SALIENCY DETECTION

Interactive Two-Stream Decoder for Accurate and Fast Saliency Detection

——ITSD Network

Huajun Zhou[1], Xiaohua Xie[1,2,3]*, Jianhuang Lai[1,2,3], Zixuan Chen[1], Lingxiao Yang[1]
[1]School of Data and Computer Science, Sun Yat-sen University, China
[2]Guangdong Province Key Laboratory of Information Security Technology, China
[3]Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China

01 Saliency Object Detection

- **Target**
  - “Salient object detection” or “salient object segmentation” is commonly interpreted in computer vision as a process that includes two stages:

  1) Detecting the most salient object
  2) Segmenting the accurate region of that object

saliency map

02 Metrics

- **Precision—Recall**

$$P = \frac{TP}{TP + FN} \qquad R = \frac{TP}{TP + FP}$$

- **F-measure**
  - $\beta^2 = 0.3$

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

- **MAE(mean absolute error)**

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

Images GT Mask

03 Dataset

- **DUTS**：10553 for train（DUTS-TR）, 5019 for test（DUTS-TE）
- **SOD**: 300 images
- **PASCAL-S**: 850 images
- **ECSSD**: 1000 images
- **HKU-IS**: 4447 images
- **DUT-O**: 5168 images

01 Interactive Two-Stream Decoder for Accurate and Fast Saliency Detection

- **CVPR 2020**

- **Motivation**

Contour information largely improves the performance of saliency detection. However, the discussion on the correlation between saliency and contour remains scarce.

- Several works introduced contour into networks by proposing a boundary-aware objective function.

BASNet: Boundary-Aware Salient Object Detection

Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan and Martin Jagersand
University of Alberta, Canada
{xuebin,vincent.zhang,chuang8,cgao3,masood1,mj7}@ualberta.ca

- Constructing a boundary-aware network becomes an impressive method in the saliency detection task.

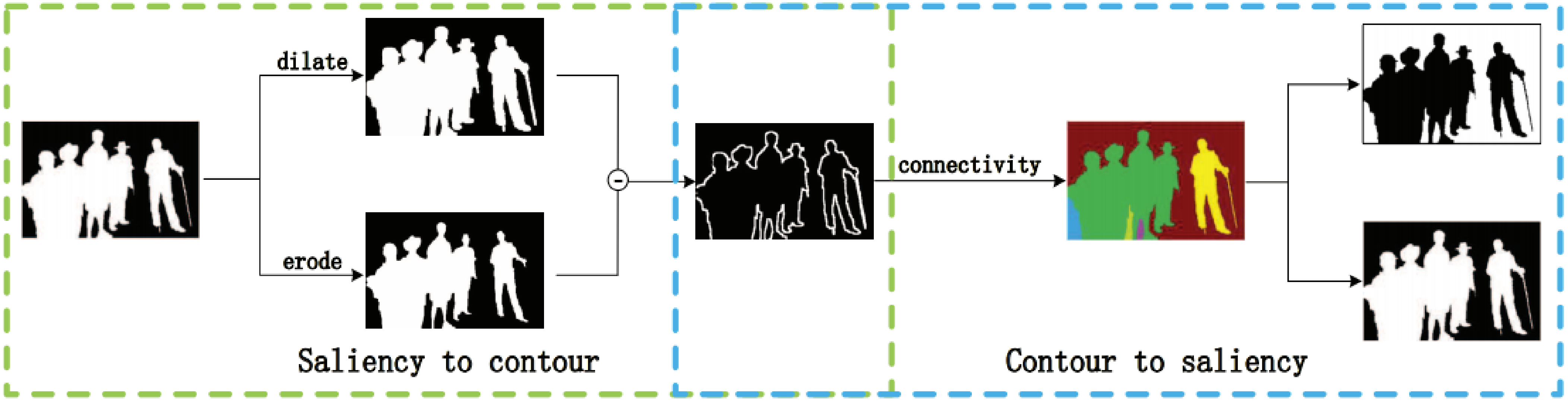
A Simple Pooling-Based Design for Real-Time Salient Object Detection

Jiang-Jiang Liu[1*] Qibin Hou[1*] Ming-Ming Cheng[1 †] Jiashi Feng[2] Jianmin Jiang[3]
[1]TKLNDST, College of CS, Nankai University [2]NUS [3]Shenzhen University
{j04.liu, andrewhoux}@gmail.com

01 Interactive Two-Stream Decoder for Accurate and Fast Saliency Detection

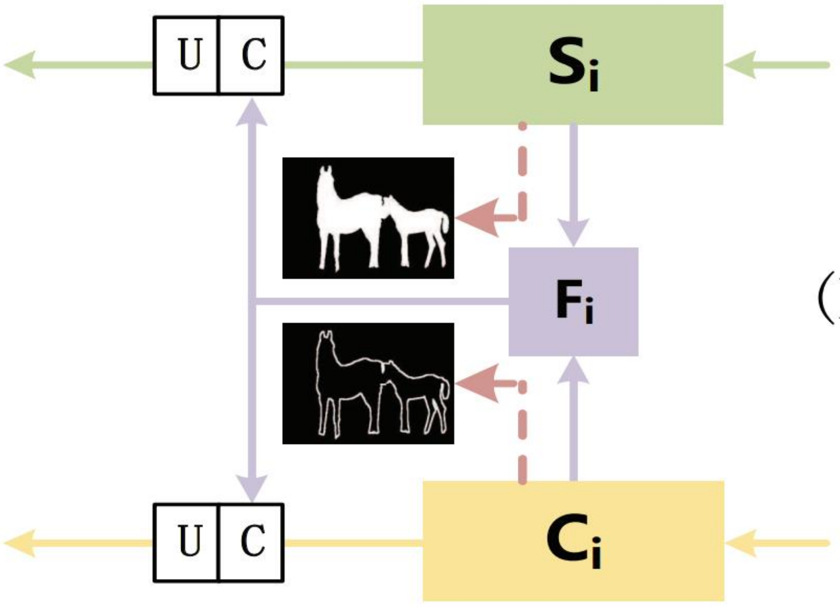
• Contribution

- Discuss the correlation between the saliency maps and corresponding contour maps.
- Propose a lightweight Interactive Two-Stream Decoder (ITSD) for saliency detection by exploring multiple cues of the saliency and contour maps.
- Develop an Adaptive ConTour (ACT) loss to improve the representation power of the learned network by taking advantage of hard examples.

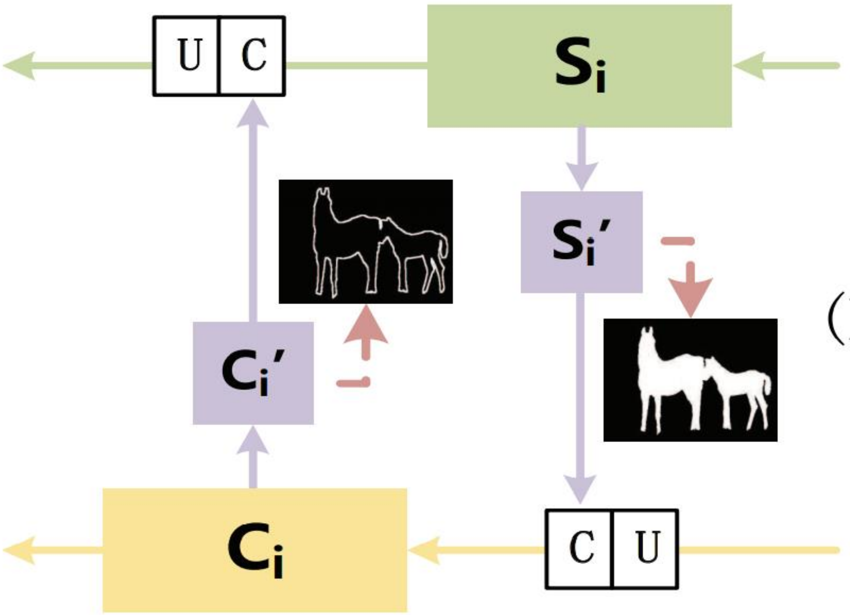


01 Interactive Two-Stream Decoder for Accurate and Fast Saliency Detection

- **Feature Correlation Fusion(FCF)**
  - FCF-a: This simple module cannot guarantee that the fused features can be complementary to each branch.
  - FCF-b: These connections(S_i' and C_i') to ensure the transferred features are related to their original branch.



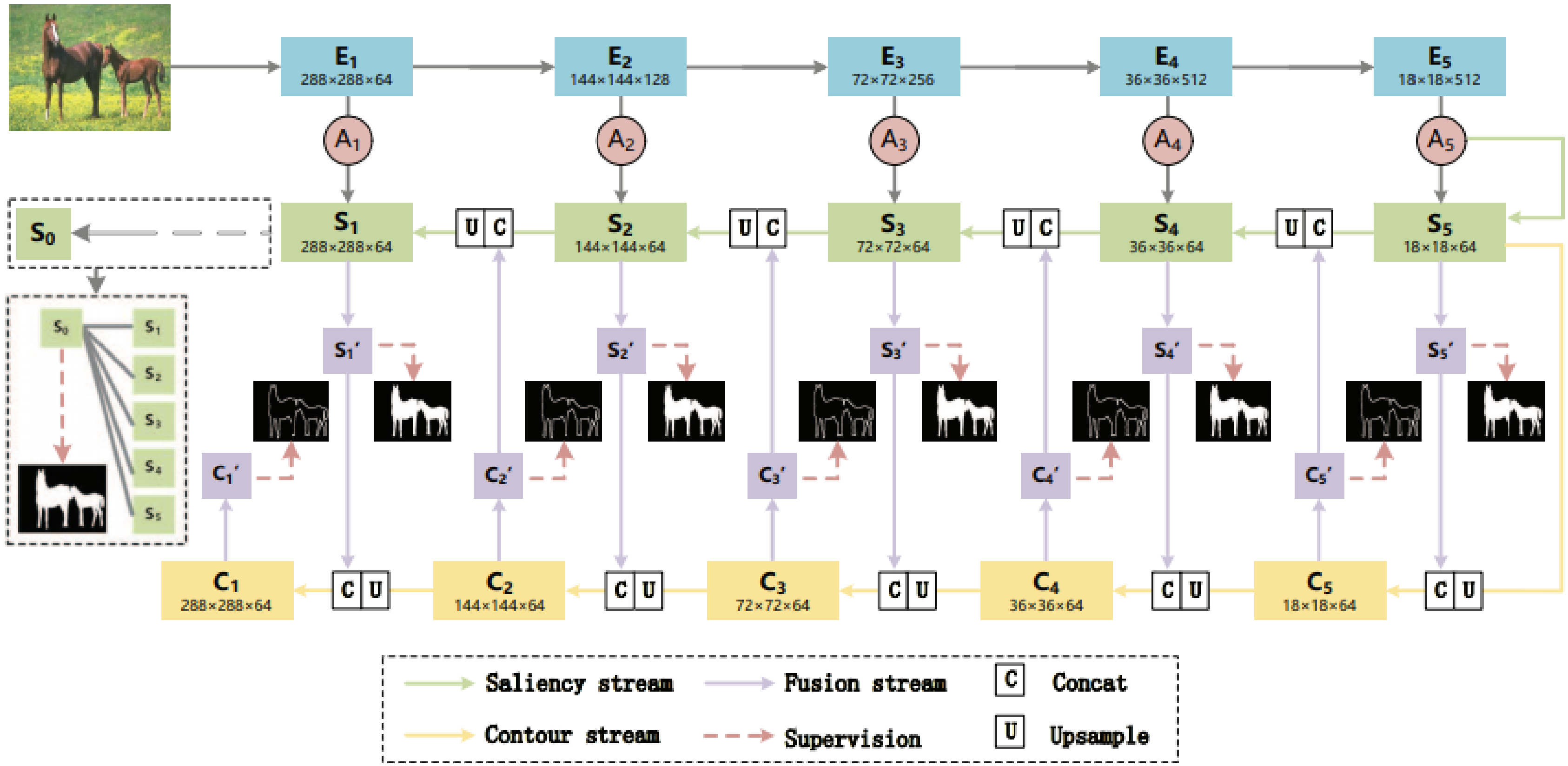
(FCF−a)



(FCF−b)

01 Interactive Two-Stream Decoder for Accurate and Fast Saliency Detection

- **Overall network architecture**
  - Feature encoder: VGG16 or ResNet50, pre-trained on ImageNet
  - Attach a **channel pooling layer** to reduce feature channels and computation loads.

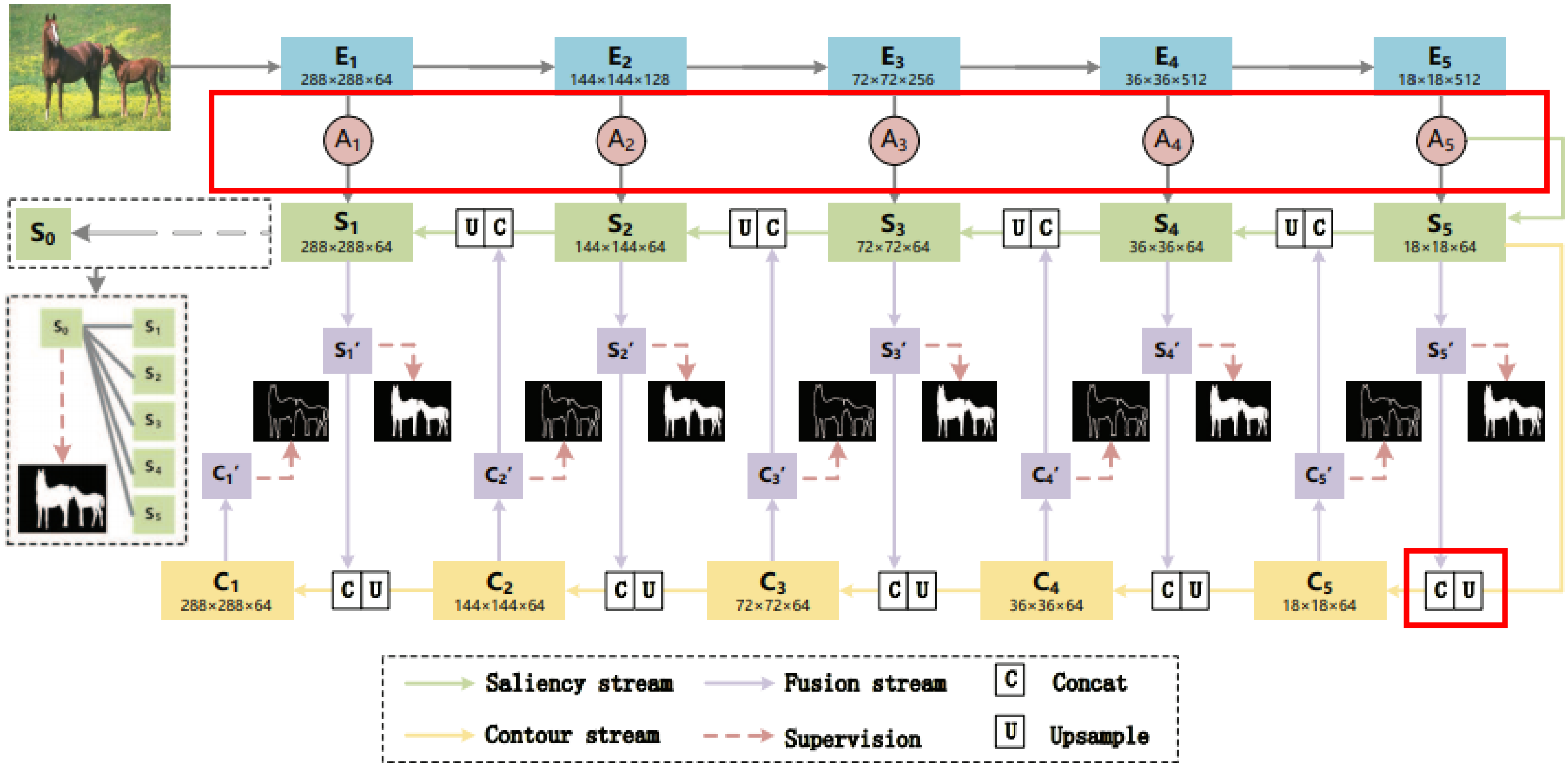


01 Interactive Two-Stream Decoder for Accurate and Fast Saliency Detection

- **Channel pooling layer**

$$A_i = cp(E_i), \quad (1)$$

$$cp(X) = collect_{j\in[0,m-1]}(max_{k\in[0,\frac{n}{m}-1]}X^{j\times\frac{n}{m}+k}), \quad (2)$$



01 Interactive Two-Stream Decoder for Accurate and Fast Saliency Detection

- **Loss for the contour branch**

$$l_{bce}(x,y) = y\log(x) + (1-y)\log(1-x)$$

$$L^c(P^c, G^c) = -\frac{1}{n}\sum_{k=1}^{n} l_{bce}(p_k^c, g_k^c)$$

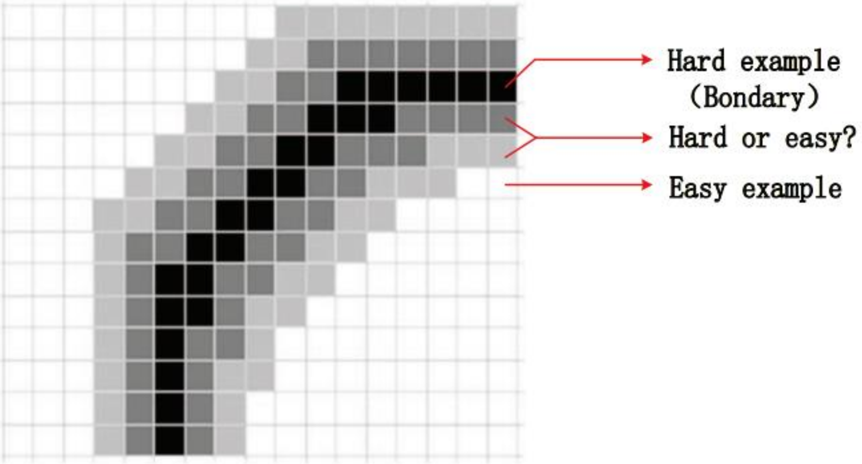
- **Loss for the saliency branch**

$$L^s(P^s, G^s, G^c) = -\frac{1}{n}\sum_{k=1}^{n}(\boxed{g_k^c} \times m + 1) l_{bce}(p_k^s, g_k^s)$$

The definition of hard examples is ambiguous.

$$L^s(P^s, P^c, G^s, G^c) = -\frac{1}{n}\sum_{k=1}^{n}(\boxed{max(p_k^c, g_k^c)} \times m + 1) l_{bce}(p_k^s, g_k^s)$$

Adaptive ConTour (ACT) loss



$$L(P^s, P^c, G^s, G^c) = \sum_{i=0}^{5} L^s(P_i^s, P_i^c, G_i^s, G_i^c) + \lambda \sum_{j=1}^{5} L^c(P_j^c, G_j^c)$$

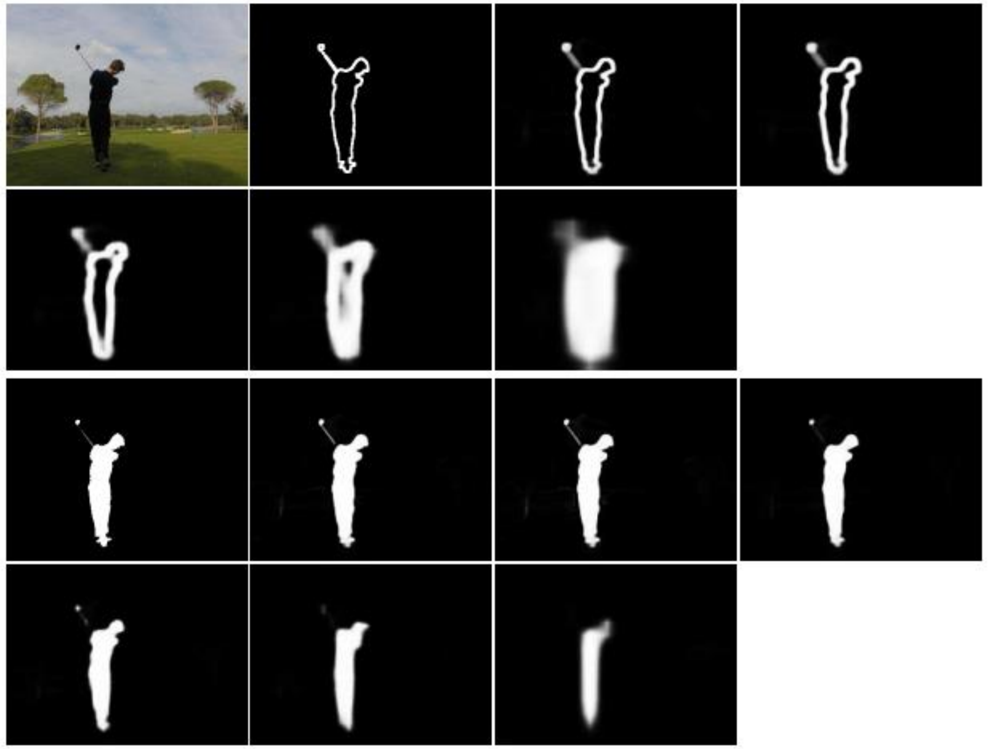
01 Interactive Two-Stream Decoder for Accurate and Fast Saliency Detection

- **Result**

Method	FPS	SOD		PASCAL-S		ECSSD		HKU-IS		DUTS-TE		DUT-O	
		F_β^*	mae	F_β^*	mae	F_β^*	mae	F_β^*	mae	F_β^*	mae	F_β^*	mae
VGG-based													
RFCN [32]	9	.807	.166	.850	.132	.898	.095	.898	.080	.783	.090	.738	.095
Amulet [45]	16	.798	.145	.837	.099	.915	.059	.897	.051	.778	.085	.743	.098
UCF [46]	23	.803	.169	.846	.128	.911	.078	.886	.074	.771	.117	.735	.132
NLDF [24]	12	.842	.125	.829	.103	.905	.063	.902	.048	.812	.066	.753	.080
DSS [12]	25	.837	.127	.828	.107	.908	.062	.900	.050	.813	.064	.760	.074
CKT [18]	23	.829	.119	.850	.086	.910	.054	.896	.048	.807	.062	.757	.071
BMP [44]	22	.851	.106	.859	.081	.928	.044	.920	.038	.850	.049	.774	.064
PAGE [35]	25	.796	.110	.835	.078	.931	.042	.930	.037	.838	.051	.791	.066
PCA [22]	5.6	.855	.108	.858	.081	.931	.047	.921	.042	.851	.054	.794	.068
CTLoss [4]	26	.861	.109	.876	.079	.933	.043	.927	.035	.872	.042	.792	.073
EGNet [48]	9	.869	.110	.863	.076	.941	.044	.929	.034	.880	.043	.826	.056
RA [3]	35	.844	.124	.834	.104	.918	.059	.913	.045	.826	.055	.786	.062
AFNet [8]	45	.855	.110	.867	.078	.935	.042	.923	.036	.862	.046	.797	.057
CPD [37]	66	.850	.114	.866	.074	.936	.040	.924	.033	.864	.043	.794	.057
PoolNet [21]	32	.859	.115	.857	.078	.936	.047	.928	.035	.876	.043	.817	.058
ITSD (Ours)	48	.869	.100	.871	.074	.939	.040	.927	.035	.877	.042	.813	.063
ResNet-based													
BasNet [26]	70	.851	.114	.854	.076	.942	.037	.928	.032	.860	.047	.805	.056
CPD [37]	62	.852	.110	.864	.072	.939	.037	.925	.034	.865	.043	.797	.056
PoolNet [21]	18	.867	.100	.863	.075	.940	.042	.934	.032	.886	.040	.830	.055
EGNet [48]	7.8	.890	.097	.869	.074	.943	.041	.937	.031	.893	.039	.842	.052
SCRN [38]	32	.860	.111	.882	.064	.950	.038	.934	.034	.888	.040	.812	.056
ITSD (Ours)	43	.880	.095	.871	.071	.947	.035	.934	.031	.883	.041	.824	.061

01 Interactive Two-Stream Decoder for Accurate and Fast Saliency Detection

- **Result**



Loss	ECSSD		DUTS-TE		HKU-IS	
	F_β	MAE	F_β	MAE	F_β	MAE
F-score	.929	.043	.845	.050	0.915	.045
BCE	.931	.040	.861	.046	0.921	.040
CTLoss	.935	.040	.872	.045	0.925	.036
ACT	.939	.039	.877	.042	0.927	.035

Module	ECSSD		DUTS-TE		HKU-IS	
	F_β	MAE	F_β	MAE	F_β	MAE
U-shape	.919	.052	.842	.062	0.913	.048
FCF-NoCS	.929	.043	.868	.045	0.921	.035
FCF-a	.930	.041	.865	.046	0.919	.036
FCF-b	.939	.040	.877	.042	0.927	.033