ECCV'20 ONLINE

23-28 AUGUST 2020

16TH EUROPEAN CONFERENCE ON COMPUTER VISION

WWW.ECCV2020.EU

Hierarchical Dynamic Filtering Network for RGB-D Salient Object Detection

Youwei Pang[1], Lihe Zhang[1*], Xiaoqi Zhao[1], Huchuan Lu[1,2]

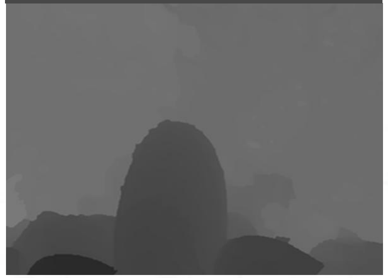
[1]Dalian University of Technology, China

[2]Peng Cheng Laboratory

IIAU-LAB

RGB-D Salient Object Detection

DEPTH IMAGE



RGB IMAGE

Why do we need depth information???

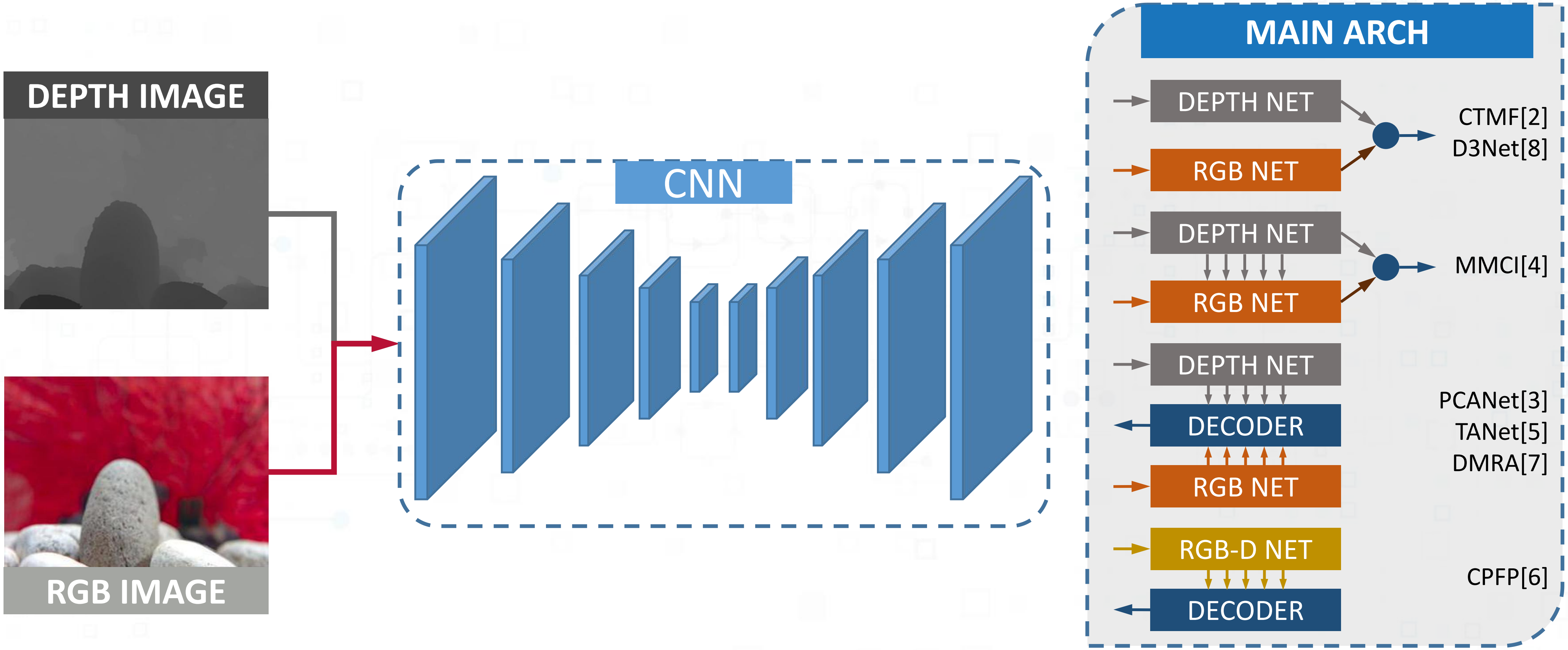
1. provide **rich spatial structure information**
2. deal with the **cluttered or low-contrast scenes**

Two forms of the depth Image:

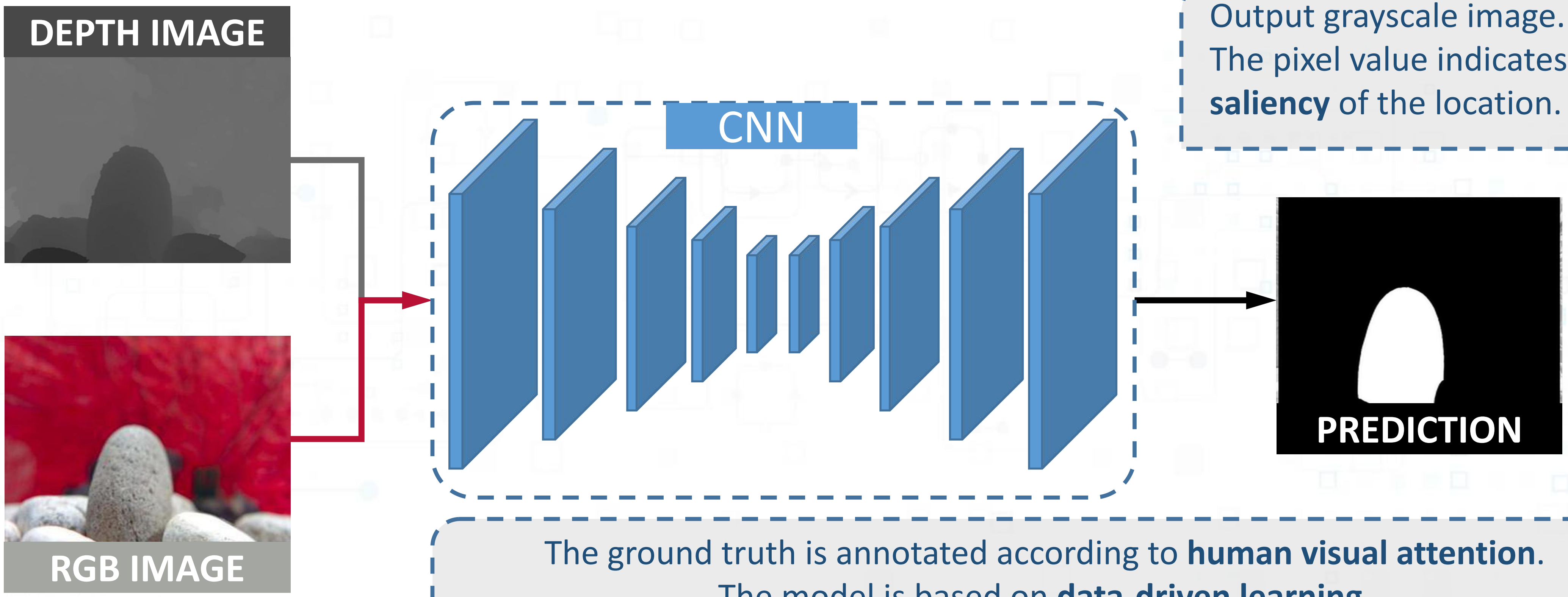
1. the original depth image with 1 channel
2. the depth image with 3 channel using HHA encoding[1]

- Depth features contain **spatial structure information**.
- RGB features contain rich **appearance and detailed information**.

RGB-D Salient Object Detection



RGB-D Salient Object Detection



Output grayscale image.
The pixel value indicates the **saliency** of the location.

The ground truth is annotated according to **human visual attention**.
The model is based on **data-driven learning**.

How to integrate depth information?

DEPTH FEATURE

How to make better use of the **characteristics** of depth information?

RGB FEATURE

Existing methods:

1. ignore **the guiding role of depth information** in the SOD task,
2. simply put it on an equal position with RGB information,
3. derectly use the fused features to make the final prediction

How to integrate depth information?

DEPTH FEATURE

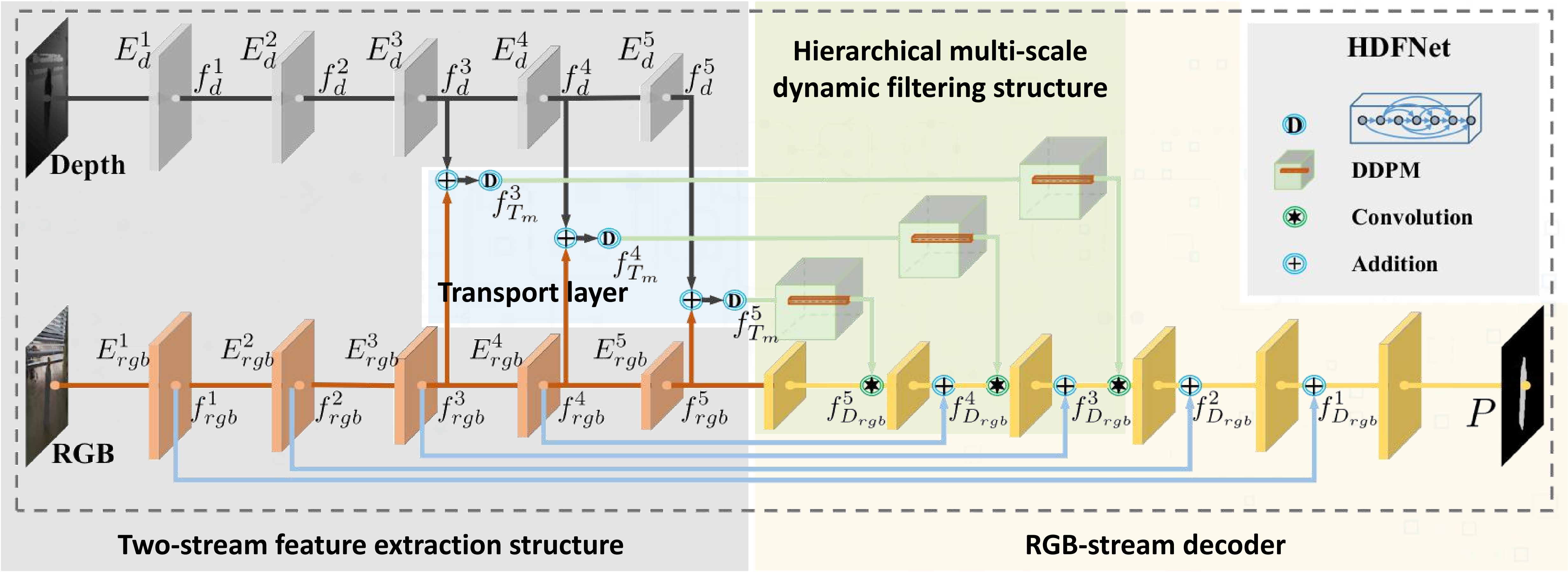
How to make better use of the **characteristics** of depth information?

RGB FEATURE

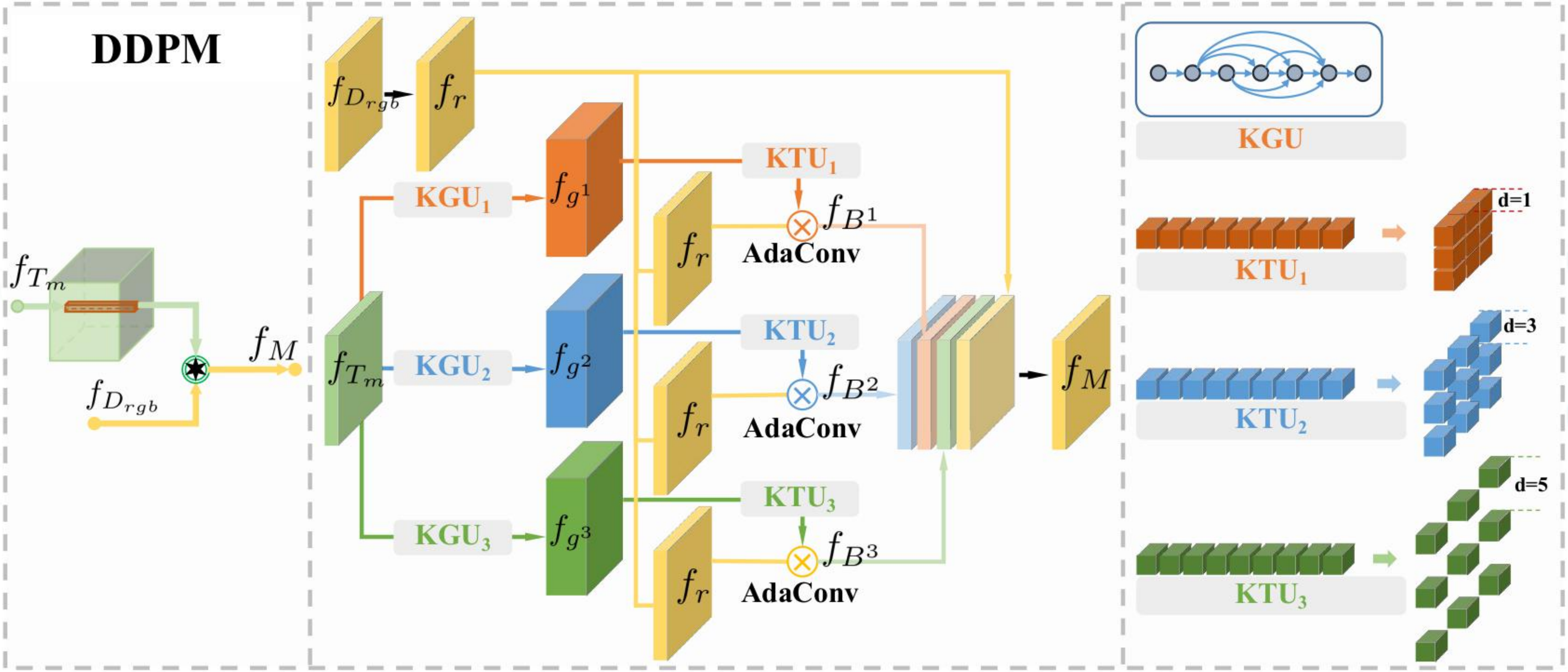
Our method:

1. gives full play to the guiding role of depth information,
2. uses fused RGB-D features to generate **sample-specific and position-specific multi-scale** filters to guide RGB features.

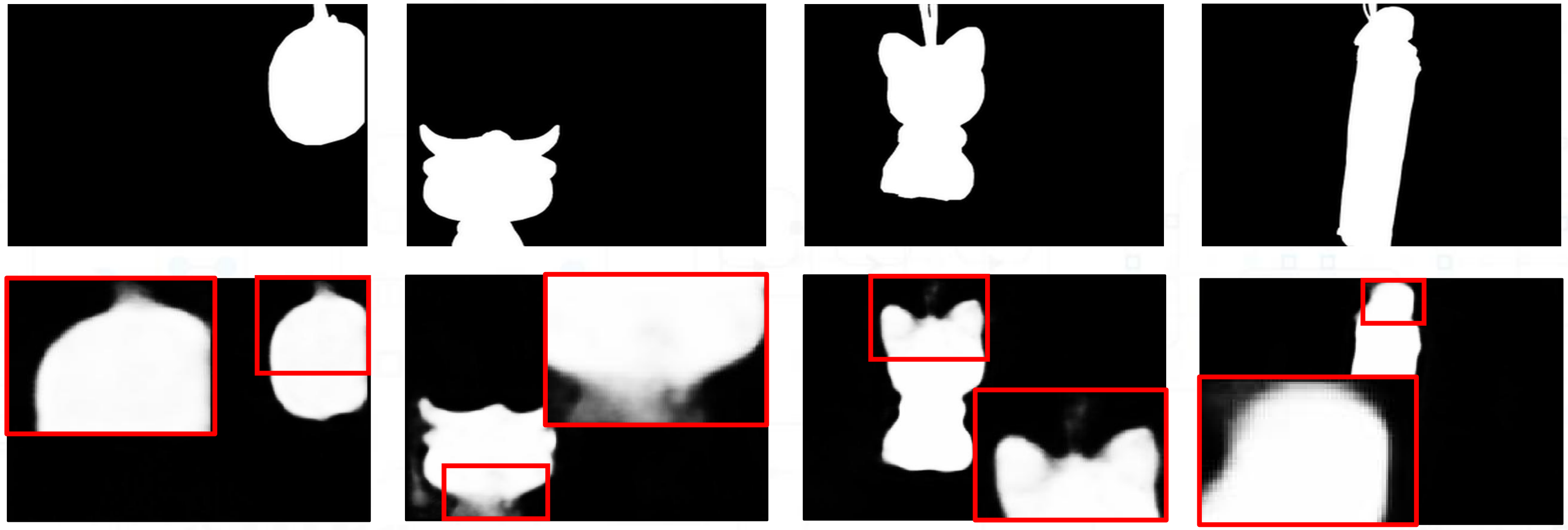
What is our method? **HDFNet!**



What is the core of our method? **DDPM!**



What are the shortcomings of our method?



We need **sharper boundaries** and **consistent saliency areas**!

How can we solve this problem? **HEL!**

$$L = L_{bce} + L_e + L_r,$$

$$L_{bce} = \frac{1}{N \times H \times W} \sum_n^N \sum_h^H \sum_w^W [g \log p + (1-g) \log(1-p)],$$

$$L_e = \frac{\sum_h^H \sum_w^W (e * |p - g|)}{\sum_h^H \sum_w^W e},$$

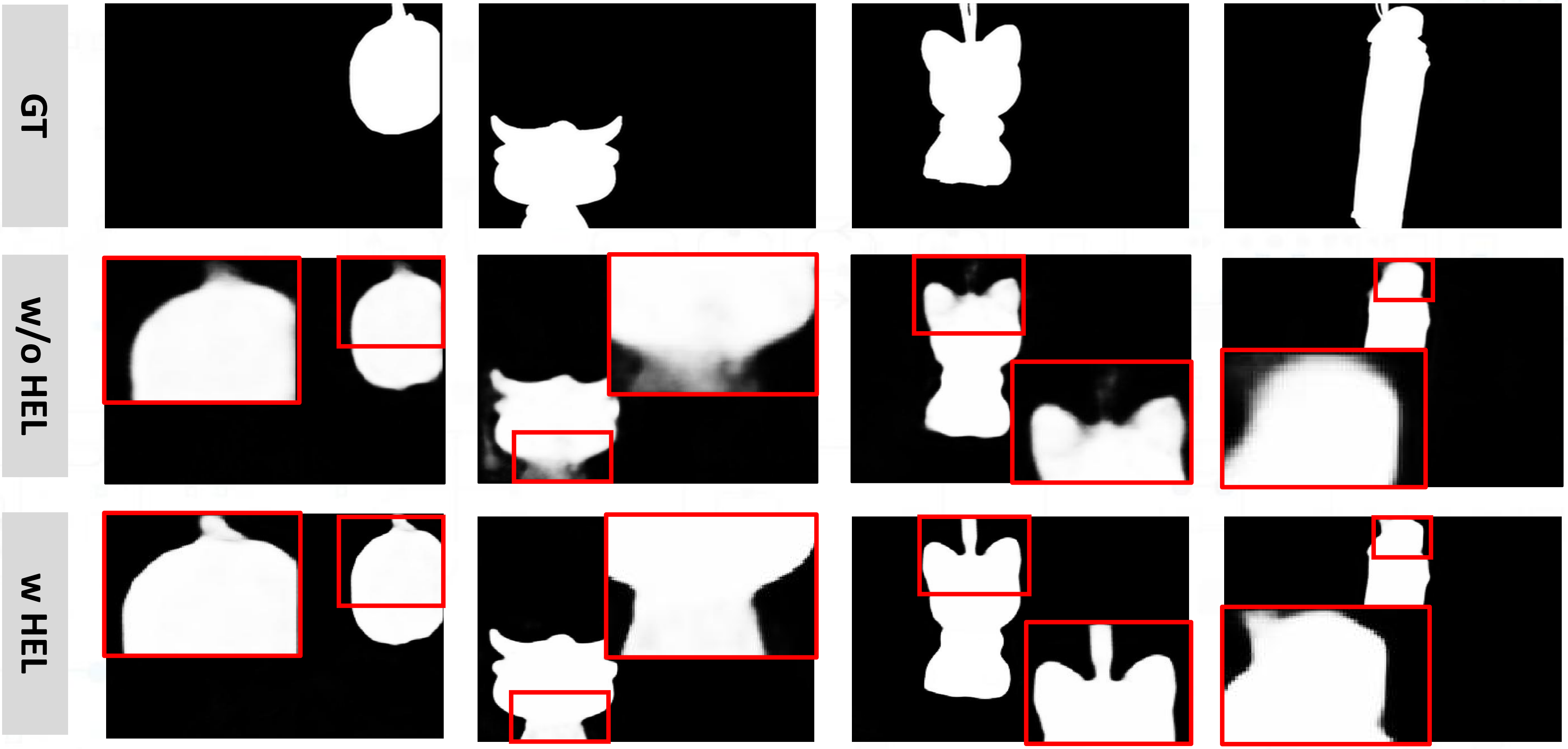
$$e = \begin{cases} 0 & \text{if } (G - \mathcal{P}(G))_{[h,w]} = 0, \\ 1 & \text{if } (G - \mathcal{P}(G))_{[h,w]} \neq 0, \end{cases}$$

$$L_r = \frac{\sum_n^N (L_f + L_b)}{N},$$

$$L_f = \frac{\sum_h^H \sum_w^W (g - g * p)}{\sum_h^H \sum_w^W g},$$

$$L_b = \frac{\sum_h^H \sum_w^W (1 - g) * p}{\sum_h^H \sum_w^W (1 - g)},$$

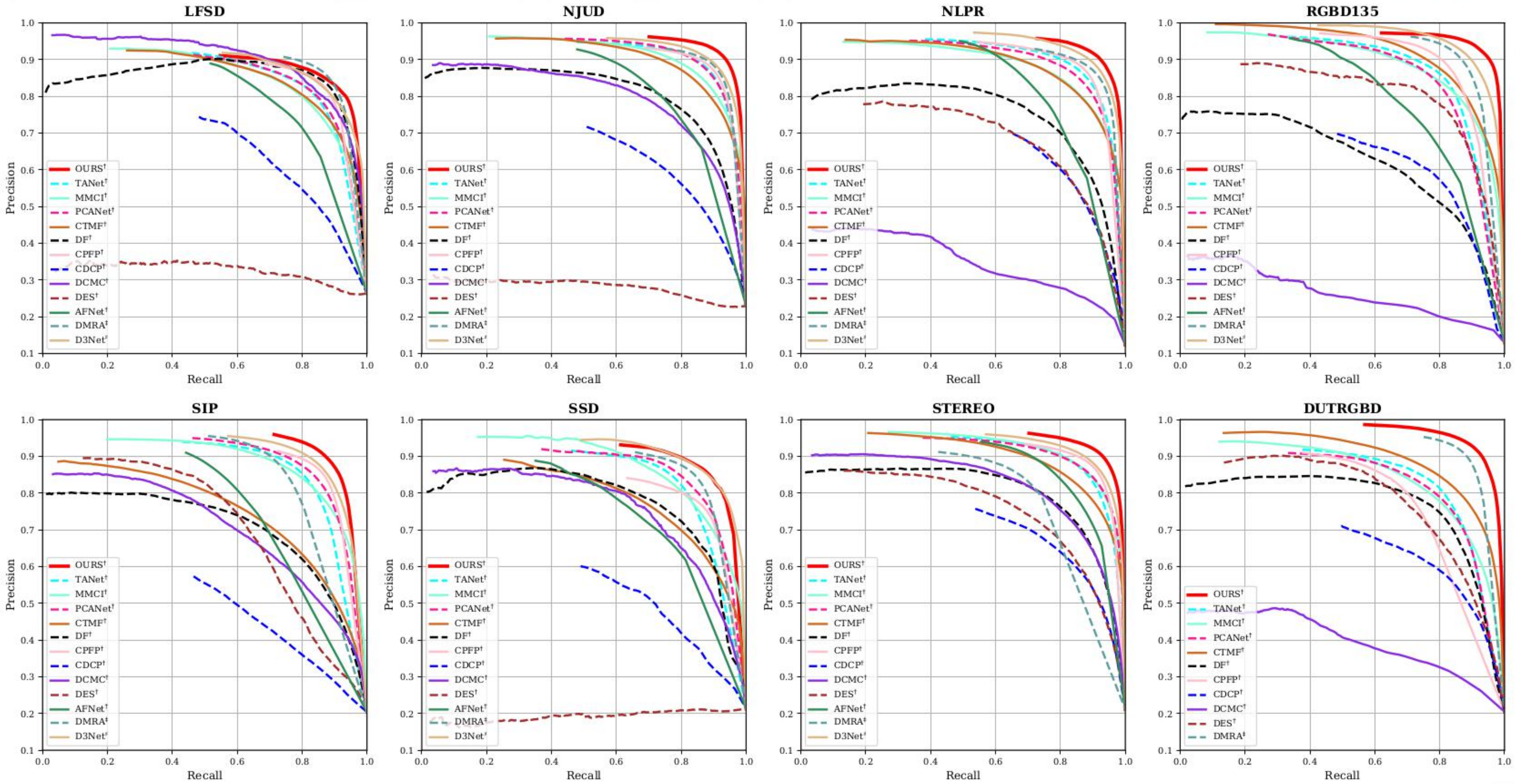
How's the effect?



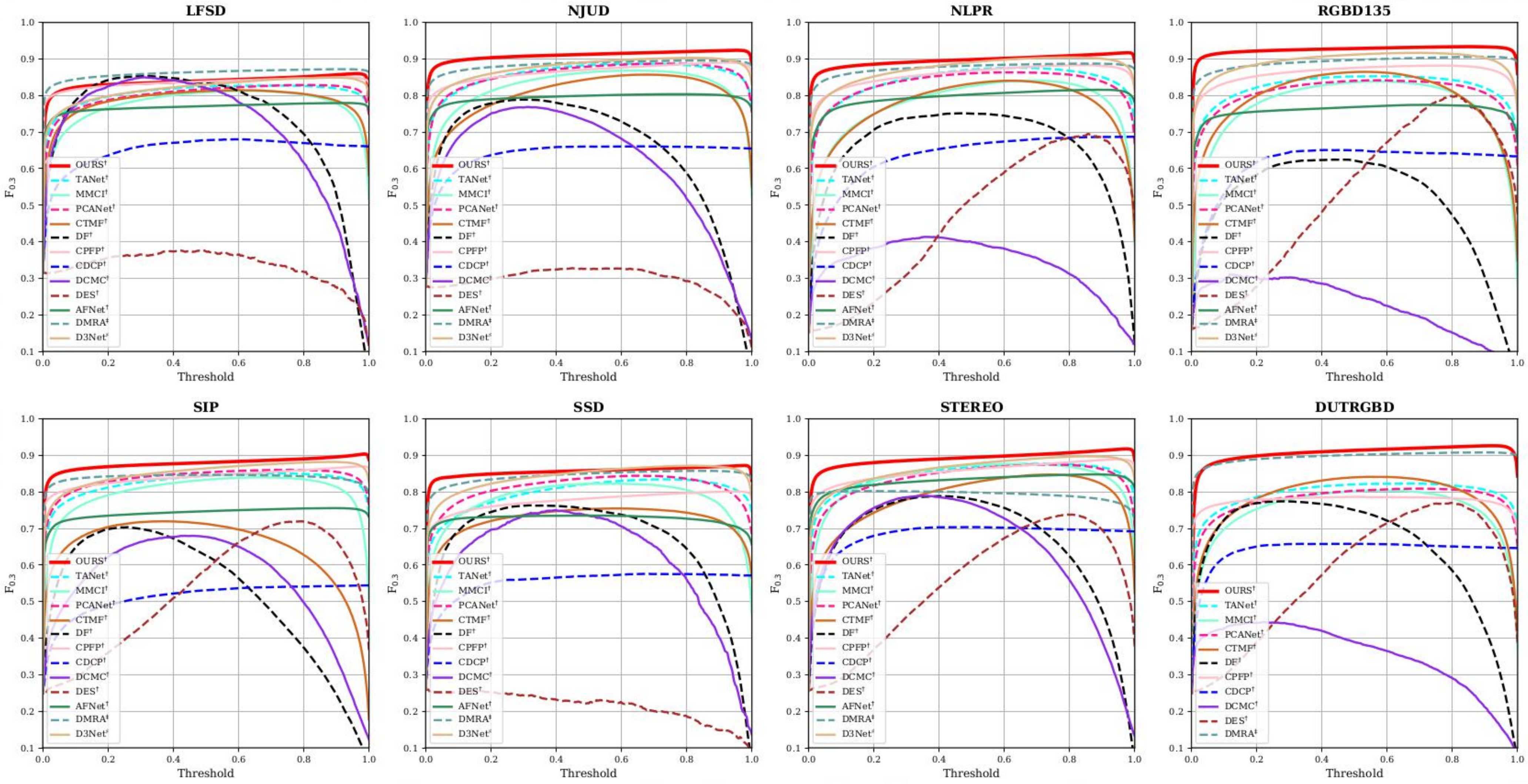
Experimental results

Dataset	Metric	DES$^\natural_{14}$ [6]	DCMC$^\natural_{16}$ [9]	CDCP$^\natural_{17}$ [54]	DF$^\dagger_{17}$ [38]	CTMF$^\dagger_{18}$ [15]	PCANet$^\dagger_{18}$ [2]	MMCI$^\dagger_{19}$ [4]	TANet$^\dagger_{19}$ [3]	AFNet$^\dagger_{19}$ [43]	CPFP$^\dagger_{19}$ [50]	OURS†	DMRA$^\ddagger_{19}$ [37]	OURS‡	D3Net$^\sharp_{19}$ [13]	OURS$^\sharp$
LFSD [25]	F_{max}	0.377	0.850	0.680	0.854	0.815	0.829	0.813	0.827	0.780	0.850	**0.860**	**0.872**	0.858	0.849	**0.883**
	F_{ada}	0.227	0.815	0.634	0.810	0.781	0.793	0.779	0.794	0.742	0.813	**0.831**	**0.849**	0.833	0.801	**0.843**
	F^ω_β	0.274	0.601	0.518	0.642	0.696	0.716	0.663	0.719	0.671	0.775	**0.792**	**0.811**	0.793	0.756	**0.806**
	MAE	0.416	0.155	0.199	0.142	0.120	0.112	0.132	0.111	0.133	0.088	**0.085**	**0.076**	0.083	0.099	**0.076**
	S_m	0.440	0.754	0.658	0.786	0.796	0.800	0.787	0.801	0.738	0.828	**0.847**	**0.847**	0.844	0.832	**0.854**
	E_m	0.492	0.842	0.737	0.841	0.851	0.856	0.840	0.851	0.810	0.867	**0.883**	**0.899**	0.886	0.860	**0.891**
NJUD [22]	F_{max}	0.328	0.769	0.661	0.789	0.857	0.887	0.868	0.888	0.804	0.890	**0.924**	0.896	**0.922**	0.903	**0.922**
	F_{ada}	0.165	0.715	0.618	0.744	0.788	0.844	0.813	0.844	0.768	0.837	**0.894**	0.872	**0.887**	0.840	**0.889**
	F^ω_β	0.234	0.497	0.510	0.545	0.720	0.803	0.739	0.805	0.696	0.828	**0.881**	0.847	**0.877**	0.833	**0.877**
	MAE	0.448	0.167	0.182	0.151	0.085	0.059	0.079	0.061	0.100	0.053	**0.037**	0.051	**0.038**	0.051	**0.038**
	S_m	0.413	0.703	0.672	0.735	0.849	0.877	0.859	0.878	0.772	0.878	**0.911**	0.885	**0.911**	0.895	**0.908**
	E_m	0.491	0.796	0.751	0.818	0.866	0.909	0.882	0.909	0.847	0.900	**0.934**	0.920	**0.932**	0.901	**0.932**
NLPR [35]	F_{max}	0.695	0.413	0.687	0.752	0.841	0.864	0.841	0.876	0.816	0.883	**0.917**	0.888	**0.919**	0.904	**0.927**
	F_{ada}	0.583	0.328	0.591	0.683	0.724	0.795	0.730	0.796	0.747	0.818	**0.878**	0.855	**0.883**	0.834	**0.889**
	F^ω_β	0.254	0.259	0.501	0.516	0.679	0.762	0.676	0.780	0.693	0.807	**0.869**	0.839	**0.871**	0.826	**0.882**
	MAE	0.300	0.196	0.114	0.100	0.056	0.044	0.059	0.041	0.058	0.038	**0.027**	0.031	**0.027**	0.034	**0.023**
	S_m	0.582	0.550	0.724	0.769	0.860	0.873	0.856	0.886	0.799	0.884	**0.916**	0.898	**0.915**	0.906	**0.923**
	E_m	0.760	0.685	0.786	0.840	0.869	0.916	0.872	0.916	0.884	0.920	**0.948**	0.942	**0.951**	0.934	**0.957**
RGBD135 [7]	F_{max}	0.800	0.311	0.651	0.625	0.865	0.842	0.839	0.853	0.775	0.882	**0.934**	0.906	**0.941**	0.917	**0.932**
	F_{ada}	0.695	0.234	0.594	0.573	0.778	0.774	0.762	0.795	0.730	0.829	**0.919**	0.867	**0.918**	0.876	**0.912**
	F^ω_β	0.301	0.169	0.478	0.392	0.686	0.711	0.650	0.740	0.641	0.787	**0.902**	0.843	**0.913**	0.831	**0.895**
	MAE	0.288	0.196	0.120	0.131	0.055	0.050	0.065	0.046	0.068	0.038	**0.020**	0.030	**0.017**	0.030	**0.021**
	S_m	0.632	0.469	0.709	0.685	0.863	0.843	0.848	0.858	0.770	0.872	**0.932**	0.899	**0.937**	0.904	**0.926**
	E_m	0.817	0.676	0.810	0.806	0.911	0.912	0.904	0.919	0.874	0.927	**0.973**	0.944	**0.976**	0.956	**0.971**
SIP [13]	F_{max}	0.720	0.680	0.544	0.704	0.720	0.860	0.840	0.851	0.756	0.870	**0.904**	0.847	**0.907**	0.882	**0.910**
	F_{ada}	0.644	0.645	0.495	0.673	0.684	0.825	0.795	0.809	0.705	0.819	**0.863**	0.815	**0.870**	0.831	**0.875**
	F^ω_β	0.342	0.413	0.397	0.406	0.535	0.768	0.711	0.748	0.617	0.788	**0.835**	0.734	**0.844**	0.793	**0.848**
	MAE	0.298	0.186	0.224	0.185	0.139	0.071	0.086	0.075	0.118	0.064	**0.050**	0.088	**0.047**	0.063	**0.047**
	S_m	0.616	0.683	0.595	0.653	0.716	0.842	0.833	0.835	0.720	0.850	**0.878**	0.800	**0.885**	0.864	**0.886**
	E_m	0.751	0.786	0.722	0.794	0.824	0.900	0.886	0.894	0.815	0.899	**0.920**	0.858	**0.924**	0.903	**0.924**
SSD [53]	F_{max}	0.260	0.750	0.576	0.763	0.755	0.844	0.823	0.834	0.735	0.801	**0.872**	0.858	**0.883**	0.872	**0.885**
	F_{ada}	0.073	0.684	0.524	0.709	0.709	0.786	0.748	0.766	0.694	0.726	**0.844**	0.821	**0.847**	0.793	**0.842**
	F^ω_β	0.172	0.480	0.429	0.536	0.622	0.733	0.662	0.727	0.589	0.708	**0.808**	0.787	**0.819**	0.780	**0.821**
	MAE	0.500	0.168	0.219	0.151	0.100	0.063	0.082	0.063	0.118	0.082	**0.048**	0.058	**0.046**	0.058	**0.045**
	S_m	0.341	0.706	0.603	0.741	0.776	0.842	0.813	0.839	0.714	0.807	**0.866**	0.856	**0.875**	0.866	**0.879**
	E_m	0.475	0.790	0.714	0.801	0.838	0.890	0.860	0.886	0.803	0.832	**0.913**	0.898	**0.911**	0.892	**0.911**
STEREO [33]	F_{max}	0.738	0.789	0.704	0.789	0.848	0.875	0.877	0.878	0.848	0.889	**0.918**	0.802	**0.916**	0.897	**0.910**
	F_{ada}	0.594	0.742	0.666	0.742	0.771	0.826	0.829	0.835	0.807	0.830	**0.879**	0.762	**0.875**	0.833	**0.867**
	F^ω_β	0.375	0.520	0.558	0.549	0.698	0.778	0.760	0.787	0.752	0.817	**0.863**	0.647	**0.859**	0.815	**0.853**
	MAE	0.295	0.148	0.149	0.141	0.086	0.064	0.068	0.060	0.075	0.051	**0.039**	0.087	**0.040**	0.054	**0.041**
	S_m	0.642	0.731	0.713	0.757	0.848	0.875	0.873	0.871	0.825	0.879	**0.906**	0.752	**0.903**	0.891	**0.900**
	E_m	0.696	0.831	0.796	0.838	0.870	0.907	0.905	0.916	0.887	0.907	**0.937**	0.816	**0.934**	0.911	**0.931**
DUTRGBD [37]	F_{max}	0.770	0.444	0.658	0.774	0.842	0.809	0.804	0.823	-	0.787	**0.926**	0.908	**0.934**	-	**0.930**
	F_{ada}	0.667	0.405	0.633	0.747	0.792	0.760	0.753	0.778	-	0.735	**0.892**	0.883	**0.894**	-	**0.885**
	F^ω_β	0.380	0.284	0.521	0.536	0.682	0.688	0.628	0.705	-	0.638	**0.865**	0.852	**0.871**	-	**0.864**
	MAE	0.280	0.243	0.159	0.145	0.097	0.100	0.112	0.093	-	0.100	**0.040**	0.048	**0.039**	-	**0.041**
	S_m	0.659	0.499	0.687	0.729	0.831	0.801	0.791	0.808	-	0.749	**0.905**	0.887	**0.911**	-	**0.907**
	E_m	0.751	0.712	0.794	0.842	0.882	0.863	0.856	0.871	-	0.815	**0.938**	0.930	**0.941**	-	**0.938**

Experimental results



Experimental results



Experimental results

Model	No.	Baseline	$+\mathbf{T}_d$	$+\mathbf{T}_{rgb}$	+DDPM	+DCM	$+\mathbf{L}_e$	$+\mathbf{L}_f$	$+\mathbf{L}_b$	F_{max}	F_{ada}	F^{ω}_{β}	MAE	S_m	E_m
Ours†	1	✔								0.875	0.819	0.768	0.067	0.865	0.898
	2	✔	✔							0.879	0.820	0.768	0.066	0.868	0.899
	3	✔	✔		✔					0.882	0.820	0.780	0.063	0.873	0.900
	4	✔		✔						0.884	0.839	0.787	0.060	0.874	0.909
	5	✔		✔	✔					0.896	0.852	0.811	0.054	0.886	0.916
	6	✔	✔	✔						0.898	0.846	0.803	0.056	0.884	0.913
	7	✔	✔	✔	✔					0.904	0.856	0.820	0.052	0.893	0.918
	8	✔	✔	✔		✔				0.878	0.823	0.777	0.064	0.871	0.903
	9	✔	✔	✔	✔		✔			0.909	0.878	0.849	0.044	0.898	0.929
	10	✔	✔	✔	✔			✔		0.909	0.845	0.827	0.050	0.887	0.916
	11	✔	✔	✔	✔				✔	0.907	0.874	0.836	0.048	0.895	0.926
	12	✔	✔	✔	✔		✔	✔	✔	0.914	0.878	0.857	0.041	0.898	0.933
R3Net18 [10]	13									0.828	0.714	0.716	0.072	0.831	0.830
	14						✔	✔	✔	0.832	0.731	0.740	0.069	0.835	0.844
CPD19 [46]	15									0.848	0.790	0.769	0.052	0.856	0.889
	16						✔	✔	✔	0.849	0.804	0.792	0.049	0.857	0.898
PoolNet19 [27]	15									0.832	0.755	0.728	0.060	0.841	0.865
	16						✔	✔	✔	0.861	0.811	0.799	0.046	0.862	0.902
GCPANet20 [5]	17									0.847	0.766	0.744	0.061	0.854	0.869
	18						✔	✔	✔	0.854	0.779	0.773	0.055	0.856	0.880

(1)

ECCV'20 ONLINE
23-28 AUGUST 2020

IIAU-LAB

Experimental results

Model	No.	Baseline	$+\mathbf{T}_d$	$+\mathbf{T}_{rgb}$	+DDPM	+DCM	$+\mathbf{L}_e$	$+\mathbf{L}_f$	$+\mathbf{L}_b$	F_{max}	F_{ada}	F_β^ω	MAE	S_m	E_m
Ours†	1	✔								0.875	0.819	0.768	0.067	0.865	0.898
	2	✔	✔							0.879	0.820	0.768	0.066	0.868	0.899
	3	✔	✔		✔					0.882	0.820	0.780	0.063	0.873	0.900
	4	✔		✔						0.884	0.839	0.787	0.060	0.874	0.909
	5	✔		✔	✔					0.896	0.852	0.811	0.054	0.886	0.916
	6	✔	✔	✔						0.898	0.846	0.803	0.056	0.884	0.913
	7	✔	✔	✔	✔					0.904	0.856	0.820	0.052	0.893	0.918
	8	✔	✔	✔		✔				0.878	0.823	0.777	0.064	0.871	0.903
	9	✔	✔	✔	✔		✔			0.909	0.878	0.849	0.044	0.898	0.929
	10	✔	✔	✔	✔			✔		0.909	0.845	0.827	0.050	0.887	0.916
	11	✔	✔	✔	✔				✔	0.907	0.874	0.836	0.048	0.895	0.926
	12	✔	✔	✔	✔		✔	✔	✔	0.914	0.878	0.857	0.041	0.898	0.933
R3Net_{18} [10]	13									0.828	0.714	0.716	0.072	0.831	0.830
	14						✔	✔	✔	0.832	0.731	0.740	0.069	0.835	0.844
CPD_{19} [46]	15									0.848	0.790	0.769	0.052	0.856	0.889
	16						✔	✔	✔	0.849	0.804	0.792	0.049	0.857	0.898
PoolNet_{19} [27]	15									0.832	0.755	0.728	0.060	0.841	0.865
	16						✔	✔	✔	0.861	0.811	0.799	0.046	0.862	0.902
GCPANet_{20} [5]	17									0.847	0.766	0.744	0.061	0.854	0.869
	18						✔	✔	✔	0.854	0.779	0.773	0.055	0.856	0.880

（2）

（3）

REFERENCE

[1] Learning Rich Features from RGB-D Images for Object Detection and Segmentation
[2] CTMF: CNNs-Based RGB-D Saliency Detection via Cross-View Transfer and Multiview Fusion
[3] PCANet: Progressively Complementarity-aware Fusion Network for RGB-D Salient Object Detection
[4] MMCI: Multi-modal fusion network with multi-scale multi-path and cross-modal interactions for RGB-D salient object detection
[5] TANet: Three-stream Attention-aware Network for RGB-D Salient Object Detection
[6] CPFP: Contrast Prior and Fluid Pyramid Integration for RGBD Salient Object Detection
[7] DMRA: Depth-induced Multi-scale Recurrent Attention Network for Saliency Detection
[8] D3Net: Rethinking RGB-D Salient Object Detection: Models, Datasets, and Large-Scale Benchmarks

ECCV'20 ONLINE

23-28 AUGUST 2020

16TH EUROPEAN CONFERENCE ON

COMPUTER VISION

WWW.ECCV2020.EU

Thanks for your attention

DALIAN UNIVERSITY OF TECHNOLOGY 1949 大连理工大学

IIAU-LAB