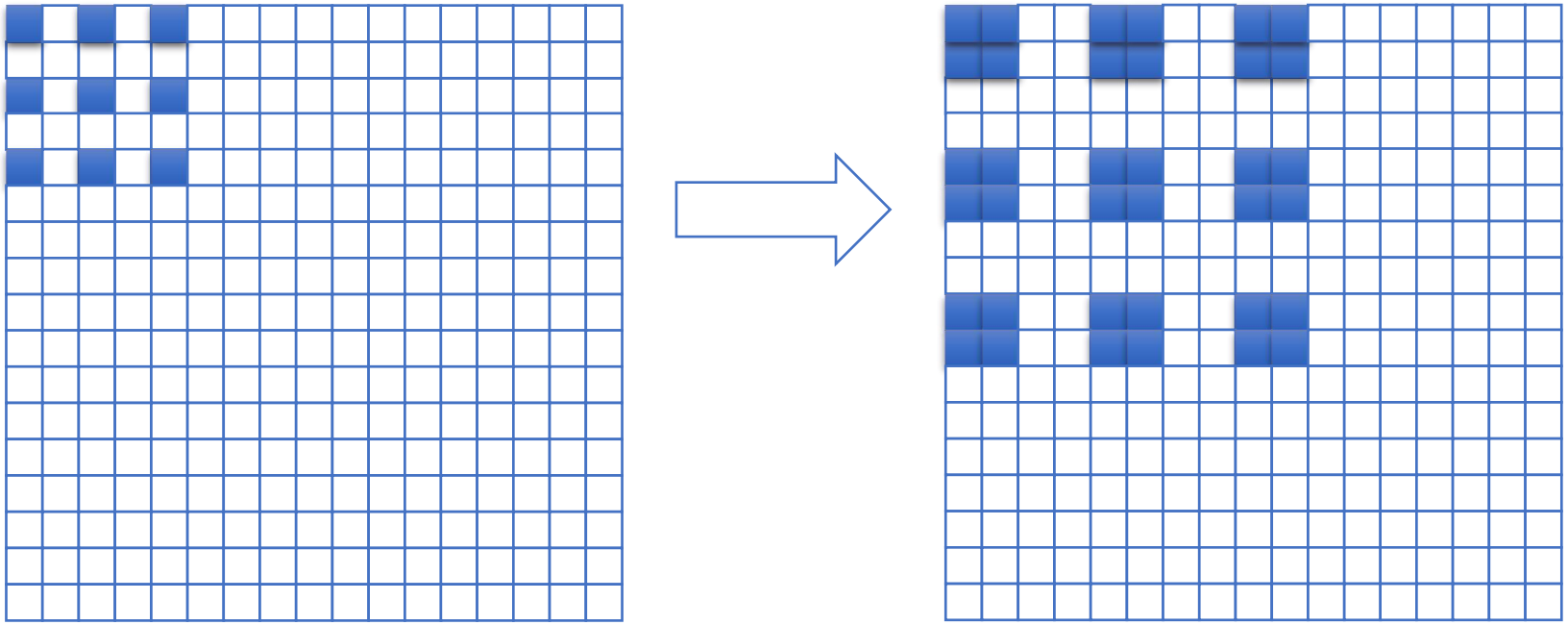
(b) group convolution[1]

(c) depthwise separable convolution[2]

[1] Xie et al., Aggregated residual transformations for deep neural networks.(ResNeXt) CVPR 2017.
[2] Chollet et al., Xception: Deep learning with depthwise separable convolutions.(Xception) CVPR 2017.

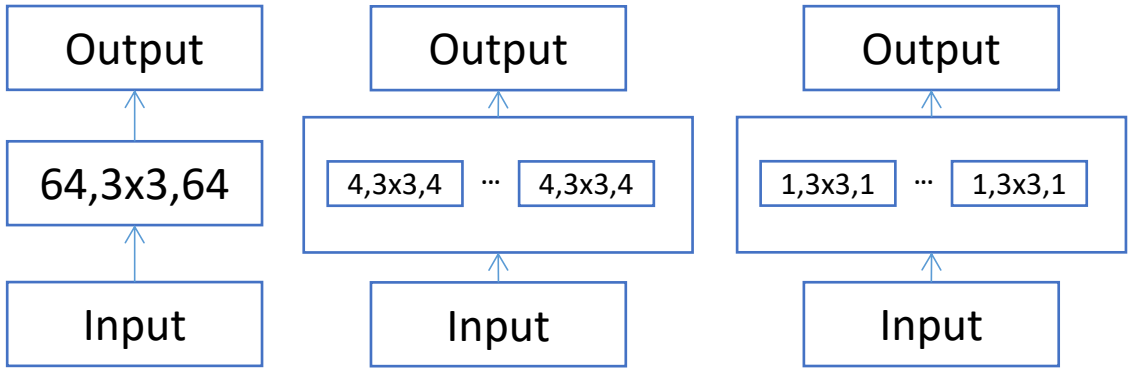
Solution

Spatial Dimension



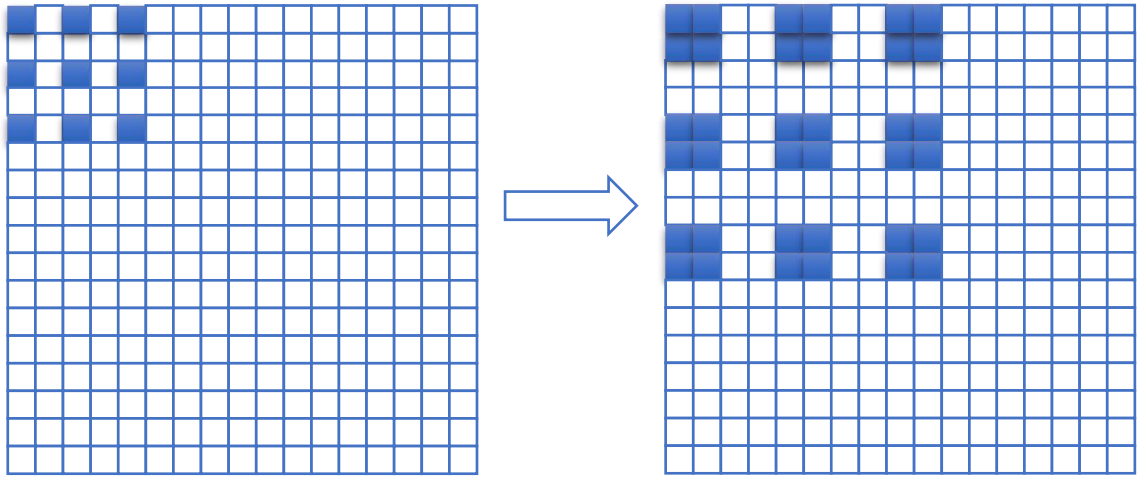
Solution

Channel Dimension



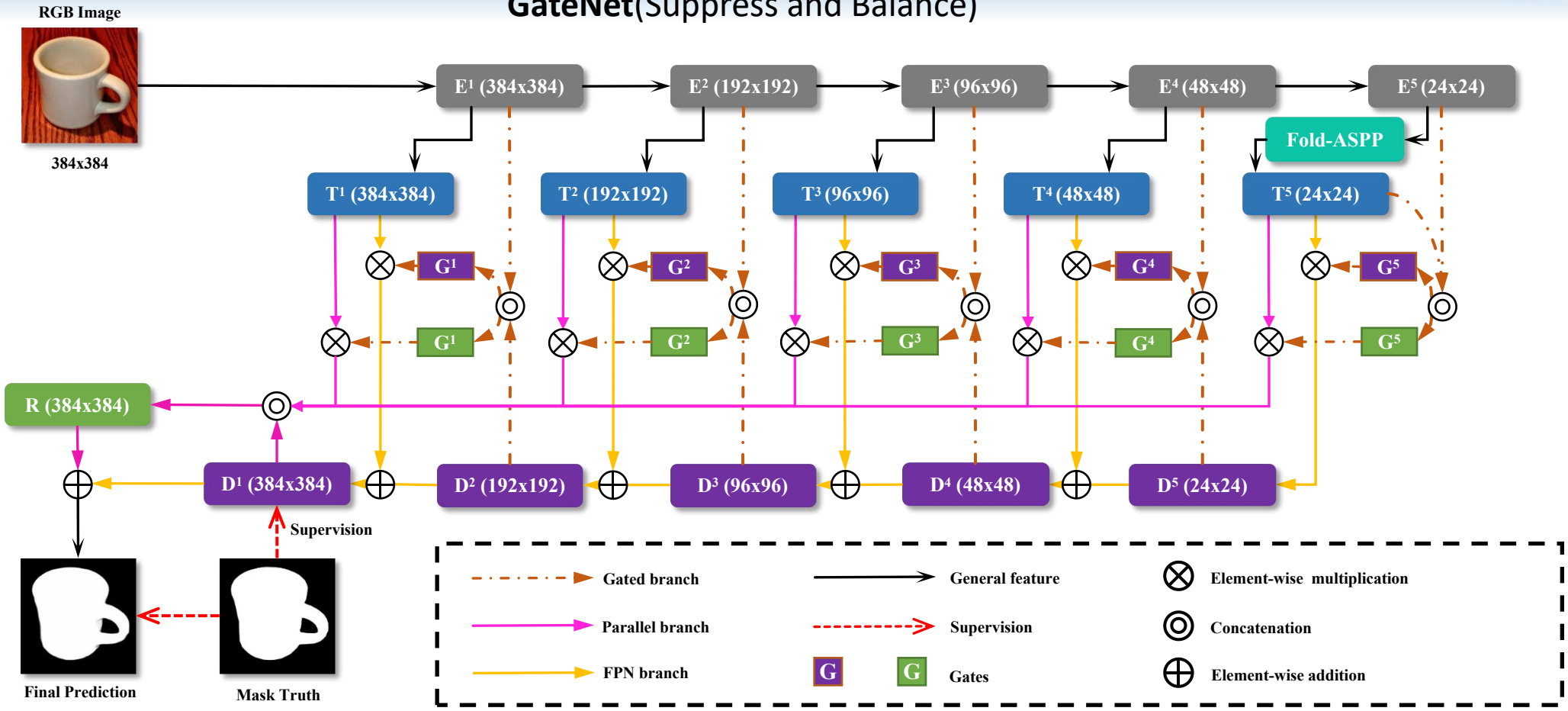
(a) vanilla convolution (b) group convolution (c) depthwise separable convolution

Spatial Dimension



Our Works

GateNet(Suppress and Balance)

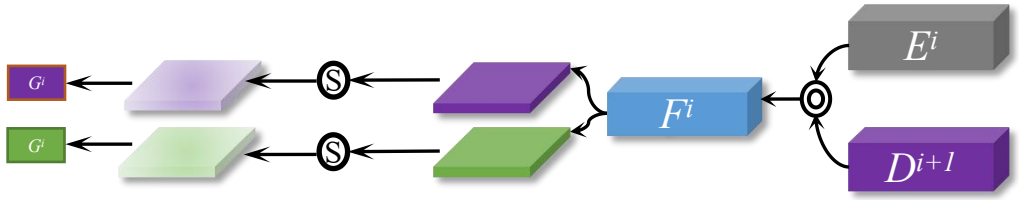


Suppress and Balance: A Simple Gated Network for Salient Object Detection. ECCV 2020(Oral).

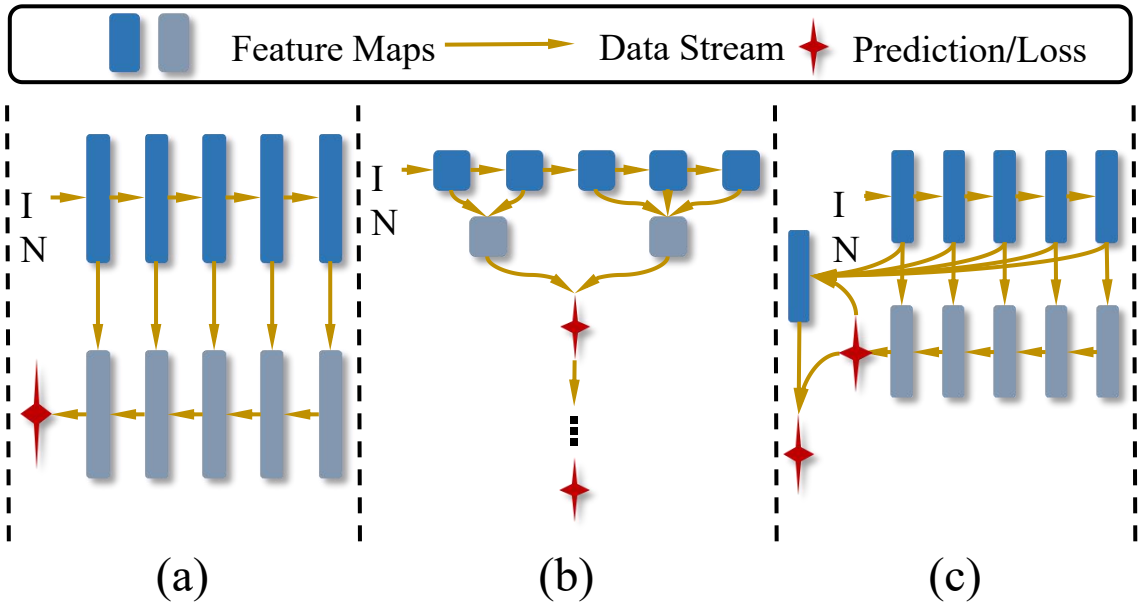
Xiaoqi Zhao, Youwei Pang, Lihe Zhang, Huchuan Lu, Lei Zhang*

Our Works

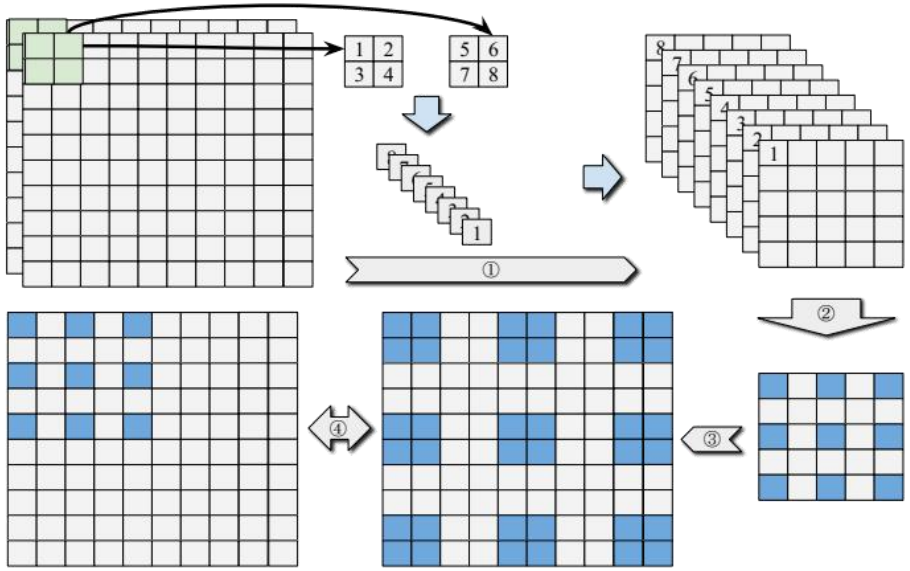
1. Gate Unit



2. Dual Branch



3. Fold-ASPP



Results and Comparisons

RGB SOD

Method	DUTS-test			DUT-OMRON			PASCAL-S			HKU-IS			ECSSD		
	$F_\beta \uparrow$	$S_m \uparrow$	MAE↓	$F_\beta \uparrow$	$S_m \uparrow$	MAE↓	$F_\beta \uparrow$	$S_m \uparrow$	MAE↓	$F_\beta \uparrow$	$S_m \uparrow$	MAE↓	$F_\beta \uparrow$	$S_m \uparrow$	MAE↓
VGG-16 backbone															
$DCL^{\dagger}_{16}$	0.782	0.796	0.088	0.757	0.770	0.080	0.829	0.793	0.109	0.907	0.877	0.048	0.901	0.868	0.068
$DSS^{\dagger}_{17}$	—	—	—	0.781	0.789	0.063	0.840	0.792	0.098	0.916	0.878	0.040	0.921	0.882	0.052
$Amulet_{17}$	0.778	0.804	0.085	0.743	0.780	0.098	0.839	0.819	0.099	0.899	0.886	0.050	0.915	0.894	0.059
$BMPM_{18}$	0.852	0.860	0.049	0.774	0.808	0.064	0.862	0.842	0.076	0.921	0.906	0.039	0.928	0.911	0.045
RAS_{18}	0.831	0.838	0.059	0.786	0.813	0.062	0.836	0.793	0.106	0.913	0.887	0.045	0.921	0.893	0.056
$PAGRN_{18}$	0.854	0.837	0.056	0.771	0.774	0.071	0.855	0.814	0.095	0.919	0.889	0.048	0.927	0.889	0.061
HRS_{19}	0.843	0.828	0.051	0.762	0.771	0.066	0.850	0.798	0.092	0.913	0.882	0.042	0.920	0.883	0.054
$MLMS_{19}$	0.852	0.861	0.049	0.774	0.808	0.064	0.864	0.844	0.075	0.921	0.906	0.039	0.928	0.911	0.045
$PAGE_{19}$	0.838	0.853	0.052	0.792	**0.824**	0.062	0.858	0.837	0.079	0.920	0.904	0.036	0.931	0.912	0.042
$BANet_{19}$	0.852	0.860	0.046	0.793	0.822	**0.061**	0.866	0.838	0.079	0.919	0.901	0.037	0.935	0.913	**0.041**
GateNet	**0.870**	**0.869**	**0.045**	**0.794**	0.820	**0.061**	**0.882**	**0.855**	**0.070**	**0.928**	**0.909**	**0.035**	**0.941**	**0.917**	**0.041**
ResNet-50 backbone															
$SRM^{\dagger}_{17}$	0.826	0.835	0.059	0.769	0.797	0.069	0.848	0.830	0.087	0.906	0.886	0.046	0.917	0.895	0.054
$DGRL_{18}$	0.828	0.841	0.050	0.774	0.805	0.062	0.856	0.836	0.073	0.911	0.895	0.036	0.922	0.903	0.041
CPD_{19}	0.865	0.868	0.043	0.797	0.824	0.056	0.870	0.844	0.074	0.925	0.906	0.034	0.939	0.918	0.037
$ICNet_{19}$	0.855	0.864	0.048	0.813	**0.837**	0.061	0.865	0.849	0.072	0.925	0.908	0.037	0.938	0.918	0.041
$BASNet_{19}$	0.860	0.864	0.048	0.805	0.835	0.057	0.860	0.834	0.079	0.930	0.907	**0.033**	0.943	0.916	0.037
$BANet_{19}$	0.872	0.878	**0.040**	0.803	0.832	0.059	0.877	0.851	0.072	0.930	0.913	**0.033**	0.944	**0.924**	**0.035**
GateNet	**0.888**	**0.884**	**0.040**	**0.818**	**0.837**	**0.055**	**0.883**	**0.857**	**0.069**	**0.933**	**0.915**	**0.033**	**0.945**	0.920	0.040
ResNet/ResNeXt-101 backbone															
$R3Net^{\dagger X}_{18}$	0.819	0.827	0.063	0.795	0.816	0.063	0.844	0.802	0.095	0.915	0.895	0.035	0.934	0.910	0.040
$Capsal^{S}_{19}$	0.819	0.818	0.063	0.639	0.673	0.101	0.869	0.837	0.074	0.883	0.851	0.058	0.863	0.826	0.077
$GateNet^{S}$	**0.893**	**0.889**	**0.038**	**0.821**	**0.844**	**0.054**	**0.883**	**0.862**	**0.067**	**0.937**	**0.920**	**0.031**	**0.951**	**0.930**	**0.035**
$GateNet^{X}$	**0.898**	**0.895**	**0.035**	**0.829**	**0.848**	**0.051**	**0.888**	**0.865**	**0.065**	**0.943**	**0.925**	**0.029**	**0.952**	**0.929**	**0.035**

Ablation Studys

RGB SOD

Table 2. Ablation analysis on the DUTS dataset.

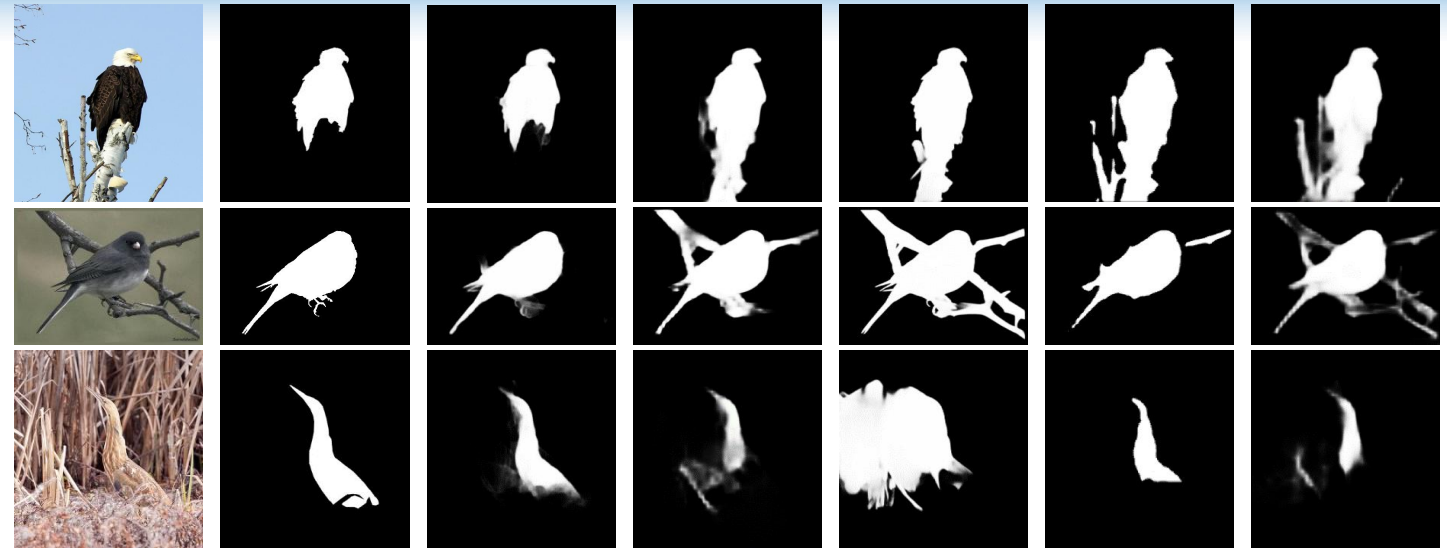
	F_β	S_m	MAE
Baseline (FPN)	0.816	0.829	0.060
+ *Gate Units*	0.840	0.847	0.053
+ *Fold-ASPP*	0.866	0.863	0.047
+ *Parallel Branch*	0.870	0.869	0.045

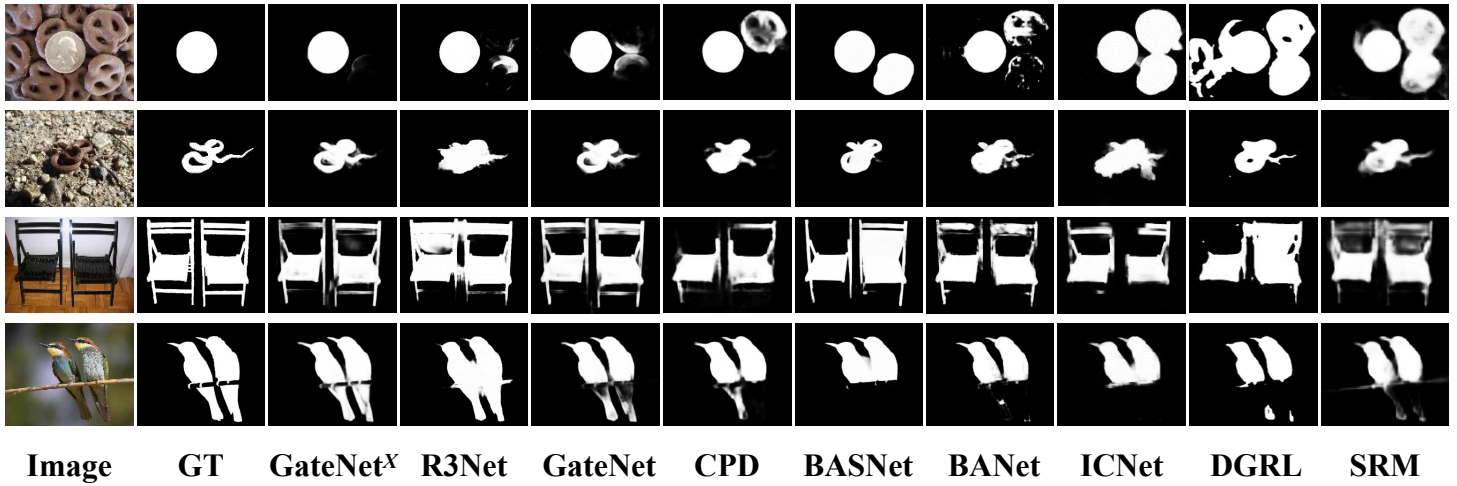
Table 3. Evaluation of the folded convolution and Fold-ASPP. (x) stands for different sampling rates of atrous convolution.

	Atrous(2)	Atrous(4)	Atrous(6)	Fold(2)	Fold(4)	Fold(6)	ASPP	Fold-ASPP
F_β	0.840	0.845	0.848	0.853	0.856	0.860	0.856	**0.866**
MAE	0.055	0.053	0.051	0.051	0.050	0.048	0.051	**0.047**
S_m	0.847	0.849	0.851	0.856	0.858	0.859	0.860	**0.863**

Qualitative Evaluation

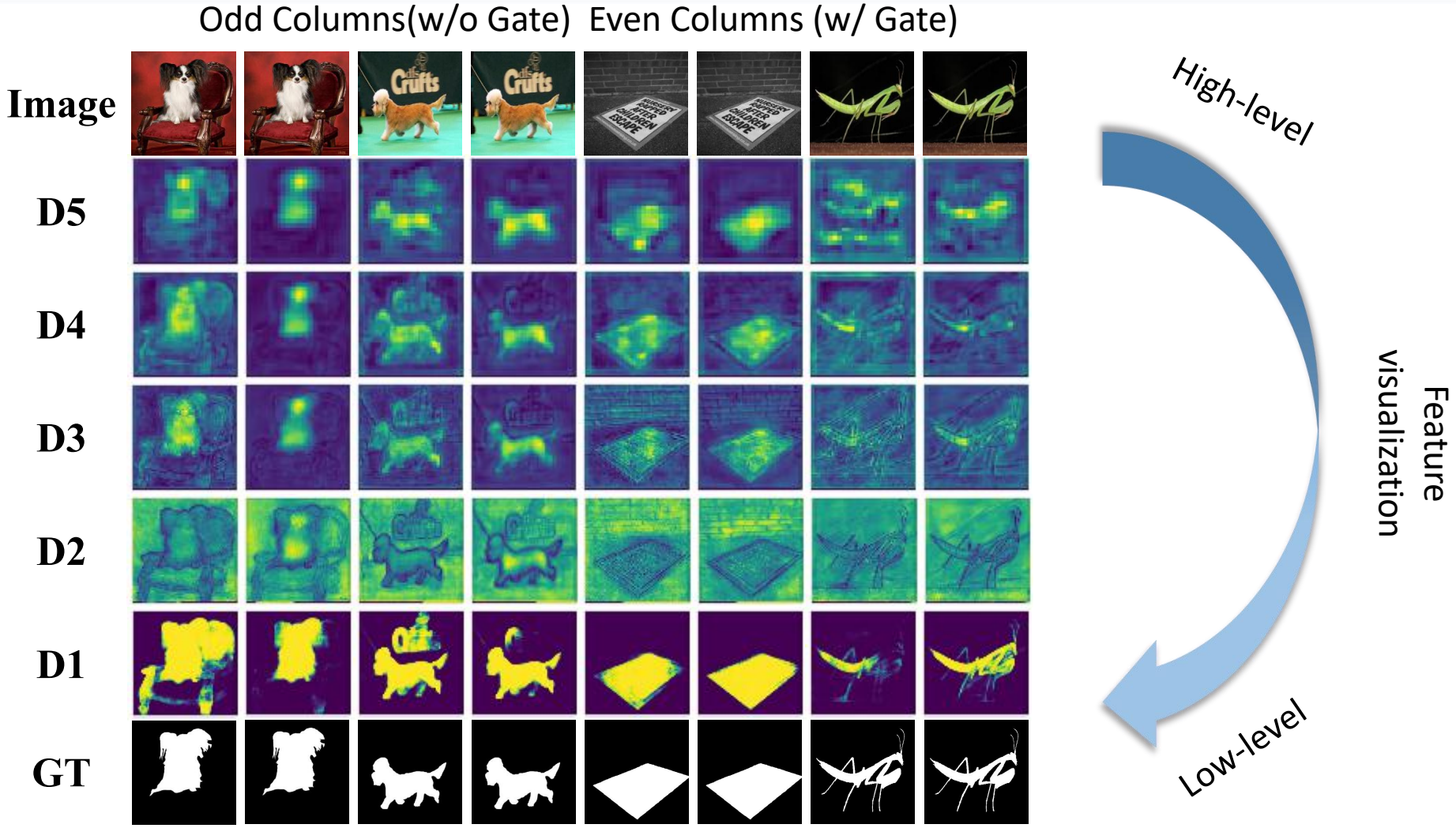
RGB SOD





Qualitative Evaluation

Gate Unit



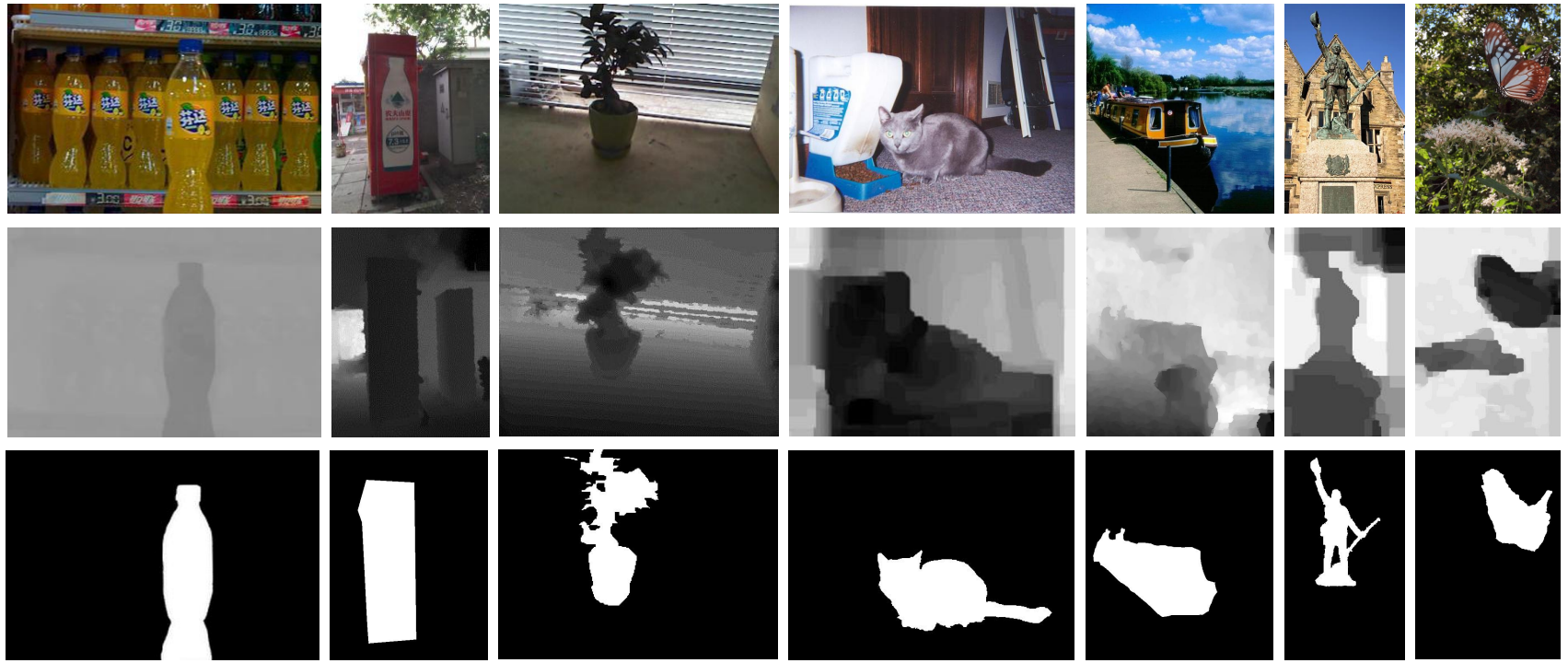
Other Tasks

RGB-D Salient Object Detection

RGB

Depth

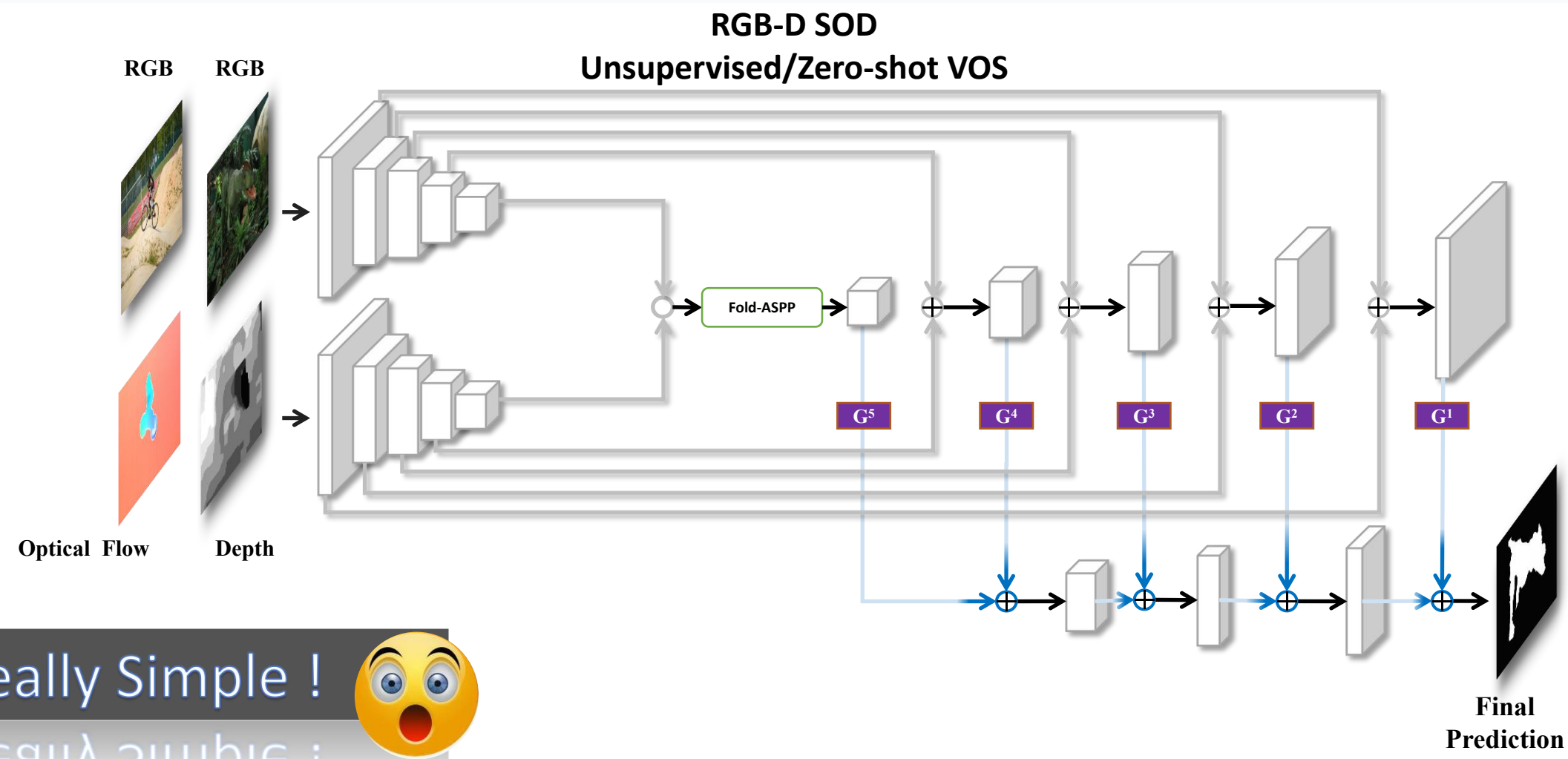
GT



Other Tasks

Video Object Segmentation

Other Tasks



Really Simple !

Other Tasks

Unsupervised/Zero-shot VOS

Table 6. Quantitative comparison of Zero-shot VOS methods on the DAVIS-16 va dation set. ↑ and ↓ indicate that the larger and smaller scores are better, respectivel The best results are shown in **red**. The subscript in each model name is the publicatic year.

2019

Metric		LVO_{17} [46]	ARP_{17} [23]	PDB_{18} [43]	$LSMO_{19}$ [47]	$MotAdapt_{19}$ [41]	AGS_{19} [55]	$COSNet_{19}$ [30]	$AnDiff_{19}$ [65]	GateNet Ours
$\mathcal{J}$	Mean↑	75.9	76.2	77.2	78.2	77.2	79.7	80.5	**81.7**	80.9
	Recall↑	89.1	91.1	90.1	89.1	87.8	91.1	93.1	90.9	**94.3**
	Decay↓	**0.0**	7.0	0.9	4.1	5.0	1.9	4.4	2.2	3.3
$\mathcal{F}$	Mean↑	72.1	70.6	74.5	75.9	77.4	77.4	79.5	**80.5**	79.4
	Recall↑	83.4	83.5	84.4	84.7	84.4	85.8	**89.5**	85.1	89.2
	Decay↓	1.3	7.9	**-0.2**	3.5	3.3	1.6	5.0	0.6	2.9

RGB-D SOD

Table 4. Quantitative comparison. ↑ and ↓ indicate that the larger and smaller scores are better, respectively. Among the CNN-based methods, the best results are shown in **red**. The subscript in each model name is the publication year.

2019

Metric		Traditional Methods			CNNs-Based Models							
		DES_{14} [8]	$DCMC_{16}$ [10]	$CDCP_{17}$ [75]	DF_{17} [37]	$CTMF_{18}$ [17]	$PCANet_{18}$ [3]	$MMCI_{19}$ [5]	$TANet_{19}$ [4]	$CPFP_{19}$ [71]	$DMRA_{19}$ [35]	GateNet Ours
SSD [74]	F_β^{max} ↑	0.260	0.750	0.576	0.763	0.755	0.844	0.823	0.835	0.801	0.858	**0.868**
	F_β^{mean} ↑	0.073	0.684	0.524	0.709	0.709	0.786	0.748	0.767	0.726	0.821	**0.822**
	F_β^{w} ↑	0.172	0.480	0.429	0.536	0.622	0.733	0.662	0.727	0.709	**0.787**	0.785
	S_m ↑	0.341	0.706	0.603	0.741	0.776	0.842	0.813	0.839	0.807	0.856	**0.870**
	E_m ↑	0.475	0.790	0.714	0.801	0.838	0.890	0.860	0.886	0.832	0.898	**0.901**
	$\mathcal{M}$ ↓	0.500	0.168	0.219	0.151	0.100	0.063	0.082	0.063	0.082	0.059	**0.055**
NJUD [22]	F_β^{max} ↑	0.328	0.769	0.661	0.789	0.857	0.888	0.868	0.888	0.890	0.896	**0.914**
	F_β^{mean} ↑	0.165	0.715	0.618	0.744	0.788	0.844	0.813	0.844	0.837	0.871	**0.879**
	F_β^{w} ↑	0.234	0.497	0.510	0.545	0.720	0.803	0.739	0.805	0.828	0.847	**0.849**
	S_m ↑	0.413	0.703	0.672	0.735	0.849	0.877	0.859	0.878	0.878	0.885	**0.902**
	E_m ↑	0.491	0.796	0.751	0.818	0.866	0.909	0.882	0.909	0.900	0.920	**0.922**
	$\mathcal{M}$ ↓	0.448	0.167	0.182	0.151	0.085	0.059	0.079	0.061	0.053	0.051	**0.047**
RGBD135 [8]	F_β^{max} ↑	0.800	0.311	0.651	0.625	0.865	0.842	0.839	0.853	0.882	0.906	**0.919**
	F_β^{mean} ↑	0.695	0.234	0.594	0.573	0.778	0.774	0.762	0.795	0.829	0.867	**0.891**
	F_β^{w} ↑	0.301	0.169	0.478	0.392	0.687	0.711	0.650	0.740	0.787	**0.843**	0.838
	S_m ↑	0.632	0.469	0.709	0.685	0.863	0.843	0.848	0.858	0.872	0.899	**0.905**
	E_m ↑	0.817	0.676	0.810	0.806	0.911	0.912	0.904	0.919	0.927	0.944	**0.966**
	$\mathcal{M}$ ↓	0.289	0.196	0.120	0.131	0.055	0.050	0.065	0.046	0.038	**0.030**	**0.030**
NLPR [33]	F_β^{max} ↑	0.695	0.413	0.687	0.752	0.841	0.864	0.841	0.876	0.884	0.888	**0.904**
	F_β^{mean} ↑	0.583	0.328	0.592	0.683	0.724	0.795	0.730	0.796	0.818	**0.855**	0.854
	F_β^{w} ↑	0.254	0.259	0.501	0.516	0.679	0.762	0.676	0.780	0.807	**0.840**	0.838
	S_m ↑	0.582	0.550	0.724	0.769	0.860	0.874	0.856	0.886	0.884	0.898	**0.910**
	E_m ↑	0.760	0.685	0.786	0.840	0.869	0.916	0.872	0.916	0.920	**0.942**	**0.942**
	$\mathcal{M}$ ↓	0.301	0.196	0.115	0.100	0.056	0.044	0.059	0.041	0.038	**0.031**	0.032
SIP [15]	F_β^{max} ↑	0.720	0.680	0.544	0.704	0.720	0.861	0.840	0.851	0.870	0.847	**0.894**
	F_β^{mean} ↑	0.644	0.645	0.495	0.673	0.684	0.825	0.795	0.809	0.819	0.815	**0.856**
	F_β^{w} ↑	0.342	0.414	0.397	0.406	0.535	0.768	0.712	0.748	0.788	0.734	**0.810**
	S_m ↑	0.616	0.683	0.595	0.653	0.716	0.842	0.833	0.835	0.850	0.800	**0.874**
	E_m ↑	0.751	0.787	0.722	0.794	0.824	0.900	0.886	0.894	0.899	0.858	**0.914**
	$\mathcal{M}$ ↓	0.298	0.186	0.224	0.185	0.139	0.071	0.086	0.075	0.064	0.088	**0.057**

ECCV'20
ONLINE
23-28 AUGUST 2020

16TH EUROPEAN CONFERENCE ON
COMPUTER VISION

WWW.ECCV2020.EU

Thanks for your attention

IIAU-LAB

Code Link

Paper Link