A Review on Salient Object Detection

Feng Lin

Salient Object Detection

- Target
  - Detect and segment salient objects in natural scenes
    a) good detection
    b) high resolution
    c) computational efficiency

- Metric
  - F-score

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

  - MAE (mean absolute error)

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

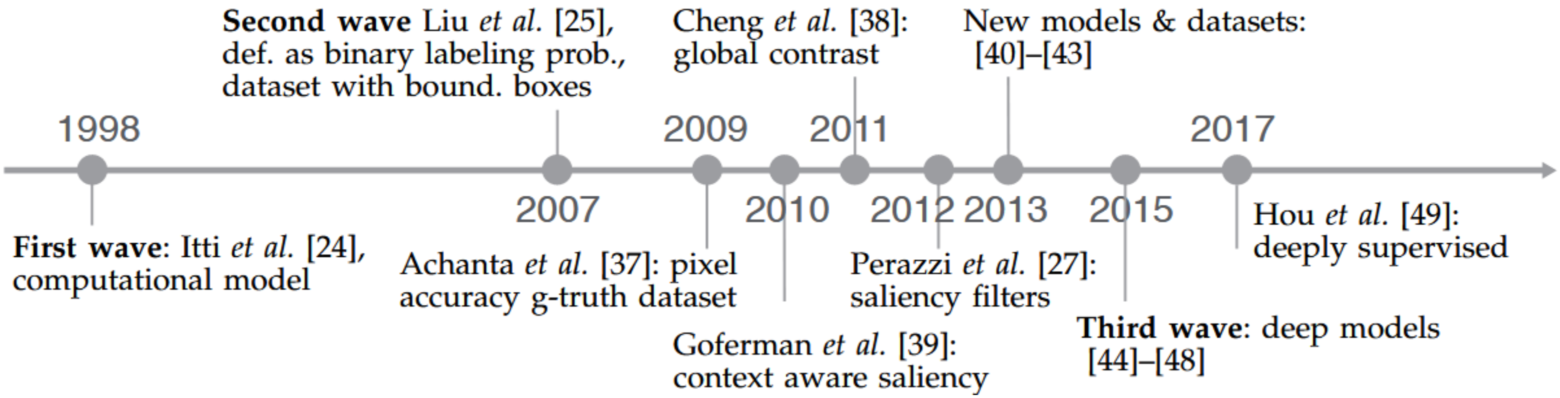
  - S-measure*

Salient Object Detection

- Dataset
  - ECSSD
  - PASCAL-S
  - SOD
  - HKU-IS
  - DUT-OMRON
  - THUR-15K
  - MSRA-10K
  - MSRA-B (2k for training)
  - DUTS (≈15k for training)
  - …

Salient Object Detection

- Method
  - Two stages: (simultaneously perform the two stages in practice)
    - a) detecting the most salient object
    - b) segmenting the accurate region of that object

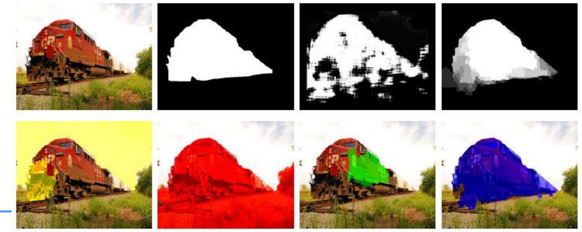


  - Supervised or unsupervised method

Salient Object Detection

- Method
  - Supervised
  - Unsupervised

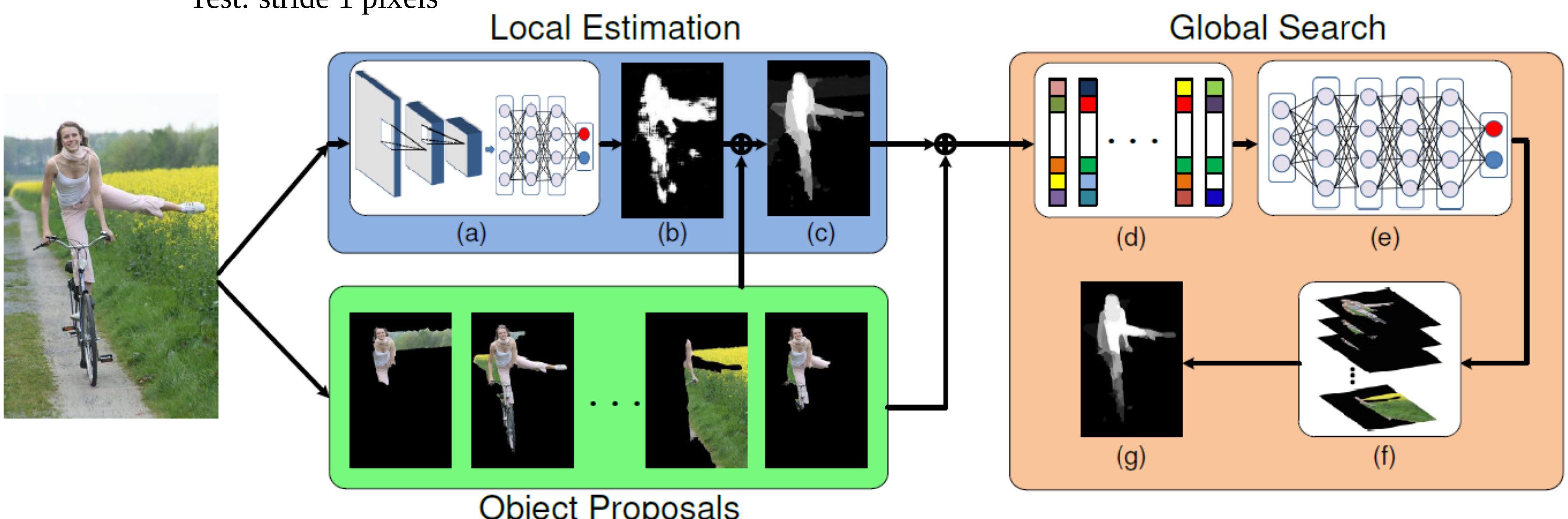
Salient Object Detection



- **Deep Networks for Saliency Detection via Local Estimation and Global Search**
  - Combining local estimation and global search
  - Utilize the geodesic object proposal (GOP)
  - Regress saliency confidence

Train: 51×51 patch, stride 10 pixels, by sliding window
Test: stride 1 pixels

Predict precision and overlap rate (IOU)



$$A_i = \frac{\sum_{x,y} \mathbf{O}_i(x,y) \times \mathbf{S}^L(x,y)}{\sum_{x,y} \mathbf{O}_i(x,y)}$$

$$C_i = \frac{\sum_{x,y} \mathbf{O}_i(x,y) \times \mathbf{S}^L(x,y)}{\sum_{x,y} \mathbf{S}^L(x,y)}$$

Average top K candidate regions

$$\mathbf{S}^G = \frac{\sum_{k=1}^{K} conf_k^G \times \hat{\mathbf{O}}_k}{\sum_{k=1}^{K} conf_k^G}$$

CVPR'15

Salient Object Detection

- **Deep Networks for Saliency Detection via Local Estimation and Global Search**

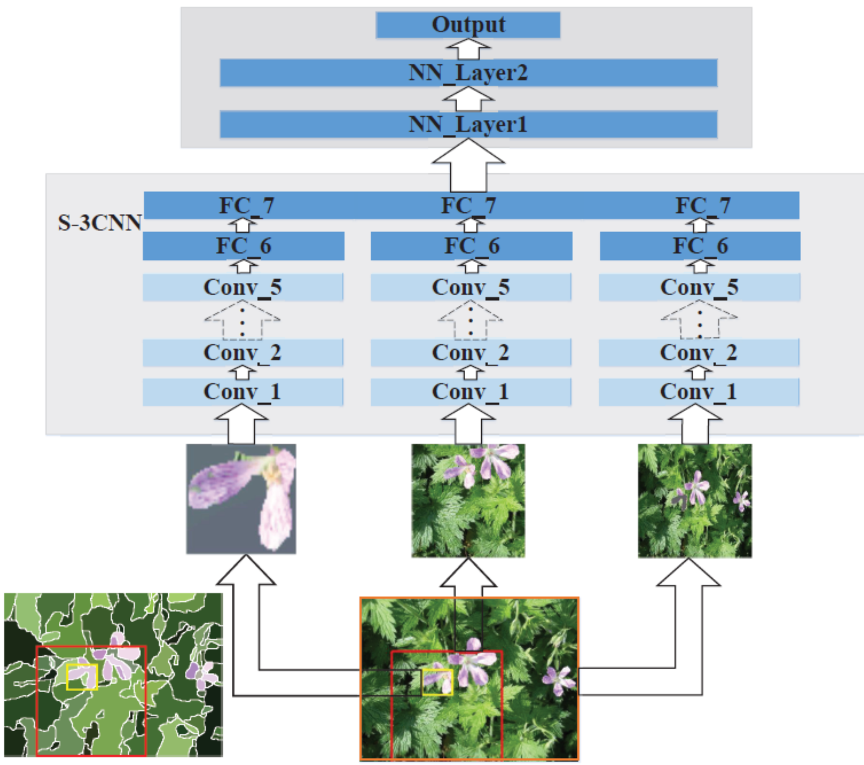
Table 4. Quantitative results using F-measure and MAE. The best and second best results are shown in **red** color and **blue** color.

Data Set	Metric	DRFI	GC	HS	MR	PCA	SVO	UFO	wCtr	CPMC-GBVS	HDCT	LEGS
SOD	F-Measure	**0.617**	0.433	0.480	0.542	0.498	0.217	0.521	0.567	–	0.511	**0.630**
	MAE	**0.230**	0.288	0.301	0.274	0.290	0.414	0.272	0.245	–	0.260	**0.205**
ECCSD	F-Measure	**0.726**	0.568	0.631	0.689	0.575	0.237	0.638	0.672	–	0.641	**0.775**
	MAE	**0.172**	0.218	0.232	0.192	0.252	0.406	0.210	0.178	–	0.204	**0.137**
PASCAL-S	F-Measure	0.619	0.496	0.536	0.600	0.531	0.266	0.552	0.611	**0.654**	0.536	**0.669**
	MAE	0.195	0.245	0.249	0.219	0.239	0.373	0.227	0.193	**0.178**	0.226	**0.170**
MSRA-5000	F-Measure	–	0.704	0.765	**0.789**	0.707	0.302	0.774	0.788	–	0.773	**0.803**
	MAE	–	0.149	0.160	0.130	0.189	0.364	0.145	**0.110**	–	0.141	**0.128**

Salient Object Detection

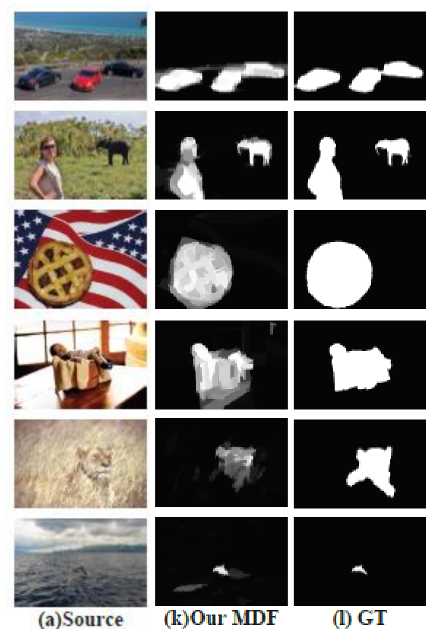
➢ Visual Saliency Based on Multiscale Deep Features

- Enclose the considered region, neighboring regions and the entire image
- Run saliency model repeatedly over every region of the image



$$A = \sum_{k=1}^{M} \alpha_k A^{(k)}$$

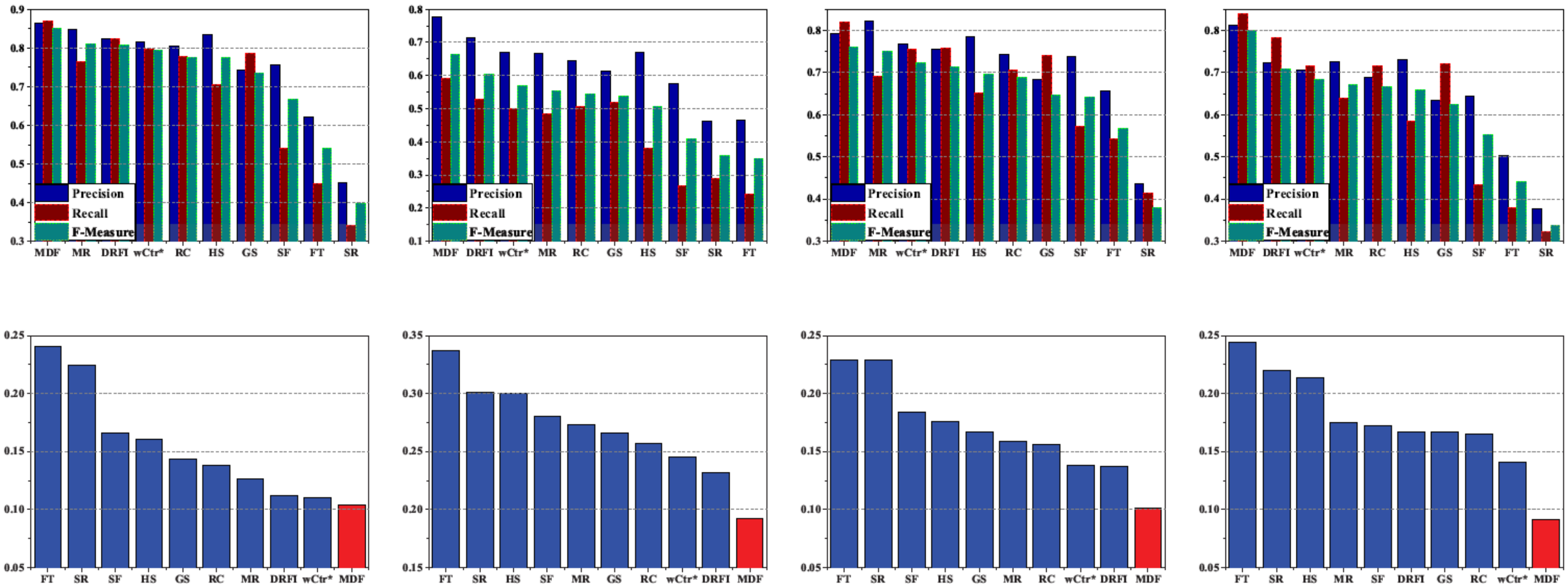
$$\text{s.t. } \{\alpha_k\}_{k=1}^{M} = \underset{\alpha_1,\alpha_2,\ldots,\alpha_M}{\operatorname{argmin}} \sum_{i \in I_v} \left\| A_i - \sum_k \alpha_k A_i^{(k)} \right\|_F^2$$



CVPR'15

Salient Object Detection

➤ **Visual Saliency Based on Multiscale Deep Features**

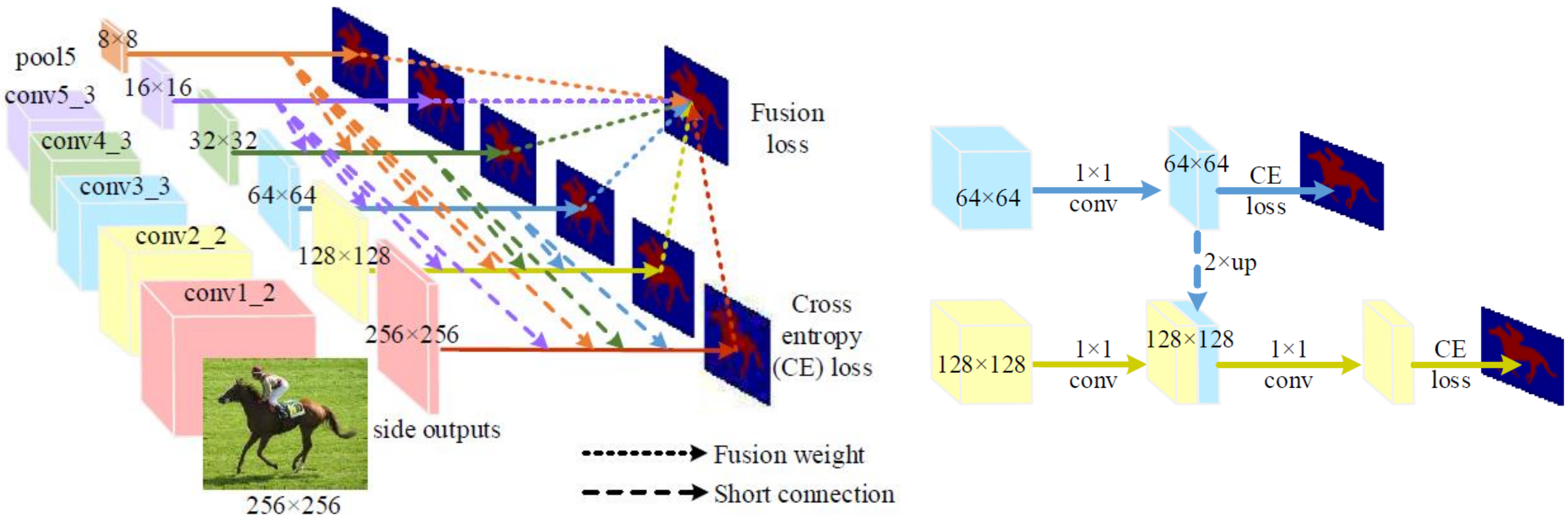


(a) (b) (c) (d)

Figure 3: Quantitative comparison of saliency maps generated from 10 different methods on 4 datasets. From left to right: (a) the MSRA-B dataset, (b) the SOD dataset, (c) the iCoSeg dataset, and (d) our new HKU-IS dataset. From top to bottom: (1st row) the precision-recall curves of different methods, (2nd row) the precision, recall and F-measure using an adaptive threshold, and (3rd row) the mean absolute error.

Salient Object Detection

- **Deeply Supervised Salient Object Detection with Short Connections**



$$\tilde{L}_{\text{final}}(\mathbf{W}, \tilde{\mathbf{w}}, \mathbf{f}, \mathbf{r}) = \tilde{L}_{\text{fuse}}(\mathbf{W}, \tilde{\mathbf{w}}, \mathbf{f}, \mathbf{r}) + \tilde{L}_{\text{side}}(\mathbf{W}, \tilde{\mathbf{w}}, \mathbf{r})$$

CVPR'17

Salient Object Detection

- **Deeply Supervised Salient Object Detection with Short Connections**

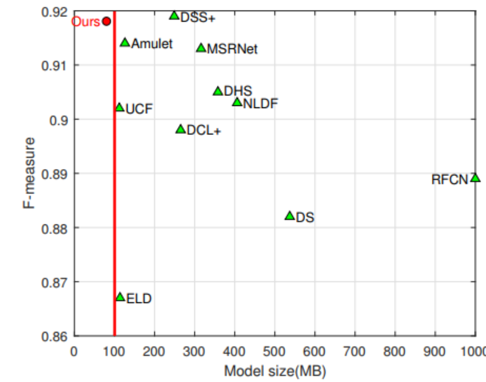
Methods \ Datasets	MSRA-B [37]		ECSSD [51]		HKU-IS [29]		PASCALS [34]		SOD [39, 40]	
	F_β	MAE	F_β	MAE	F_β	MAE	F_β	MAE	F_β	MAE
RC [7]	0.817	0.138	0.741	0.187	0.726	0.165	0.640	0.225	0.657	0.242
CHM [31]	0.809	0.138	0.722	0.195	0.728	0.158	0.631	0.222	0.655	0.249
DSR [32]	0.812	0.119	0.737	0.173	0.735	0.140	0.646	0.204	0.655	0.234
DRFI [24]	0.855	0.119	0.787	0.166	0.783	0.143	0.679	0.221	0.712	0.215
MC [52]	0.872	0.062	0.822	0.107	0.781	0.098	0.721	0.147	0.708	0.184
ELD [13]	0.914	0.042	0.865	0.981	0.844	0.071	0.767	0.121	0.760	0.154
MDF [29]	0.885	0.104	0.833	0.108	0.860	0.129	0.764	0.145	0.785	0.155
DS [13]	-	-	0.810	0.160	-	-	0.818	0.170	0.781	0.150
RFCN [47]	0.926	0.062	0.898	0.097	0.895	0.079	0.827	0.118	0.805	0.161
DHS [36]	-	-	0.905	0.061	0.892	0.052	0.820	0.091	0.823	0.127
DCL [30]	0.916	0.047	0.898	0.071	0.907	0.048	0.822	0.108	0.832	0.126
Ours	0.927	0.028	0.915	0.052	0.913	0.039	0.830	0.080	0.842	0.118

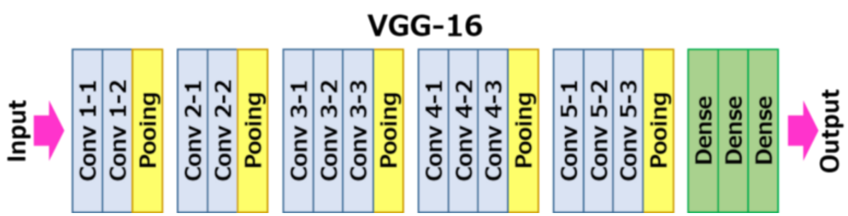
Table 3: Quantitative comparisons with 11 methods on 5 popular datasets. The top three results are highlighted in red, green, and blue, respectively.

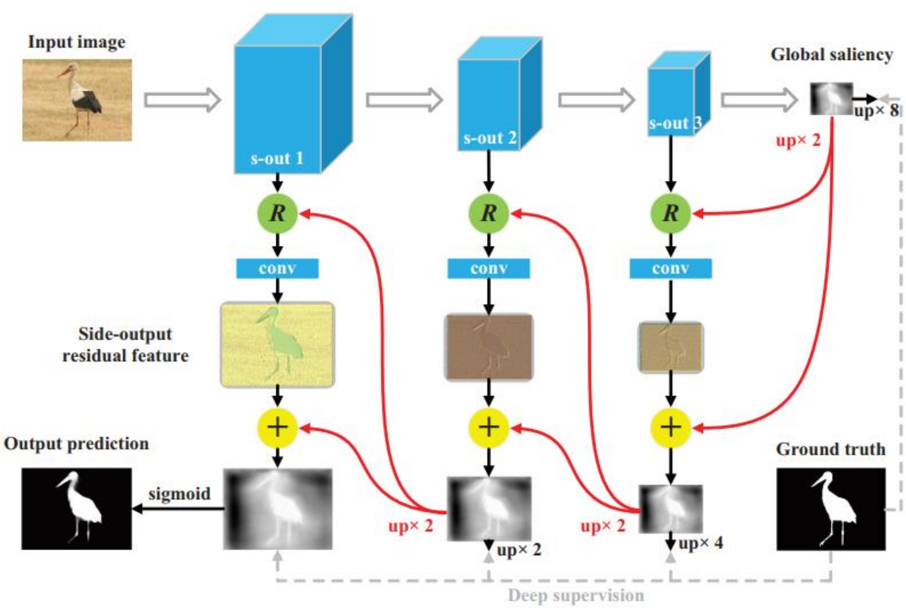
Salient Object Detection

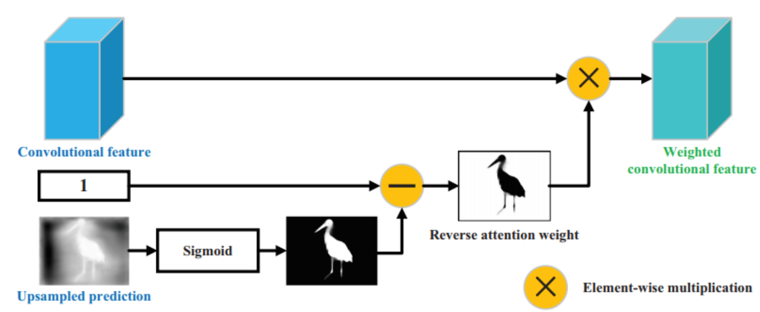
➢ Reverse Attention for Salient Object Detection

- Fine boundary, efficiency (45 FPS) and light weight (81 MB)
- Learn redundant features inside object without RA



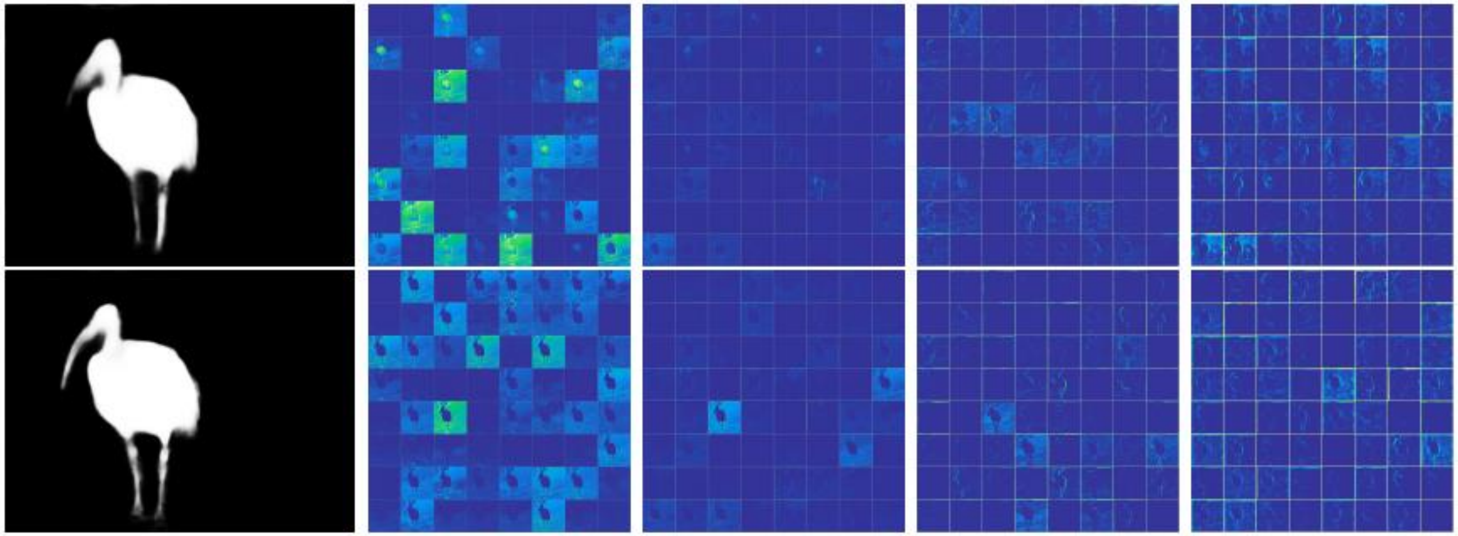






Salient Object Detection

- **Reverse Attention for Salient Object Detection**



Salient Object Detection

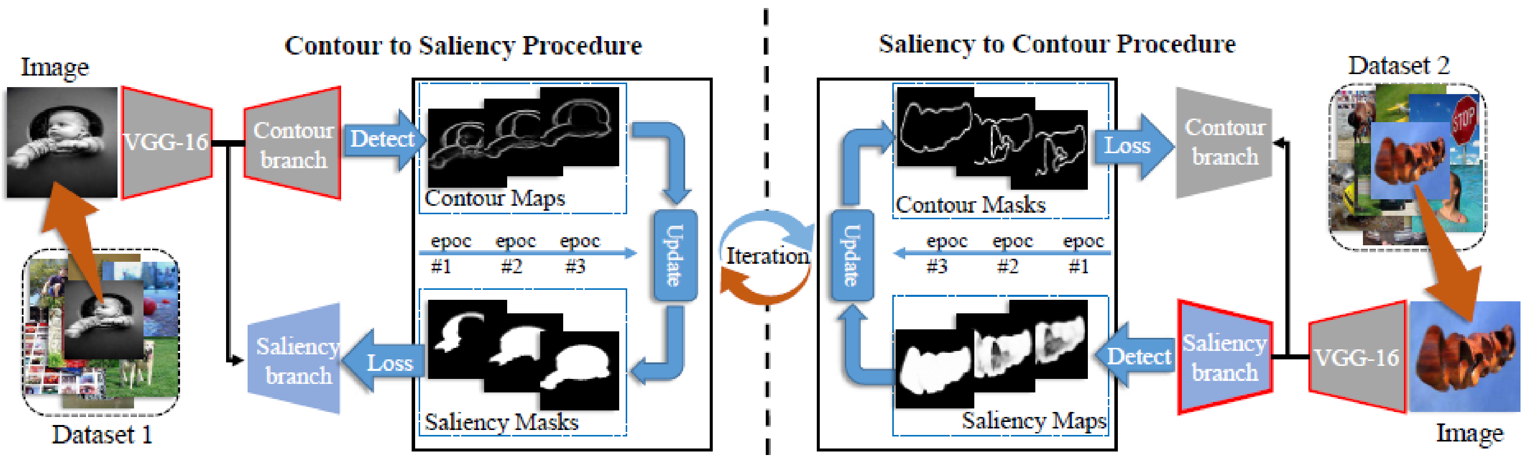
- **Reverse Attention for Salient Object Detection**

	Training Dataset	Training #Images	MSRA-B	HKU-IS	ECSSD	PASCAL-S	SOD	DUT-OMRON
DRFI [13]	MB	2.5k	0.851	0.775	0.784	0.690	0.699	0.664
			0.123	0.146	0.172	0.210	0.223	0.150
DCL+ [22]	MB	2.5k	0.918	0.907	0.898	0.810	0.831	0.757
			0.047	0.048	0.071	0.115	0.131	0.080
DHS[26]	MK+D	9.5k×12	-	0.892	0.905	0.824	0.823	-
			-	0.052	0.061	**0.094**	0.127	-
SSD[16]	MB	2.5k	0.902	-	0.865	0.774	0.793	0.754
			0.160	-	0.193	0.220	0.222	0.193
RFCN[39]	MK	10k	-	0.894	0.889	0.829	0.799	0.744
			-	0.088	0.109	0.133	0.169	0.111
DLS[10]	MK	10k	-	0.835	0.852	0.753	-	0.687
			-	0.070	0.088	0.132	-	0.090
NLDF[30]	MB	2.5k×2	0.911	0.902	0.903	0.826	0.837	0.753
			0.048	0.048	0.065	0.099	0.123	0.080
Amulet[45]	MK	10k×8	-	0.899	0.914	0.832	0.795	0.743
			-	0.050	0.061	0.100	0.144	0.098
UCF[46]	MK	10k×8	-	0.888	0.902	0.818	0.805	0.730
			-	0.061	0.071	0.116	0.148	0.120
DSS[8]	MB	2.5k×2	0.920	0.900	0.908	0.826	0.834	0.764
			0.043	0.050	0.063	0.102	0.126	0.072
DSS+[8]	MB	2.5k×2	**0.929**	**0.916**	**0.919**	**0.835**	**0.843**	**0.781**
			0.034	**0.040**	**0.055**	**0.095**	**0.122**	**0.063**
Ours w/o RA	MB	2.5k×2	0.919	0.898	0.905	0.818	0.839	0.762
			0.042	0.049	0.063	0.106	0.126	0.071
Ours	MB	2.5k×2	**0.931**	**0.913**	**0.918**	**0.834**	**0.844**	**0.786**
			0.036	**0.045**	**0.059**	0.104	**0.124**	**0.062**

Salient Object Detection

➢ Contour Knowledge Transfer for Salient Object Detection

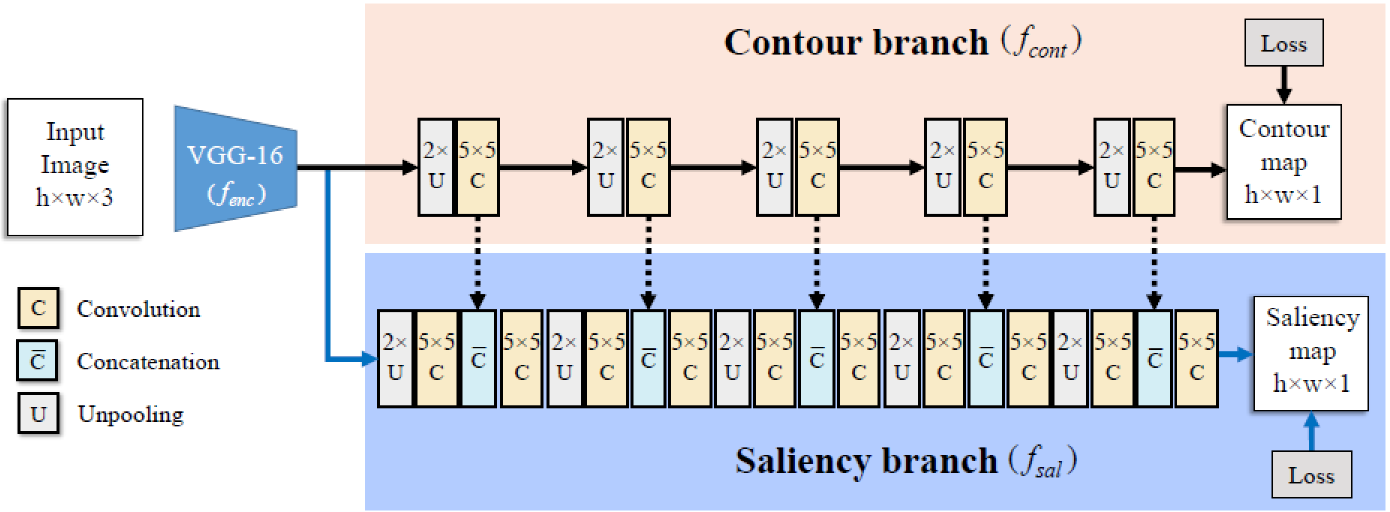
- Automatically convert an existing deep contour detection model into a salient object detection model without using any manual salient object masks
- An alternating training pipeline to update the network parameters



Salient Object Detection

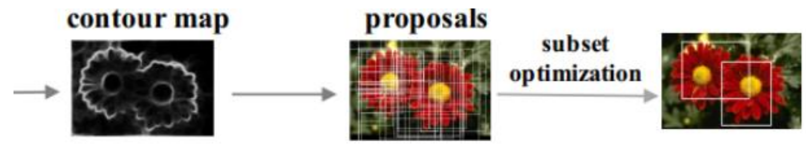
- **Contour Knowledge Transfer for Salient Object Detection**

$$\min_{\theta_c} \sum_i e_{cont}(\mathcal{L}_{cont}(\mathcal{I}_i), C(\mathcal{F}_i; \theta_c))$$



$$\min_{\theta_s} \sum_i e_{sal}(\mathcal{L}_{sal}(\mathcal{I}_i), S(\mathcal{F}_i; \theta_s))$$

Salient Object Detection



Contour Knowledge Transfer for Salient Object Detection

- Contour to Saliency:

  utilize a large collection of unlabeled images to generate corresponding salient object masks, via Multiscale Combinatorial Grouping (MCG)

$$\max_{\mathcal{B}}\{S(\mathcal{B}) - \alpha \cdot O(\mathcal{B}) - \kappa \cdot N(\mathcal{B})\} \quad s.t. \quad \mathcal{B} \subseteq \mathcal{C}$$

$$\Rightarrow \max\{\sum_{b_i \subseteq C} S_i c_i - \alpha \cdot \sum_{\substack{b_i, b_j \in C \\ i \neq j}} K(b_i, b_j) c_i c_j - \kappa \cdot \sum_{b_i \subseteq C} c_i\} \quad s.t. \quad c_i, c_j = 0 \quad or \quad 1$$

- Saliency to Contour:

  compute gradient on the binary region mask

- Alternating Training:

  use two different sets of unlabeled images (M and N) to interactively train the saliency branch and contour branch

WTA: $S_i = K(cnt(b_i), C^{e_r}) + \gamma \cdot K(b_i, S^{e_r})$

Salient Object Detection

set lambda = 0 in the first epoch, and 1 in the following epochs

Contour Knowledge Transfer for Salient Object Detection

Table 1. Analysis of the proposed method. Our results are obtained on ECSSD. "CDC" denotes the cross domain connections that used in our C2S-Net. "AVG-P" means the two-stage strategy, "WTA" denotes the "winner-take-all" strategy, and "CTS" refers to the contour-to-saliency transferring method used in this paper. "SCJ" denotes that we optimize the parameters of two branches jointly, and "AT$_{(i)}$" means that *i-th* alternating training iterations are used to update network parameters. "† " denotes the model used in this paper for comparing with fully supervised models. Weighted F-measure (F^w_β): the higher the better; MAE: the lower the better.

Method	data/annotations	F^w_β	MAE
C2S-Net	5K w/ masks	0.793	0.103
C2S-Net + CDC	5K w/ masks	0.812	0.081
C2S-Net + CDC + AVG-P	5K w/o masks	0.665	0.121
C2S-Net + CDC + WTA	5K w/o masks	0.732	0.112
C2S-Net + CDC + CTS	5K w/o masks	0.743	0.093
C2S-Net + CDC + CTS + SCJ	10K w/o masks	0.759	0.088
C2S-Net + CDC + CTS + AT$_{(1)}$	10K w/o masks	0.778	0.080
C2S-Net + CDC + CTS + AT$_{(3)}$	10K w/o masks	0.837	0.059
C2S-Net + CDC + CTS + AT$_{(5)}$	10K w/o masks	0.838	0.059
C2S-Net + CDC + CTS + AT$_{(3)}$	20K w/o masks	0.849	0.056
† C2S-Net + CDC + CTS + AT$_{(3)}$	30K w/o masks	0.852	0.054

Salient Object Detection

➢ Contour Knowledge Transfer for Salient Object Detection

Table 2. Quantitative comparisons with 10 leading CNN-Based methods on five widely-used benchmarks. The top three results are shown in Red, Blue, and Green, respectively. F_β: the higher the better; MAE: the lower the better.

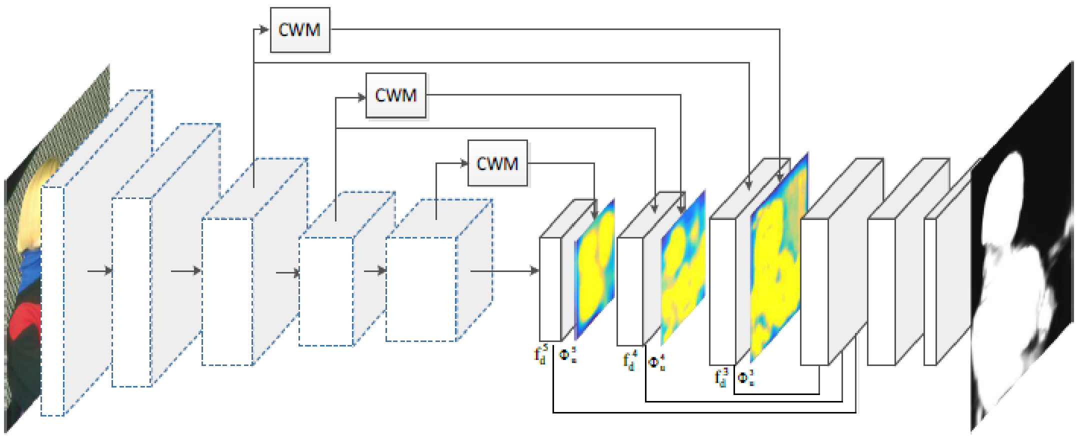
Methods	ECSSD		PASCAL-S		DUT		HKU-IS		DUTS-TE	
	F_β	MAE	F_β	MAE	F_β	MAE	F_β	MAE	F_β	MAE
SBF [37]	0.852	0.880	0.765	0.130	0.685	0.108	0.842	0.075	0.698	0.107
WSS [29]	0.856	0.103	0.770	0.139	0.689	0.110	0.860	0.079	0.737	0.100
Ours(10K)	0.896	0.059	0.835	0.086	0.733	0.079	0.883	0.051	0.790	0.066
MC [39]	0.822	0.107	0.721	0.147	0.703	0.088	0.781	0.098	-	-
MDF [14]	0.832	0.105	0.759	0.142	0.694	0.092	0.860	0.129	0.768	0.099
DS [19]	0.882	0.122	0.757	0.172	0.716	0.120	0.866	0.079	0.776	0.090
ELD [12]	0.869	0.098	0.777	0.121	0.720	0.091	0.767	0.071	0.758	0.097
DHS [23]	0.902	0.061	0.820	0.092	-	-	0.892	0.052	0.812	0.065
DCL [15]	0.887	0.072	0.798	0.109	0.718	0.094	0.879	0.059	0.771	0.079
DSS [8]	0.903	0.062	0.821	0.101	0.761	0.074	0.899	0.051	0.813	0.064
UCF [38]	0.910	0.078	0.819	0.127	0.735	0.132	0.885	0.074	0.771	0.117
Amulet [37]	0.915	0.059	0.828	0.100	0.743	0.098	0.895	0.052	0.778	0.085
Ours(30K)	0.910	0.054	0.846	0.081	0.757	0.071	0.896	0.048	0.807	0.062

Salient Object Detection

- **Detect Globally, Refine Locally: A Novel Approach to Saliency Detection**
  - Directly applying concatenation or element-wise operation to different feature maps are suboptimal (is cluttered)
  - A spatial response map to adaptively weight the features maps for each position
  - Consider the relations between the center point and its n × n neighbors
  - Recurrent Localization Network + Boundary Refinement Network

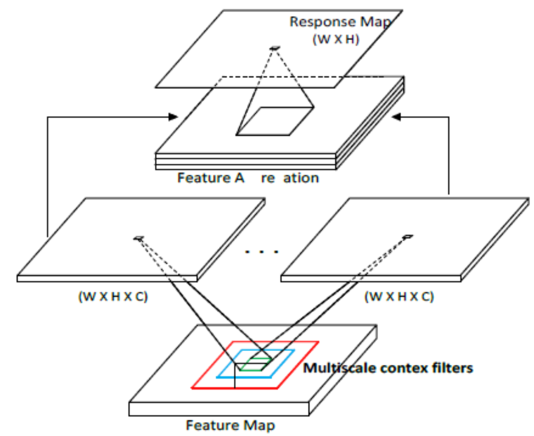
$$\mathbf{\Phi}^k(x,y) = \frac{\exp(\mathbf{M}^k(x,y))}{\sum_{(x',y')}\exp(\mathbf{M}^k(x',y'))}$$

Contextual Weighting Module



Recurrent Module

Recurrent Localization Network

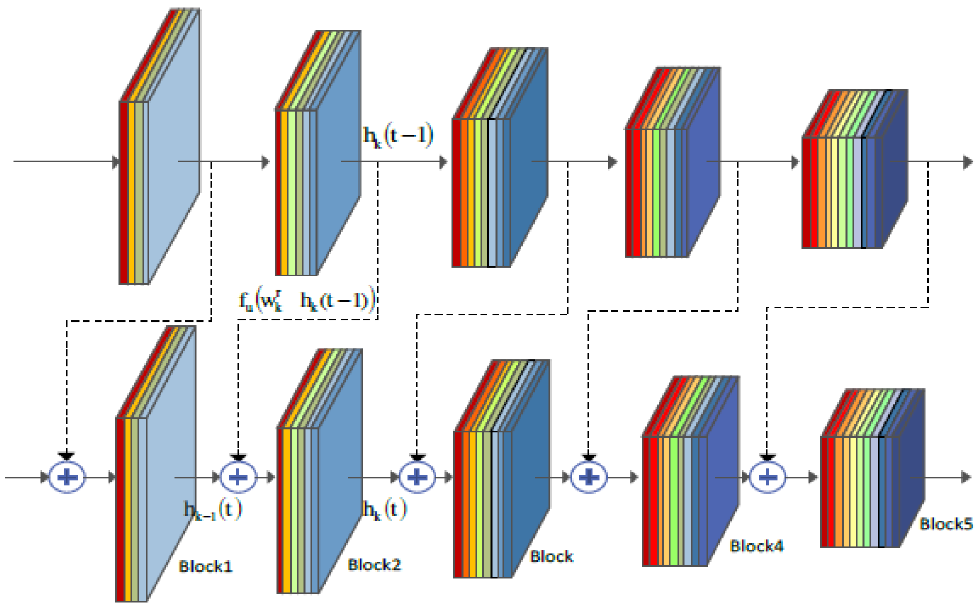


kernel sizes (3×3, 5×5, 7×7)

CVPR'18

Salient Object Detection

- **Detect Globally, Refine Locally: A Novel Approach to Saliency Detection**

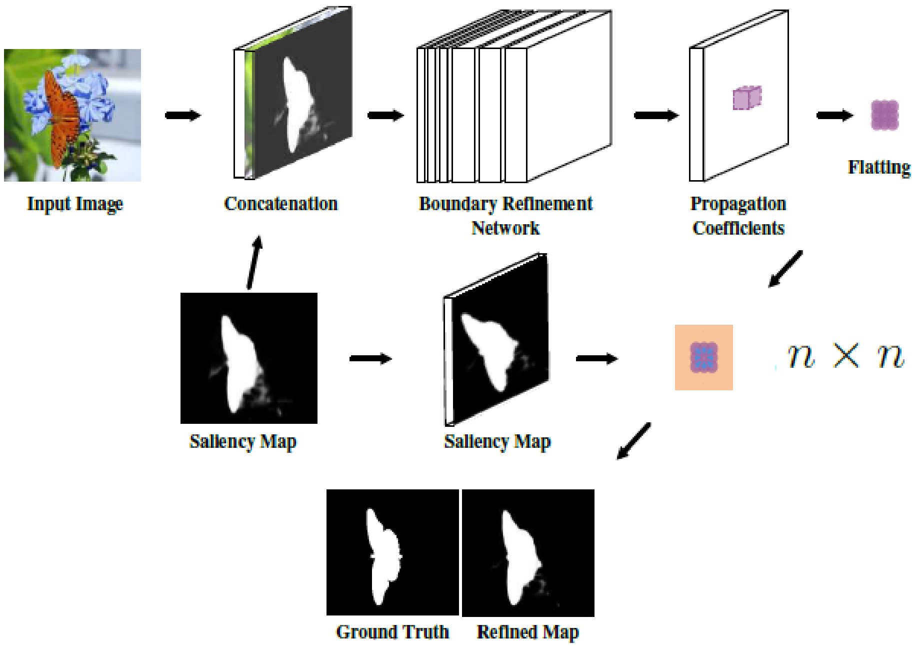


Recurrent Module

- Absorb the contextual and structural information with the hidden convolution units
- Increase the depth of traditional CNNs without increasing the number of parameters

Salient Object Detection

- **Detect Globally, Refine Locally: A Novel Approach to Saliency Detection**



Layer	Channel	Kernel size	Bias size
1	64	$(K+3) \times 64 \times 3 \times 3$	64
2	64	$64 \times 64 \times 3 \times 3$	64
3	64	$64 \times 64 \times 3 \times 3$	64
4	128	$64 \times 128 \times 3 \times 3$	128
5	128	$128 \times 128 \times 3 \times 3$	128
6	128	$128 \times 128 \times 3 \times 3$	128
7	$n \times n$	$128 \times (n \times n) \times 3 \times 3$	$n \times n$

$$\mathbf{s}_i' = \sum_{d=1}^{n \times n} \mathbf{v}_i^d \cdot \mathbf{s}_i^d, d \in 1, 2, ..., n \times n$$

Salient Object Detection

- **Detect Globally, Refine Locally: A Novel Approach to Saliency Detection**

*	ECSSD [31]		THUR15K [5]		HKU-IS [17]		DUTS [27]		DUT-OMRON [32]	
	F-measure	MAE	F-measure	MAE	F-measure	MAE	F-measure	MAE	F-measure	MAE
Ours	0.903	0.045	0.716	0.077	0.882	0.037	0.768	0.051	0.709	0.063
SRM [29]	0.892	0.056	0.708	0.077	0.874	0.046	0.757	0.059	0.707	0.069
Amulet [33]	0.869	0.061	0.670	0.094	0.839	0.052	0.676	0.085	0.647	0.098
UCF [34]	0.841	0.080	0.645	0.112	0.808	0.074	0.629	0.117	0.613	0.132
KSR [30]	0.782	0.135	0.604	0.123	0.747	0.120	0.602	0.121	0.591	0.131
RFCN [28]	0.834	0.109	0.627	0.100	0.835	0.089	0.712	0.090	0.627	0.111
DS [20]	0.821	0.124	0.626	0.116	0.785	0.078	0.632	0.091	0.603	0.120
DCL [18]	0.827	0.151	0.676	0.161	0.853	0.136	0.714	0.149	0.684	0.157
DHS [22]	0.871	0.063	0.673	0.082	0.852	0.054	0.724	0.067	-	-
LEGS [26]	0.785	0.119	0.607	0.125	0.732	0.119	0.585	0.138	0.592	0.133
MCDL [36]	0.796	0.102	0.620	0.103	0.757	0.092	0.594	0.105	0.625	0.089
MDF [17]	0.805	0.108	0.636	0.109	-	-	0.673	0.100	0.644	0.092
BL [25]	0.684	0.217	0.532	0.219	0.660	0.207	0.490	0.238	0.499	0.239
DRFI [12]	0.733	0.166	0.576	0.150	0.722	0.145	0.541	0.175	0.550	0.138

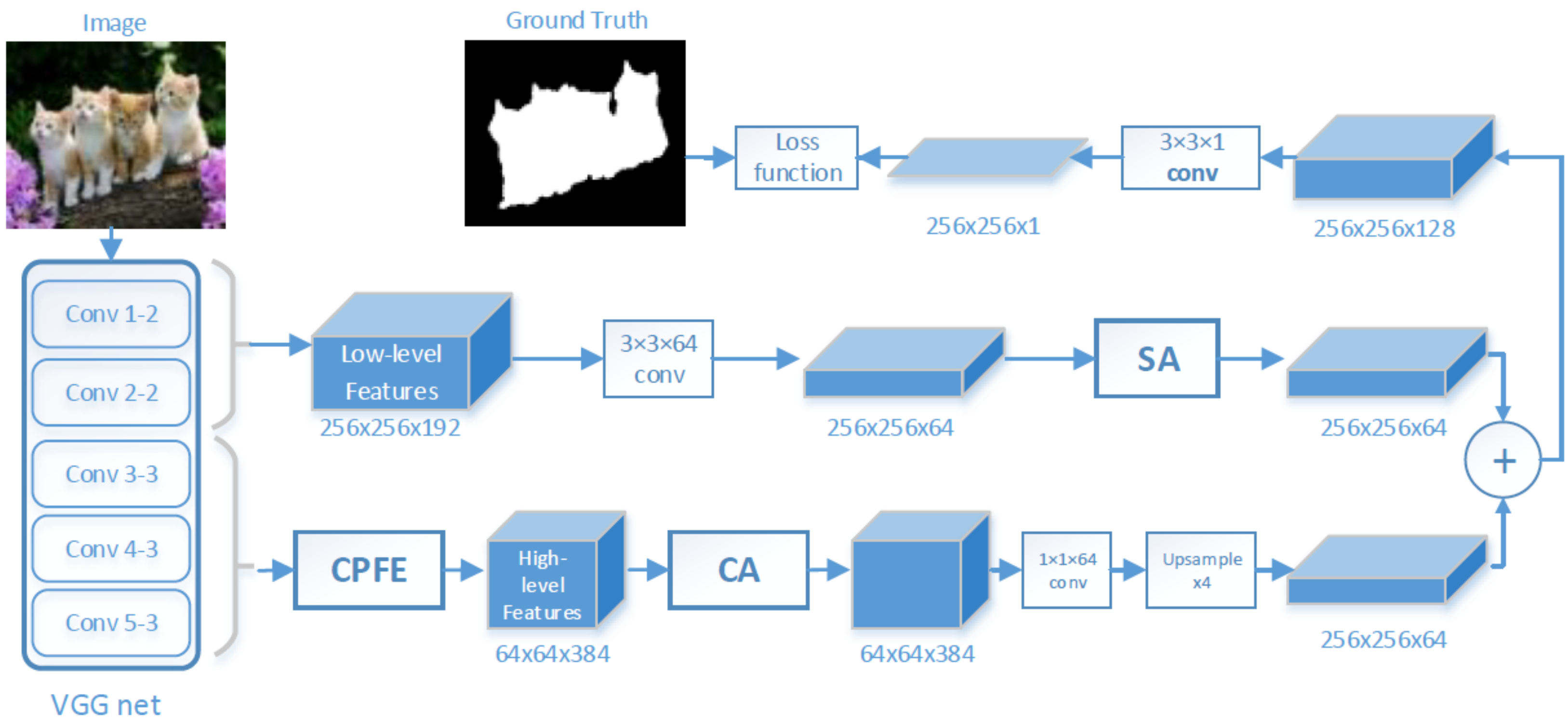
Table 1. Quantitative evaluation in terms of F-measure and MAE scores. The best two scores are shown in red and blue colors, respectively.

*	ECSSD		THUR15K		HKU-IS		DUTS		DUT-OMRON	
	F-measure	MAE	F-measure	MAE	F-measure	MAE	F-measure	MAE	F-measure	MAE
Baseline	0.861	0.058	0.659	0.099	0.838	0.050	0.696	0.073	0.643	0.092
CWM	0.867	0.054	0.667	0.084	0.840	0.047	0.716	0.060	0.661	0.075
RM	0.893	0.048	0.702	0.080	0.875	0.041	0.760	0.054	**0.712**	0.066
BRN	**0.903**	**0.045**	**0.716**	**0.077**	**0.882**	**0.037**	**0.768**	**0.051**	0.709	**0.063**

Table 3. Performance of the proposed modules.

Salient Object Detection

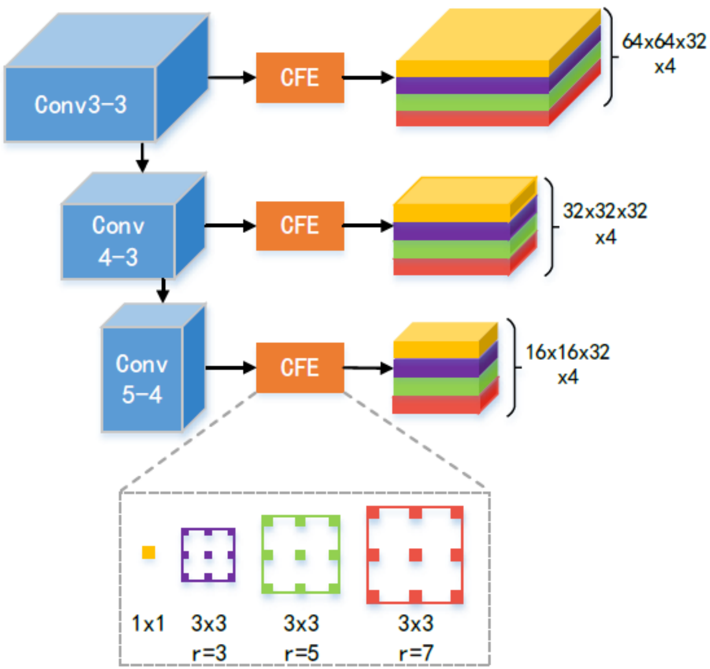
- **Pyramid Feature Attention Network for Saliency Detection**
  - ASPP + Channel Attention Block (*CVPR'18*), actually
  - Edge information as the previous works
  - Impressive performance



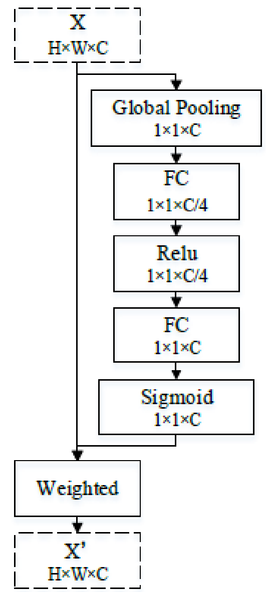
CVPR'19

Salient Object Detection

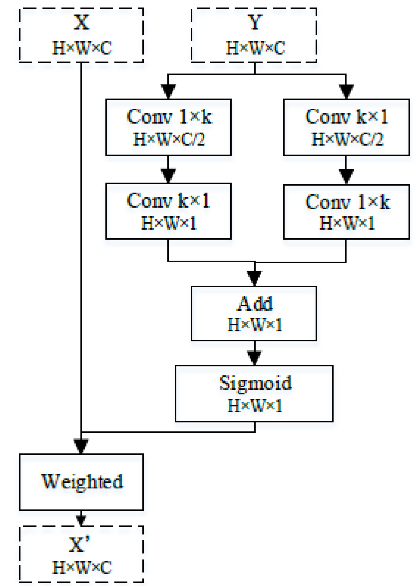
- **Pyramid Feature Attention Network for Saliency Detection**



context-aware feature extraction module (CPFE)



Channel-wise Attention



Spacial Attention

Salient Object Detection

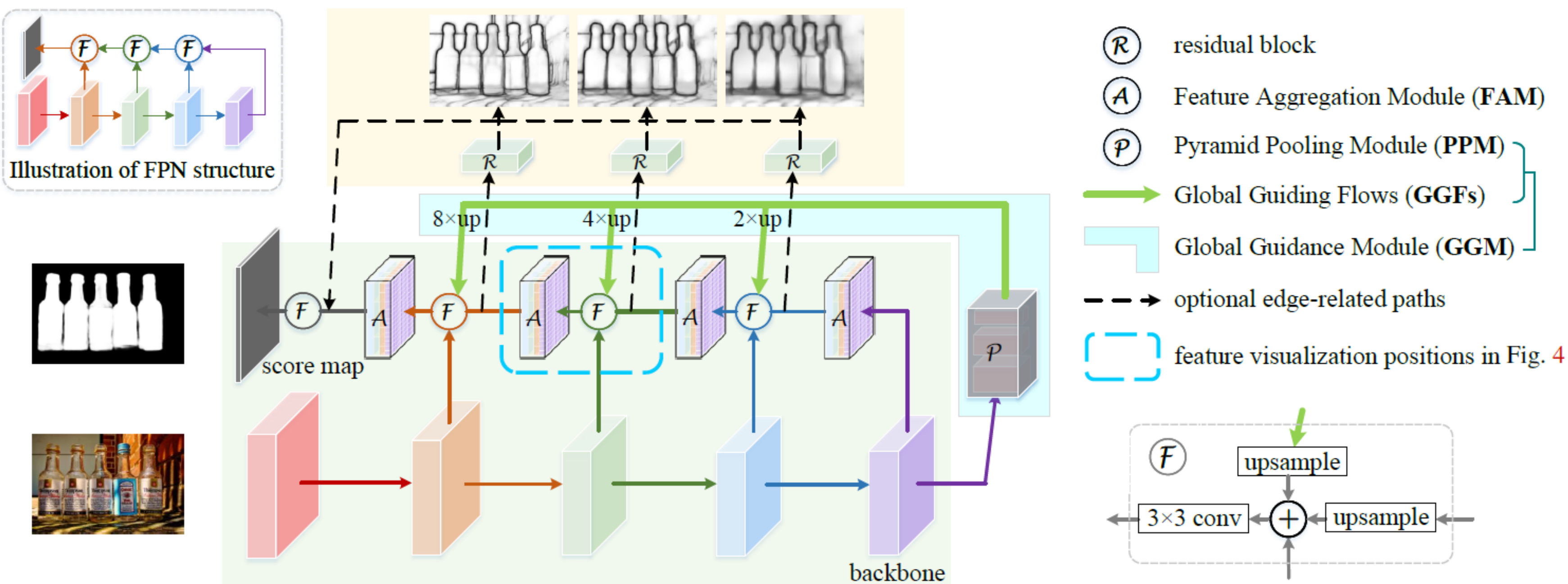
- **Pyramid Feature Attention Network for Saliency Detection**

Table 1. The wF_β and MAE of different salient object detection approaches on all test datasets. The best three results are shown in red, blue, and green.

Methods	DUTS-test		ECSSD		HKU-IS		PASCAL-S		DUT-OMRON	
	wF_β	MAE	wF_β	MAE	wF_β	MAE	wF_β	MAE	wF_β	MAE
Ours	0.8702	0.0405	0.9313	0.0328	0.9264	0.0324	0.8922	0.0677	0.8557	0.0414
BDMPM[42]	0.8508	0.0484	0.9249	0.0478	0.9200	0.0392	0.8806	0.0788	0.7740	0.0635
GRL[33]	0.8341	0.0509	0.9230	0.0446	0.9130	0.0377	0.8811	0.0799	0.7788	0.0632
PAGRN[45]	0.8546	0.0549	0.9237	0.0643	0.9170	0.0479	0.8690	0.0940	0.7709	0.0709
Amulet[43]	0.7773	0.0841	0.9138	0.0604	0.8968	0.0511	0.8619	0.0980	0.7428	0.0976
SRM[32]	0.8269	0.0583	0.9158	0.0564	0.9054	0.0461	0.8677	0.0859	0.7690	0.0694
UCF[44]	0.7723	0.1112	0.9018	0.0704	0.8872	0.0623	0.8492	0.1099	0.7296	0.1203
DCL[20]	0.7857	0.0812	0.8959	0.0798	0.8899	0.0639	0.8457	0.1115	0.7567	0.0863
DHS[22]	0.8114	0.0654	0.9046	0.0622	0.8901	0.0532	0.8456	0.0960	-	-
DSS[15]	0.8135	0.0646	0.8959	0.0647	0.9011	0.0476	0.8506	0.0998	0.7603	0.0751
ELD[18]	0.7372	0.0924	0.8674	0.0811	0.8409	0.0734	0.7882	0.1228	0.7195	0.0909
NLDF[24]	0.8125	0.0648	0.9032	0.0654	0.9015	0.0481	0.8518	0.1004	0.7532	0.0796
RFCN[31]	0.7826	0.0893	0.8969	0.0972	0.8869	0.0806	0.8554	0.1159	0.7381	0.0945

Salient Object Detection

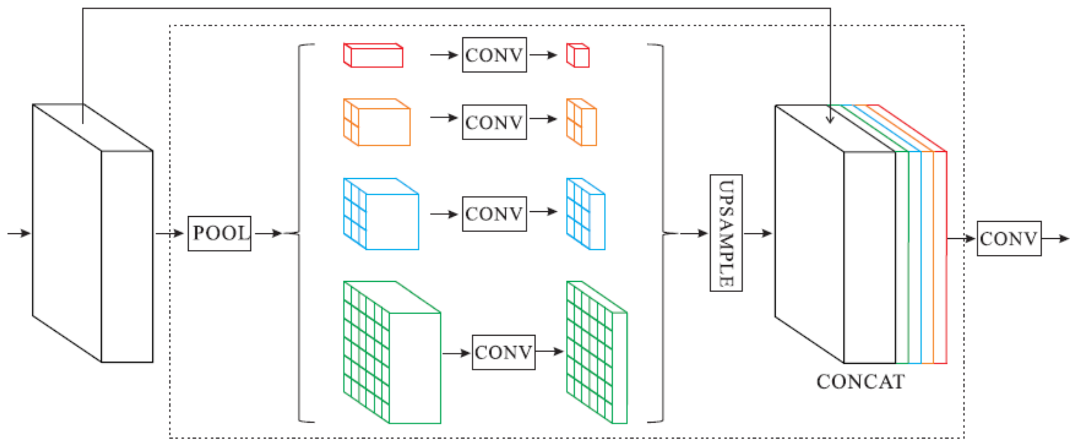
- **A Simple Pooling-Based Design for Real-Time Salient Object Detection**
  - Use edge detection dataset, train alternatively
  - Use PSP / modified PSP blocks



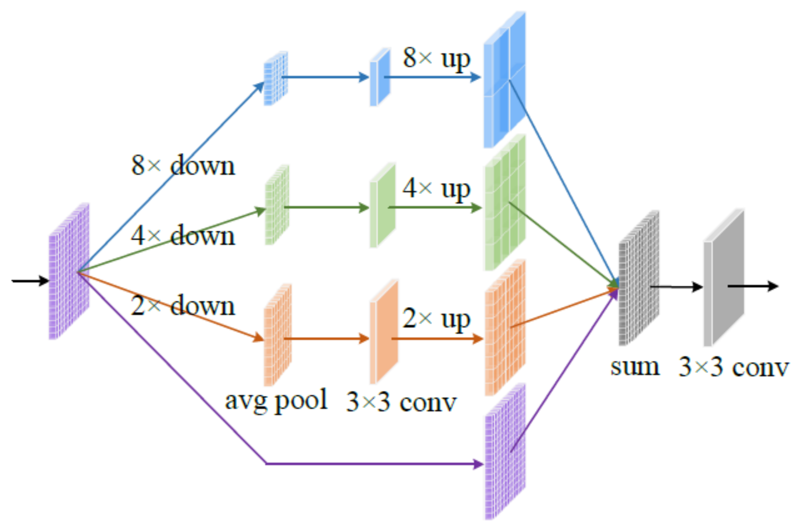
Salient Object Detection

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

PSPNet



FAM



Salient Object Detection

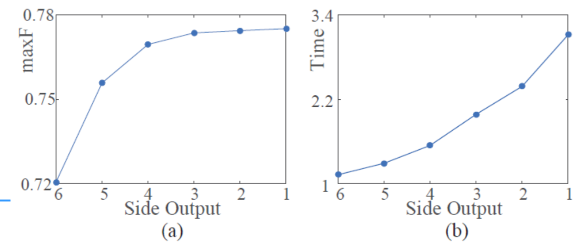
- **A Simple Pooling-Based Design for Real-Time Salient Object Detection**

Model	Training #Images	Training Dataset	ECSSD [41] MaxF ↑	ECSSD [41] MAE ↓	PASCAL-S [21] MaxF ↑	PASCAL-S [21] MAE ↓	DUT-O [42] MaxF ↑	DUT-O [42] MAE ↓	HKU-IS [18] MaxF ↑	HKU-IS [18] MAE ↓	SOD [30] MaxF ↑	SOD [30] MAE ↓	DUTS-TE [35] MaxF ↑	DUTS-TE [35] MAE ↓
VGG-16 backbone														
DCL [19]	2,500	MB	0.896	0.080	0.805	0.115	0.733	0.094	0.893	0.063	0.831	0.131	0.786	0.081
RFCN [36]	10,000	MK	0.898	0.097	0.827	0.118	0.747	0.094	0.895	0.079	0.805	0.161	0.786	0.090
DHS [23]	9,500	MK+DTO	0.905	0.062	0.825	0.092	-	-	0.892	0.052	0.823	0.128	0.815	0.065
MSR [17]	5,000	MB + H	0.903	0.059	0.839	0.083	0.790	0.073	0.907	0.043	0.841	0.111	0.824	0.062
DSS [9]	2,500	MB	0.906	0.064	0.821	0.101	0.760	0.074	0.900	0.050	0.834	0.125	0.813	0.065
NLDF [28]	3,000	MB	0.903	0.065	0.822	0.098	0.753	0.079	0.902	0.048	0.837	0.123	0.816	0.065
UCF [45]	10,000	MK	0.908	0.080	0.820	0.127	0.735	0.131	0.888	0.073	0.798	0.164	0.771	0.116
Amulet [44]	10,000	MK	0.911	0.062	0.826	0.092	0.737	0.083	0.889	0.052	0.799	0.146	0.773	0.075
GearNet[10]	5,000	MB + H	0.923	0.055	-	-	0.790	0.068	0.934	0.034	0.853	0.117	-	-
PAGR [46]	10,553	DTS	0.924	0.064	0.847	0.089	0.771	0.071	0.919	0.047	-	-	0.854	0.055
PiCANet [24]	10,553	DTS	0.930	0.049	0.858	0.078	0.815	0.067	0.921	0.042	0.863	**0.102**	0.855	0.053
PoolNet (Ours)	2,500	MB	0.918	0.057	0.828	0.098	0.783	0.065	0.908	0.044	0.846	0.124	0.819	0.062
PoolNet (Ours)	5,000	MB + H	0.930	0.053	0.838	0.093	0.806	0.063	**0.936**	**0.032**	0.861	0.118	0.855	0.053
PoolNet (Ours)	10,553	DTS	0.936	0.047	0.857	0.078	0.817	0.058	0.928	0.035	0.859	0.115	0.876	0.043
PoolNet† (Ours)	10,553	DTS	**0.937**	**0.044**	**0.865**	**0.072**	**0.821**	**0.056**	0.931	0.033	**0.866**	0.105	**0.880**	**0.041**
ResNet-50 backbone														
SRM [37]	10,553	DTS	0.916	0.056	0.838	0.084	0.769	0.069	0.906	0.046	0.840	0.126	0.826	0.058
DGRL [38]	10,553	DTS	0.921	0.043	0.844	0.072	0.774	0.062	0.910	0.036	0.843	0.103	0.828	0.049
PiCANet [24]	10,553	DTS	0.932	0.048	0.864	0.075	0.820	0.064	0.920	0.044	0.861	0.103	0.863	0.050
PoolNet (Ours)	10,553	DTS	0.940	0.042	0.863	0.075	0.830	0.055	0.934	0.032	0.867	**0.100**	0.886	0.040
PoolNet† (Ours)	10,553	DTS	**0.945**	**0.038**	**0.880**	**0.065**	**0.833**	**0.053**	**0.935**	**0.030**	**0.882**	0.102	**0.892**	**0.036**

MB: MSRA-B [25], MK: MSRA10K [3], DTO: DUT-OMRON [42], H: HKU-IS [18], DTS: DUTS-TR [35].

Table 3. Quantitative salient object detection results on 6 widely used datasets. The best results with different backbones are highlighted in **blue** and **red**, respectively. †: joint training with edge detection. As can be seen, our approach achieves the best results on nearly all datasets in terms of F-measure and MAE.

Salient Object Detection



- **Cascaded Partial Decoder for Fast and Accurate Salient Object Detection**

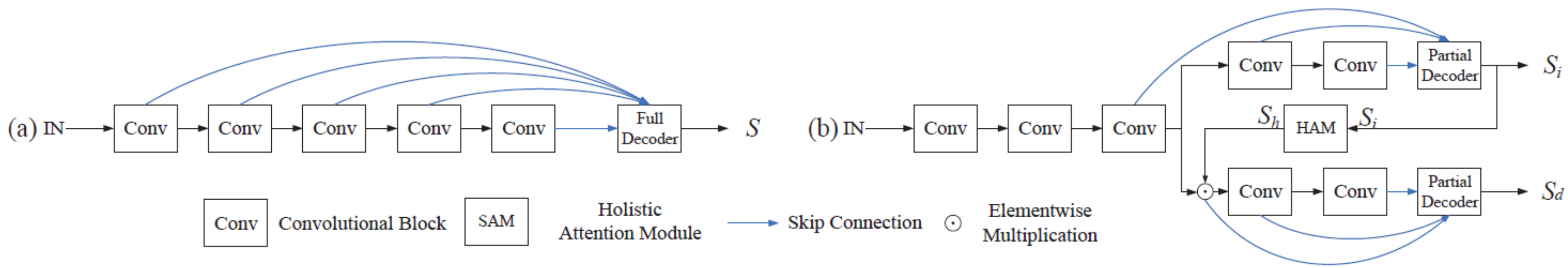


Figure 3: *(a) Traditional encoder-decoder framework, (b) The proposed cascaded partial decoder framework. We use VGG16 [29] as the backbone network. Traditional framework generates saliency map S by adopting full decoder which integrates all level features. The proposed framework adopts partial decoder, which only integrates features of deeper layers, and generates an initial saliency map S_i and the final saliency map S_d.*

Gaussian blur for the attention map:

$$S_h = MAX(f_{min\_max}(Conv_g(S_i, k)), S_i)$$

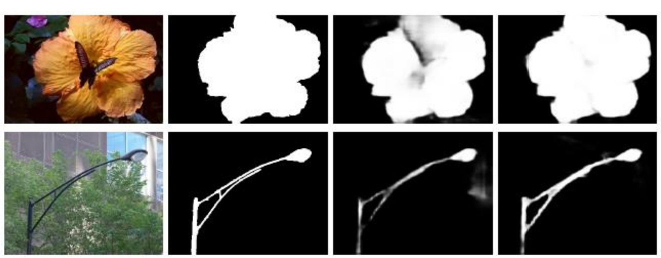
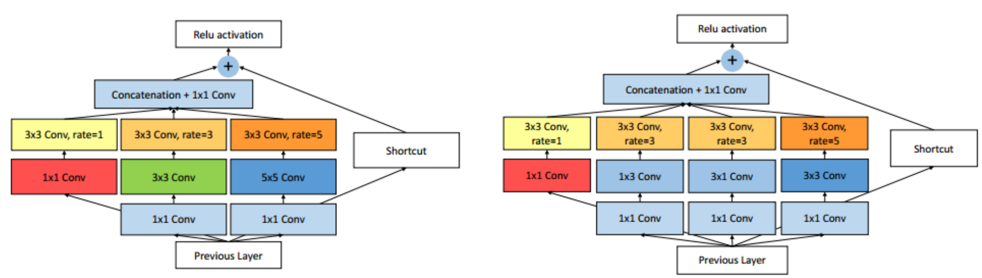


Image GT Initial Attention Holistic Attention

Partial decoder: a RFB-like block



(a) RFB (b) RFB-s

Salient Object Detection

➢ Cascaded Partial Decoder for Fast and Accurate Salient Object Detection

Method	Backbone	FPS	ECSSD [39]			HKU-IS [16]			DUT-OMRON [40]			DUTS [33]			PASCAL-S [19]		
			maxF	avgF	MAE	maxF	avgF	MAE	maxF	avgF	MAE	maxF	avgF	MAE	maxF	avgF	MAE
Amulet [42]	VGG16	21	0.922	0.881	0.057	0.909	0.863	0.047	0.791	0.699	0.072	0.832	0.738	0.062	0.839	0.780	0.095
NLDF [25]	VGG16	20	0.915	0.886	0.051	0.908	0.871	0.041	0.759	0.694	0.071	0.830	0.759	0.055	0.840	0.792	0.083
DSS [9]	VGG16	23	0.928	0.889	0.051	0.915	0.867	0.043	0.781	0.692	0.065	0.858	0.757	0.050	0.859	0.796	0.081
BMPM [41]	VGG16	28	0.928	0.894	0.044	0.920	0.875	0.039	0.775	0.693	0.063	0.850	0.768	0.049	0.862	0.770	**0.074**
PAGR [43]	VGG19	-	0.927	0.894	0.061	0.918	0.886	0.048	0.771	0.711	0.072	0.855	0.788	0.055	0.851	0.803	0.092
PiCANet [21]	VGG16	7	0.931	0.885	0.046	0.921	0.870	0.042	**0.794**	0.710	0.068	0.851	0.749	0.054	0.862	0.796	0.076
CPD-A (ours)	VGG16	**105**	0.928	0.906	0.045	0.918	0.884	0.037	0.781	0.721	0.061	0.854	0.787	0.047	0.859	0.814	0.077
CPD (ours)	VGG16	66	**0.936**	**0.915**	**0.040**	**0.924**	**0.896**	**0.033**	**0.794**	**0.745**	**0.057**	**0.864**	**0.813**	**0.043**	**0.866**	**0.825**	**0.074**
SRM [35]	ResNet50	37	0.917	0.892	0.054	0.903	0.871	0.047	0.769	0.707	0.069	0.827	0.757	0.059	0.847	0.796	0.085
DGRL [36]	ResNet50	6	0.925	0.903	0.043	0.914	0.882	0.037	0.779	0.709	0.063	0.834	0.764	0.051	0.853	0.807	0.074
PiCANet-R [21]	ResNet50	5	0.935	0.886	0.046	0.919	0.870	0.043	**0.803**	0.717	0.065	0.860	0.759	0.051	0.863	0.798	0.075
CPD-RA (ours)	ResNet50	**104**	0.934	0.907	0.043	0.918	0.882	0.038	0.783	0.725	0.059	0.852	0.776	0.048	0.855	0.807	0.077
CPD-R (ours)	ResNet50	62	**0.939**	**0.917**	**0.037**	**0.925**	**0.891**	**0.034**	0.797	**0.747**	**0.056**	**0.865**	**0.805**	**0.043**	**0.864**	**0.824**	**0.072**

Table 1: *Comparison of different methods on five benchmark datasets and four metrics including FPS, MAE (lower is better), max F-measure (higher is better) and average F-measure. The comparison is under two settings (with VGG [29] and ResNet50 [8] backbone netowrk). The best result of each setting is shown in* ***Red****. "-R" means using ResNet50 as the backbone. "-A" means the results of the attention branch. All method are the trained on training set of DUTS [33]. There is not available code of PAGR [43] and the author only provides the saliency maps.*

Salient Object Detection

➢ **Cascaded Partial Decoder for Fast and Accurate Salient Object Detection**

Method	FPS	ECSSD [39]			HKU-IS [16]			DUT-OMRON [40]			DUTS [33]			PASCAL-S [19]		
		maxF	avgF	MAE	maxF	avgF	MAE	maxF	avgF	MAE	maxF	avgF	MAE	maxF	avgF	MAE
BMPM [41]	28	0.928	0.894	0.044	0.920	0.875	0.039	0.775	0.693	0.063	0.850	0.768	0.049	0.862	0.803	0.074
BMPM-CPD-A	**82**	0.932	0.901	0.046	0.920	0.882	0.037	0.796	0.731	0.057	0.864	0.799	0.046	0.861	0.817	0.074
BMPM-CPD	47	**0.935**	**0.907**	**0.043**	**0.925**	**0.888**	**0.035**	**0.804**	**0.740**	**0.056**	**0.870**	**0.808**	**0.044**	**0.868**	**0.822**	**0.072**
NLDF [25]	21	0.915	0.886	0.051	0.908	0.871	0.041	0.759	0.694	0.071	0.830	0.759	0.055	0.840	0.792	0.083
NLDF-CPD-A	**75**	0.918	0.889	0.049	0.914	0.873	0.039	0.775	0.710	0.061	0.837	0.773	0.050	0.841	0.793	0.083
NLDF-CPD	48	**0.922**	**0.896**	**0.044**	**0.916**	**0.880**	**0.036**	**0.781**	**0.721**	**0.060**	**0.842**	**0.786**	**0.048**	**0.843**	**0.800**	**0.080**
Amulet [42]	21	0.922	0.881	0.057	0.909	0.863	0.047	0.791	0.699	0.072	0.832	0.738	0.062	0.839	0.780	0.095
Amulet-CPD-A	**61**	0.925	0.889	0.053	0.910	0.864	0.045	0.790	0.708	0.070	0.832	0.747	0.060	0.842	0.784	0.091
Amulet-CPD	45	**0.934**	**0.901**	**0.047**	**0.920**	**0.878**	**0.040**	**0.805**	**0.735**	**0.063**	**0.845**	**0.771**	**0.055**	**0.851**	**0.801**	**0.085**

Table 2: *Comparison of the original models and the improved models (-CPD-A and -CPD).*

Salient Object Detection

- Method
  - Supervised
  - Unsupervised

Salient Object Detection

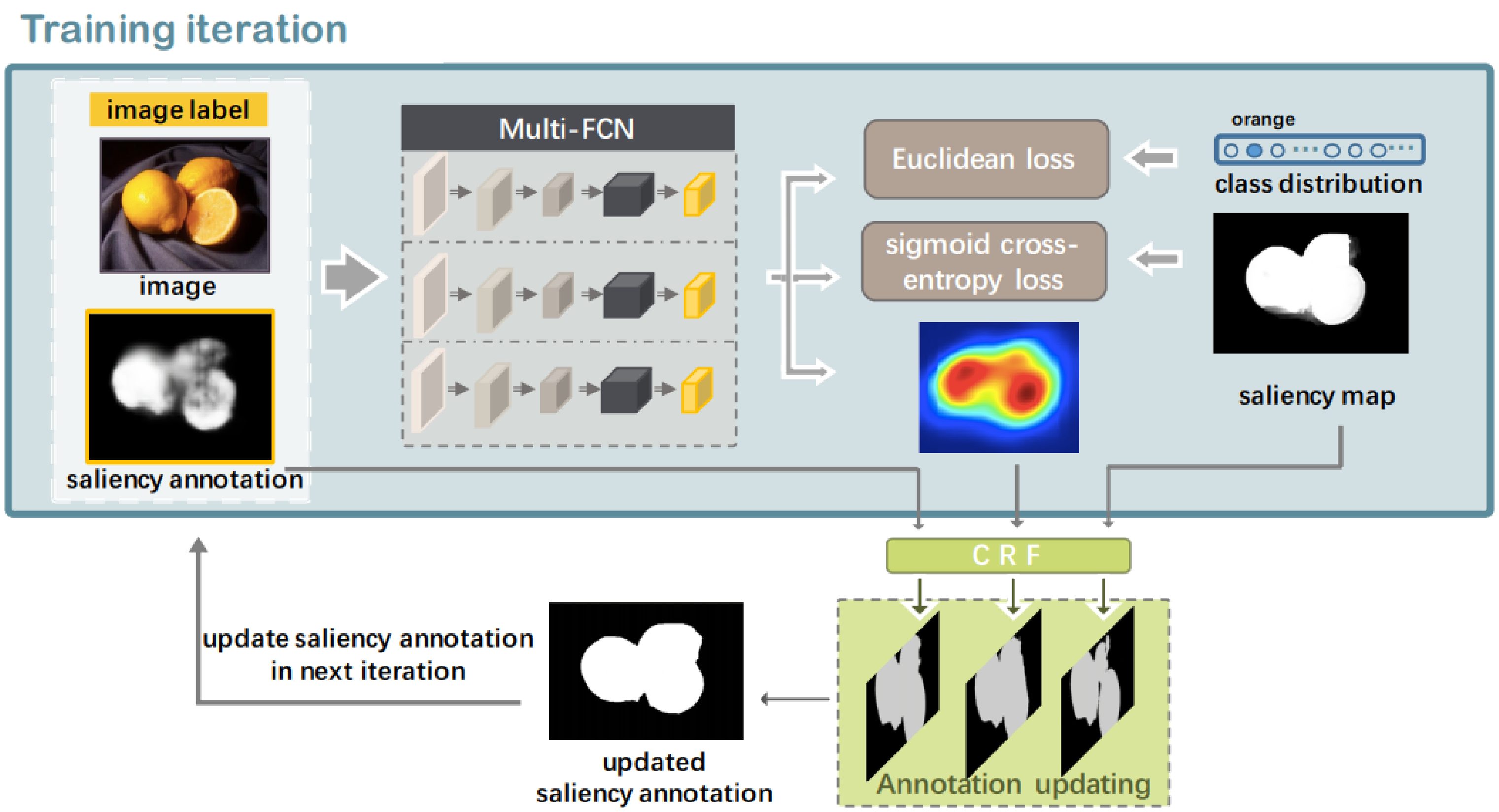
➢ Weakly Supervised Salient Object Detection Using Image Labels

- MB+ generate training saliency maps (hard for low contrast and complex background)
- Multi-FCN simultaneously learns pixel-wise saliency and class distribution
- Initial saliency, predicted saliency and average top-three CAMs map + CRF
- Iteratively training (lowest validation error for each iteration)
- Finetune saliency prediction stream guided by offline CAM without annotations
- Multiple input scales (0.5, 0.75, 1)
- Probability maps are resized to raw size, summed up to get final probability (sigmoid)
- MS COCO with multiple class labels + MSRA-B and HKU-IS without annotations

$$M_c(x,y) = \sum_k w_k^c f_k(x,y)$$

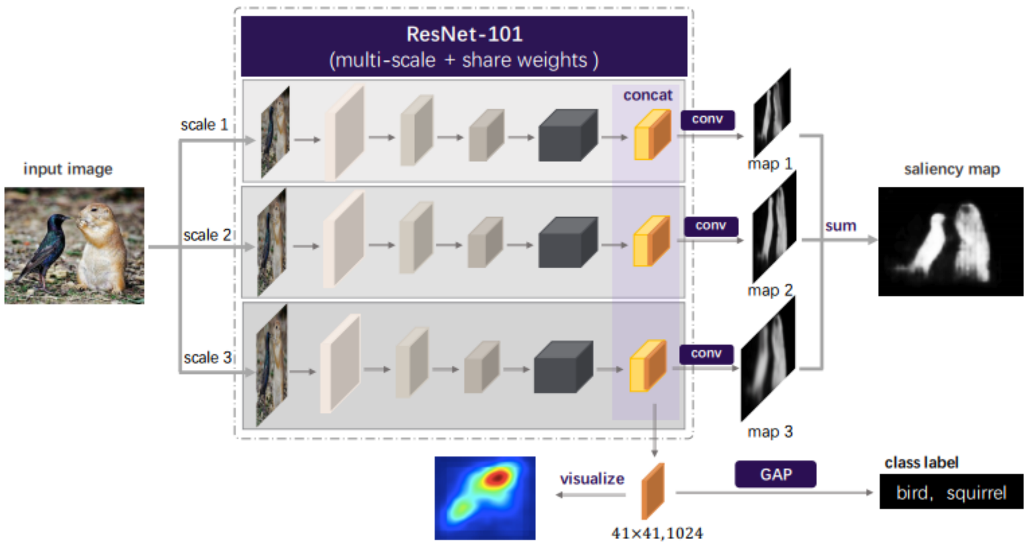
Salient Object Detection

- **Weakly Supervised Salient Object Detection Using Image Labels**



Salient Object Detection

➢ **Weakly Supervised Salient Object Detection Using Image Labels**



Algorithm 1 Saliency Annotations Updating

Require: Current saliency map annotation S_{anno}, the predicted saliency map $S_{predict}$, CRF output of current saliency map annotation C_{anno}, CRF output of the predicted saliency map $C_{predict}$ and CRF output of the class activation map C_{cam}

Ensure: The updated saliency map annotation S_{update}.

1: **if** $\mathrm{MAE}(C_{anno}, C_{predict}) \le \alpha$ **then**

2: $\quad S_{update} = \mathrm{CRF}\left(\frac{S_{anno}+S_{predict}}{2}\right)$

3: **else if** $\mathrm{MAE}(C_{anno}, C_{cam}) > \beta$ and $\mathrm{MAE}(C_{predict}, C_{cam}) > \beta$ **then**

4: $\quad$ Discard the training sample in next iteration

5: **else if** $\mathrm{MAE}(C_{anno}, C_{cam}) \le \mathrm{MAE}(C_{predict}, C_{cam})$ **then**

6: $\quad S_{update} = C_{anno}$

7: **else**

8: $\quad S_{update} = C_{predict}$

9: **end if**

Salient Object Detection

➢ **Weakly Supervised Salient Object Detection Using Image Labels**

Data Set	Metric	GS	SF	HS	MR	GC	BSCA	MB+	MST	ASMO	ASMO+
MSRA-B	maxF	0.777	0.700	0.813	**0.824**	0.719	0.830	0.822	0.809	**0.890**	**0.896**
	MAE	0.144	0.166	0.161	0.127	0.159	0.130	0.133	**0.098**	**0.067**	**0.068**
ECSSD	maxF	0.661	0.548	0.727	0.736	0.597	**0.758**	0.736	0.724	**0.837**	**0.845**
	MAE	0.206	0.219	0.228	0.189	0.233	0.183	0.193	**0.155**	**0.110**	**0.112**
HKU-IS	maxF	0.682	0.590	0.710	0.714	0.588	0.723	**0.727**	0.707	**0.846**	**0.855**
	MAE	0.166	0.173	0.213	0.174	0.211	0.174	0.180	**0.139**	**0.086**	**0.088**
DUT-OMRON	maxF	0.556	0.495	0.616	0.610	0.495	0.617	**0.621**	0.588	**0.722**	**0.732**
	MAE	0.173	**0.147**	0.227	0.187	0.218	0.191	0.193	0.161	**0.101**	**0.100**
PASCAL-S	maxF	0.620	0.493	0.641	0.661	0.539	0.666	**0.673**	0.657	**0.752**	**0.758**
	MAE	0.223	0.240	0.264	0.223	0.266	0.224	0.228	**0.194**	**0.152**	**0.154**
SOD	maxF	0.620	0.516	0.646	0.636	0.526	0.654	**0.658**	0.647	**0.751**	**0.758**
	MAE	0.251	0.267	0.283	0.259	0.284	0.251	0.255	**0.223**	**0.185**	**0.187**

Table 1: Comparison of quantitative results including maximum F-measure (larger is better) and MAE (smaller is better). The best three results on each dataset are shown in **red**, **blue**, and **green** , respectively.

Table 4: Evaluation of different benchmark methods in alternate saliency map optimization on DUT-OMRON dataset.

Metric	MB+	ASMO (MB+)	BSCA	ASMO (BSCA)	MST	ASMO (MST)
maxF	0.621	0.722	0.617	0.685	0.588	0.691
MAE	0.193	0.101	0.191	0.121	0.161	0.126

Salient Object Detection

➢ **Weakly Supervised Salient Object Detection Using Image Labels**

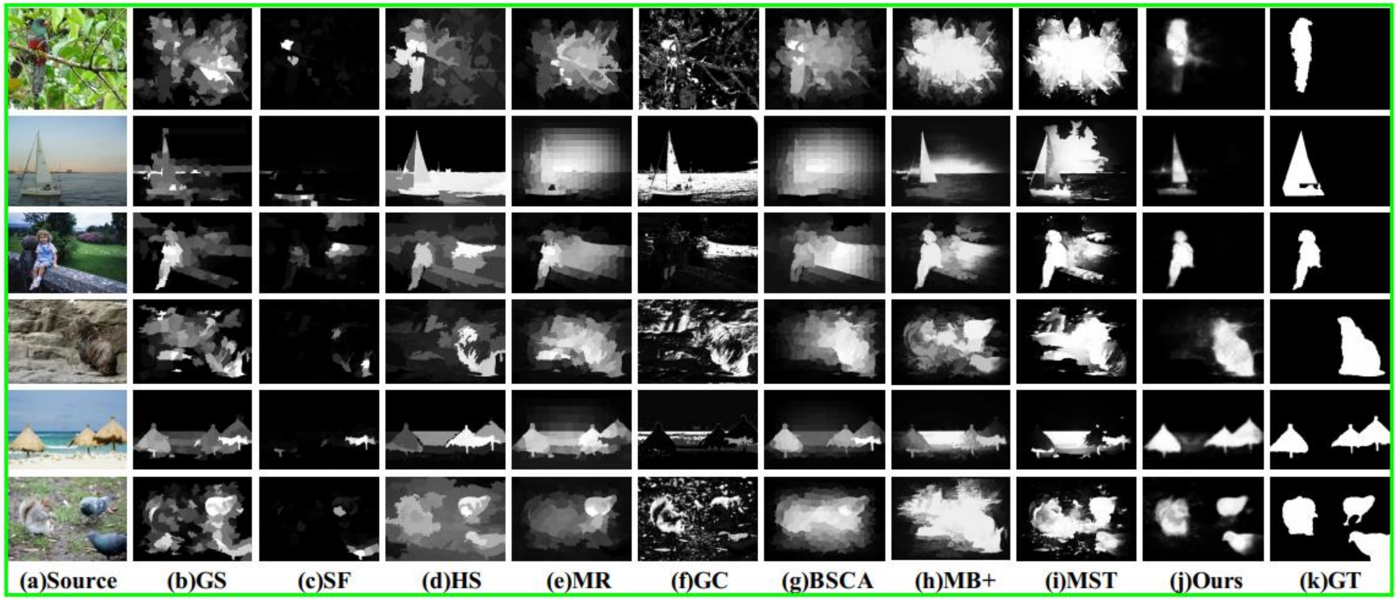


Figure 5: Visual comparison of saliency maps from state-of-the-art methods. The ground truth (GT) is shown in the last column. Our proposed method consistently produces saliency maps closest to the ground truth.

Salient Object Detection

Reference

[1] Salient Object Detection: A Survey, *TPAMI’17*
[2] Deep Networks for Saliency Detection via Local Estimation and Global Search, *CVPR’15*
[3] Visual Saliency Based on Multiscale Deep Features, *CVPR’15*
[4] Deeply Supervised Salient Object Detection with Short Connections, *CVPR’17*
[5] Reverse Attention for Salient Object Detection, *ECCV’18*
[6] Contour Knowledge Transfer for Salient Object Detection, *ECCV’18*
[7] Weakly Supervised Salient Object Detection Using Image Labels, *AAAI’18*
[8] Detect Globally, Refine Locally: A Novel Approach to Saliency Detection, *CVPR’18*
[9] Pyramid Feature Attention Network for Saliency detection, *CVPR’19*
[10] A Simple Pooling-Based Design for Real-Time Salient Object Detection, *CVPR’19*
[11] Cascaded Partial Decoder for Fast and Accurate Salient Object Detection, *CVPR’19*