

# U2Net论文笔记

这篇论文是一篇效果非常好的显著性目标检测的论文，效果非常好，这里先放上实验效果图，体验一下它的效果。

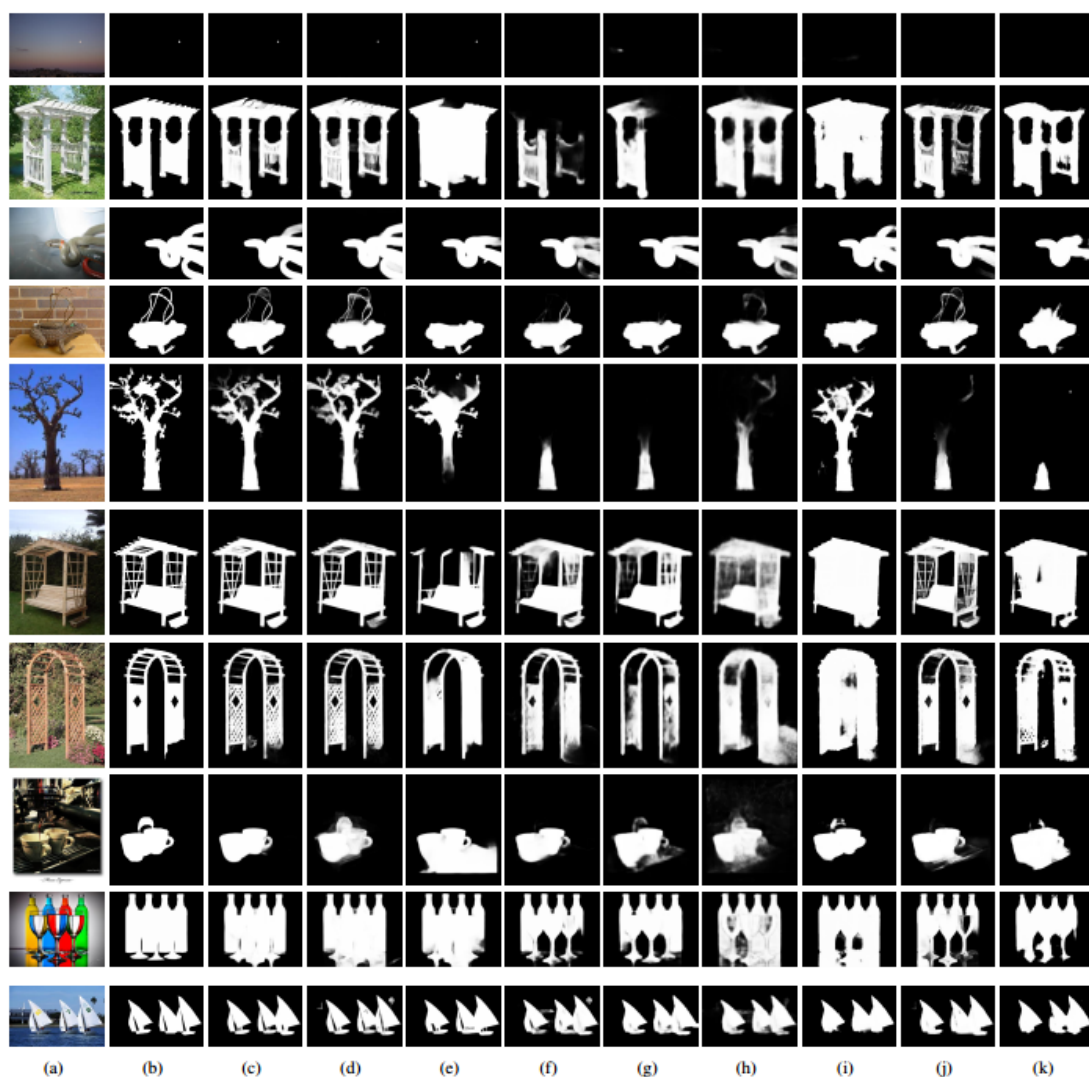


Figure 7. Qualitative comparison of the proposed method with seven other SOTA methods: (a) image, (b) GT, (c) Ours, (d) Ours<sup>†</sup>, (e) BASNet, (f) PoolNet, (g) CPD, (h) PiCANetR, (i) R<sup>3</sup>Net+, (j) AFNet, (k) DSS+, where '+' indicates the CRF post-processing.

## 1. 论文摘要

在这篇论文中，我们设计了一个简单但是非常实用的深度学习显著性目标检测网络U2Net。U2Net是一个两级的嵌套U型结构，这种设计和方式有以下几个优点：

1. 利用来自残差U型模块（RSU）的不同尺度不同感受野的混合，能够捕捉来自更多的不同尺度的上下文信息（全局信息）
2. 采用RSU中的池化操作，U2Net在不增加计算复杂度的基础上，可以提升整个模型架构的深度。
3. 使用这种结构可以从头训练网络而不依赖于图像分类的算法网络

针对U2Net论文中提出了两种架构，第一个是U2Net（网络大小为176.3MB，在GTX 1080Ti上推理速度为30FPS），第二个模型为U2Net<sup>+</sup>，网络大小为4.7MB,推理速度为40FPS。利用这两种模型来解决不同环境的使用，在6个显著性目标检测数据集上两种模型均实现了竞争性的结果。代码的开源地址为：<https://github.com/NathanUA/U-2-Net>

U2Net算法模型的总体架构：

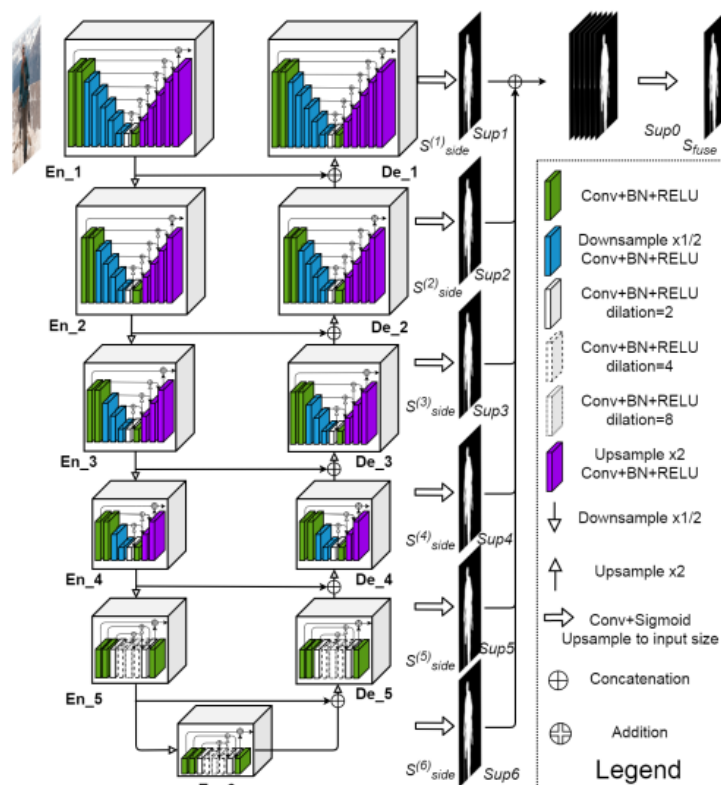


Figure 5. Illustration of our proposed U<sup>2</sup>-Net architecture. The main architecture is a U-Net like Encoder-Decoder, where each stage consists of our newly proposed residual U-block (RSU). For example, En\_1 is based on our RSU block shown in Fig. 2(e). Detailed configuration of RSU block of each stage is given in the last two rows of Table 1.

## 2. 论文主要思想

文章的出发点主要针对以下几个问题：

1. 传统的深度学习的基础网络采用VGG，ResNet，Inception，DensNet，这种结构经过一系列的下采样提取了原始图像的全局语义信息，然而对于局部的细节以及全局对比信息的关注度不够（由于这些网络结构主要针对图像分类任务设计），而且使用这种模型一般需要采用ImageNet上的预训练模型，如果目标数据的分布与ImageNet的数据集分布差距较大的话，效果相对较差。**怎样设计一种能够从头训练的新显著性目标检测的网络架构，并且算法效果与已经存在的效果相比更好，就是一个需要考虑的问题。**
2. 传统的网络架构采用一系列的下采样（最大池化或者stride大于1的卷积）来减小feature map的分辨率，从而降低算大的计算复杂度。然而针对分割任务而言，高分辨率的feature map也是一个重要的因素，**因此设计一个在较低的计算复杂度的情况下，随着网络的加深能够维持较高的分辨率就是一个重要的问题。**

一个非常简单的架构U2Net很好的解决了上边的问题，**U2Net是一个两级的嵌入U型网络结构，这是一种新型的网络结构不需要进行预训练，从头训练可以得到很好的结果。而且，这种网络结构实现了在网络层数加深的同时能够保持较高的分辨率。**

### 3. 算法介绍

为了实现上边的功能，算法引入了以下的几个主要解决方案（创新点）。

两级的U型嵌套结构：

1. 底层是利用RSU在网络层数加深的情况下依然维持较高的分辨率
2. 高层，利用类似UNet的网络结构

#### 3.1 残差U型单元（RSU）

几种基本的卷积神经网络的模块结构图对比

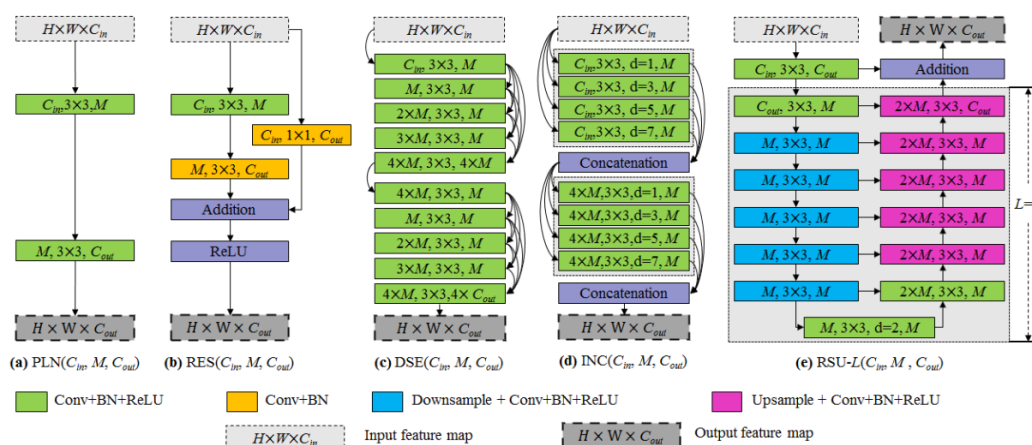


Figure 2. Illustration of existing convolution blocks and our proposed residual U-block RSU: (a) Plain convolution block PLN, (b) Residual-like block RES, (c) Inception-like block INC, (d) Dense-like block DSE and (e) Our residual U-block RSU.

从左到有依次是，基本的卷积模块，残差模块，Inception模块，Dense模块，残差U型模块。

看（e）的RSU模块，利用U型单元模块，实现了低层高分辨率的局部信息与高层低分辨率全局信息的融合。

下边是残差模块与残差u型模块的对比：

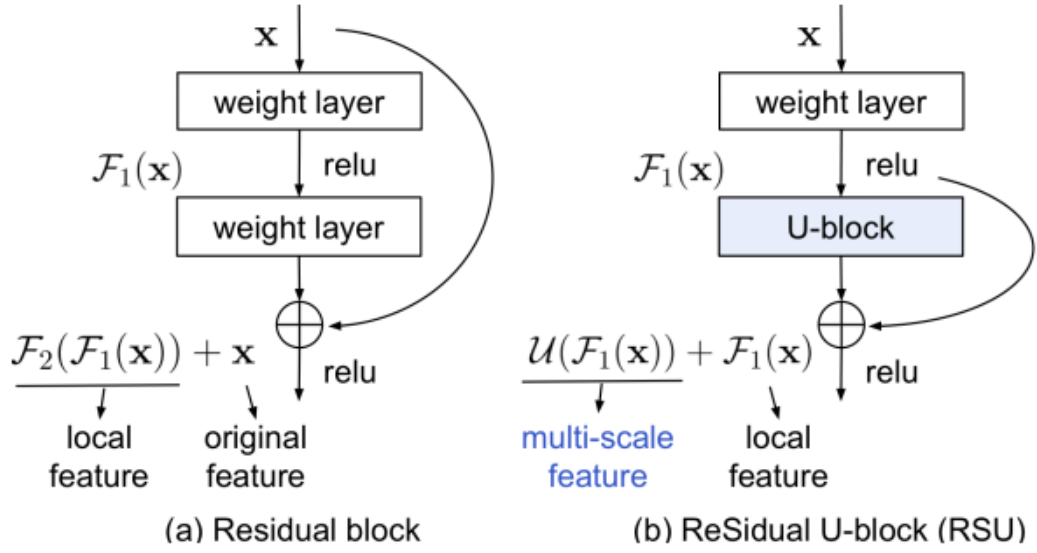


Figure 3. Comparison of the residual block and our RSU.

### 3.2 U2Net的架构

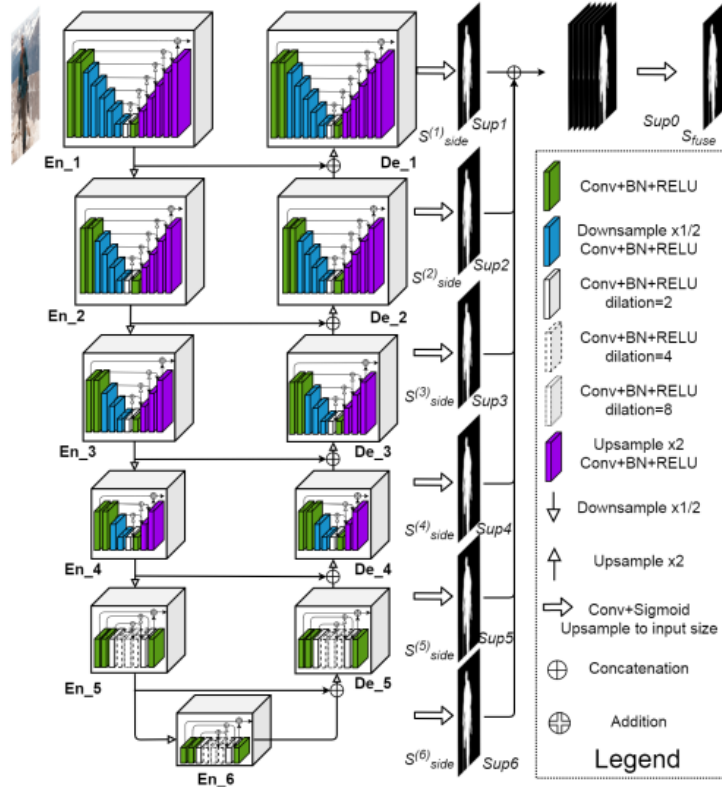


Figure 5. Illustration of our proposed  $U^2$ -Net architecture. The main architecture is a U-Net like Encoder-Decoder, where each stage consists of our newly proposed residual U-block (RSU). For example, **En\_1** is based on our RSU block shown in Fig. 2(e). Detailed configuration of RSU block of each stage is given in the last two rows of Table 1.

算法的整体是一个U型的网络结构，其中每个模块都采用的残差U型模块堆叠而成。

## 算法实验结果

Table 3. Comparison of our method and 20 SOTA methods on DUT-OMRON, DUTS-TE, HKU-IS in terms of model size,  $maxF_{\beta}$  ( $\uparrow$ ),  $MAE$  ( $\downarrow$ ), weighted  $F_{\beta}^w$  ( $\uparrow$ ), structure measure  $S_m$  ( $\uparrow$ ) and relax boundary F-measure  $relaxF_{\beta}^b$  ( $\uparrow$ ). Red, Green, and Blue indicate the best, second best and third best performance.

Method	Backbone	Size(MB)	DUT-OMRON (5168)					DUTS-TE (5019)					HKU-IS (4447)				
			$maxF_{\beta}$	$MAE$	$F_{\beta}^w$	$S_m$	$relaxF_{\beta}^b$	$maxF_{\beta}$	$MAE$	$F_{\beta}^w$	$S_m$	$relaxF_{\beta}^b$	$maxF_{\beta}$	$MAE$	$F_{\beta}^w$	$S_m$	$relaxF_{\beta}^b$
MDF <sub>TIP16</sub>	AlexNet	112.1	0.694	0.142	0.565	0.721	0.406	0.729	0.099	0.543	0.723	0.447	0.860	0.129	0.564	0.810	0.594
UCF <sub>ICCV17</sub>	VGG-16	117.9	0.730	0.120	0.573	0.760	0.480	0.773	0.112	0.596	0.777	0.518	0.888	0.062	0.779	0.875	0.679
Amulet <sub>ICCV17</sub>	VGG-16	132.6	0.743	0.098	0.626	0.781	0.528	0.778	0.084	0.658	0.796	0.568	0.897	0.051	0.817	0.886	0.716
NLDF <sub>CVPR17</sub>	VGG-16	428.0	0.753	0.080	0.634	0.770	0.514	0.813	0.065	0.710	0.805	0.591	0.902	0.048	0.838	0.879	0.694
DSS <sub>CVPR17</sub>	VGG-16	237.0	0.781	0.063	0.697	0.790	0.559	0.825	0.056	0.755	0.812	0.606	0.916	0.040	0.867	0.878	0.706
RAS <sub>ICCV18</sub>	VGG-16	81.0	0.786	0.062	0.695	0.814	0.615	0.831	0.059	0.740	0.828	0.656	0.913	0.045	0.843	0.887	0.748
PAGRN <sub>CVPR18</sub>	VGG-19	-	0.771	0.071	0.622	0.775	0.582	0.854	0.055	0.724	0.825	0.692	0.918	0.048	0.820	0.887	0.762
BMPM <sub>CVPR18</sub>	VGG-16	-	0.774	0.064	0.681	0.809	0.612	0.852	0.048	0.761	0.851	0.699	0.921	0.039	0.859	0.907	0.773
PiCANet <sub>CVPR18</sub>	VGG-16	153.3	0.794	0.068	0.691	0.826	0.643	0.851	0.054	0.747	0.851	0.704	0.921	0.042	0.847	0.906	0.784
MLMS <sub>CVPR19</sub>	VGG-16	263.0	0.774	0.064	0.681	0.809	0.612	0.852	0.048	0.761	0.851	0.699	0.921	0.039	0.859	0.907	0.773
AFNet <sub>CVPR19</sub>	VGG-16	143.0	0.797	0.057	0.717	0.826	0.635	0.862	0.046	0.785	0.855	0.714	0.923	0.036	0.869	0.905	0.772
MSWS <sub>CVPR19</sub>	Dense-169	48.6	0.718	0.109	0.527	0.756	0.362	0.767	0.908	0.586	0.749	0.376	0.856	0.084	0.685	0.818	0.438
R <sup>3</sup> Net <sub>ICAA18</sub>	ResNeXt	215.0	0.795	0.063	0.728	0.817	0.599	0.828	0.058	0.763	0.817	0.601	0.915	0.036	0.877	0.895	0.740
CapSal <sub>CVPR19</sub>	ResNet-101	-	0.699	0.101	0.482	0.674	0.396	0.823	0.072	0.691	0.808	0.605	0.882	0.062	0.782	0.850	0.654
SRM <sub>ICCV17</sub>	ResNet-50	189.0	0.769	0.069	0.658	0.798	0.523	0.826	0.058	0.722	0.824	0.592	0.906	0.046	0.835	0.887	0.680
DGRL <sub>CVPR18</sub>	ResNet-50	646.1	0.779	0.063	0.697	0.810	0.584	0.834	0.051	0.760	0.836	0.656	0.913	0.037	0.865	0.897	0.744
PiCANet <sub>CVPR18</sub>	ResNet-50	197.2	0.803	0.065	0.695	0.832	0.632	0.860	0.050	0.755	0.859	0.696	0.918	0.043	0.840	0.904	0.765
CPD <sub>CVPR19</sub>	ResNet-50	183.0	0.797	0.056	0.719	0.825	0.655	0.865	0.043	0.795	0.858	0.741	0.925	0.034	0.875	0.905	0.795
PoolNet <sub>CVPR19</sub>	ResNet-50	273.3	0.808	0.056	0.729	0.836	0.675	0.880	0.040	0.807	0.871	0.765	0.932	0.033	0.881	0.917	0.811
BASNet <sub>CVPR19</sub>	ResNet-34	348.5	0.805	0.056	0.751	0.836	0.694	0.860	0.047	0.803	0.853	0.758	0.928	0.032	0.889	0.909	0.807
U <sup>2</sup> -Net (Ours)	RSU	176.3	0.823	0.054	0.757	0.847	0.702	0.873	0.044	0.804	0.861	0.765	0.935	0.031	0.890	0.916	0.812
U <sup>2</sup> -Net <sup>†</sup> (Ours)	RSU	4.7	0.813	0.060	0.731	0.837	0.676	0.852	0.054	0.763	0.847	0.723	0.928	0.037	0.867	0.908	0.794

Table 4. Comparison of our method and 20 SOTA methods on ECSSD, PASCAL-S, SOD in terms of model size,  $maxF_{\beta}$  ( $\uparrow$ ),  $MAE$  ( $\downarrow$ ), weighted  $F_{\beta}^w$  ( $\uparrow$ ), structure measure  $S_m$  ( $\uparrow$ ) and relax boundary F-measure  $relaxF_{\beta}^b$  ( $\uparrow$ ). Red, Green, and Blue indicate the best, second best and third best performance.

Method	Backbone	Size(MB)	ECSSD (1000)					PASCAL-S (850)					SOD (300)				
			$maxF_{\beta}$	$MAE$	$F_{\beta}^w$	$S_m$	$relaxF_{\beta}^b$	$maxF_{\beta}$	$MAE$	$F_{\beta}^w$	$S_m$	$relaxF_{\beta}^b$	$maxF_{\beta}$	$MAE$	$F_{\beta}^w$	$S_m$	$relaxF_{\beta}^b$
MDF <sub>TIP16</sub>	AlexNet	112.1	0.832	0.105	0.705	0.776	0.472	0.759	0.142	0.589	0.696	0.343	0.746	0.192	0.508	0.643	0.311
UCF <sub>ICCV17</sub>	VGG-16	117.9	0.903	0.069	0.806	0.884	0.669	0.814	0.115	0.694	0.805	0.493	0.808	0.148	0.675	0.762	0.471
Amulet <sub>ICCV17</sub>	VGG-16	132.6	0.915	0.059	0.840	0.894	0.711	0.828	0.100	0.734	0.818	0.541	0.798	0.144	0.677	0.753	0.454
NLDF <sub>CVPR17</sub>	VGG-16	428.0	0.905	0.063	0.839	0.897	0.666	0.822	0.098	0.737	0.798	0.495	0.841	0.125	0.709	0.755	0.475
DSS <sub>CVPR17</sub>	VGG-16	237.0	0.921	0.052	0.872	0.882	0.696	0.831	0.093	0.759	0.798	0.499	0.846	0.124	0.710	0.743	0.444
RAS <sub>ICCV18</sub>	VGG-16	81.0	0.921	0.056	0.857	0.893	0.741	0.829	0.101	0.736	0.799	0.560	0.851	0.124	0.720	0.764	0.544
PAGRN <sub>CVPR18</sub>	VGG-19	-	0.927	0.061	0.834	0.889	0.747	0.847	0.090	0.738	0.822	0.594	-	-	-	-	-
BMPM <sub>CVPR18</sub>	VGG-16	-	0.928	0.045	0.871	0.911	0.770	0.850	0.074	0.779	0.845	0.617	0.856	0.108	0.726	0.786	0.562
PiCANet <sub>CVPR18</sub>	VGG-16	153.3	0.931	0.046	0.865	0.914	0.784	0.856	0.078	0.772	0.848	0.612	0.854	0.103	0.722	0.789	0.572
MLMS <sub>CVPR19</sub>	VGG-16	263.0	0.928	0.045	0.871	0.911	0.770	0.855	0.074	0.779	0.844	0.620	0.856	0.108	0.726	0.786	0.562
AFNet <sub>CVPR19</sub>	VGG-16	143.0	0.935	0.042	0.887	0.914	0.776	0.863	0.070	0.798	0.849	0.626	0.856	0.111	0.723	0.774	-
MSWS <sub>CVPR19</sub>	Dense-169	48.6	0.878	0.096	0.716	0.828	0.411	0.786	0.133	0.614	0.768	0.289	0.800	0.167	0.573	0.700	0.231
R <sup>3</sup> Net <sub>ICAA18</sub>	ResNeXt	215.0	0.934	0.040	0.902	0.910	0.759	0.834	0.092	0.761	0.807	0.538	0.850	0.125	0.735	0.759	0.431
CapSal <sub>CVPR19</sub>	ResNet-101	-	0.874	0.077	0.771	0.826	0.574	0.861	0.073	0.786	0.837	0.527	0.773	0.148	0.597	0.695	0.404
SRM <sub>ICCV17</sub>	ResNet-50	189.0	0.917	0.054	0.853	0.895	0.672	0.838	0.084	0.758	0.834	0.509	0.843	0.128	0.670	0.741	0.392
DGRL <sub>CVPR18</sub>	ResNet-50	646.1	0.925	0.042	0.883	0.906	0.753	0.848	0.074	0.787	0.839	0.569	0.848	0.106	0.731	0.773	0.502
PiCANet <sub>CVPR18</sub>	ResNet-50	197.2	0.935	0.046	0.867	0.917	0.775	0.857	0.076	0.777	0.854	0.598	0.856	0.104	0.724	0.790	0.528
CPD <sub>CVPR19</sub>	ResNet-50	183.0	0.939	0.037	0.898	0.918	0.811	0.861	0.071	0.800	0.848	0.639	0.860	0.112	0.714	0.767	0.556
PoolNet <sub>CVPR19</sub>	ResNet-50	273.3	0.944	0.039	0.896	0.921	0.813	0.865	0.075	0.798	0.832	0.644	0.871	0.102	0.759	0.797	0.606
BASNet <sub>CVPR19</sub>	ResNet-34	348.5	0.942	0.037	0.904	0.916	0.826	0.856	0.076	0.798	0.838	0.660	0.851	0.113	0.730	0.769	0.603
U <sup>2</sup> -Net (Ours)	RSU	176.3	0.951	0.033	0.910	0.928	0.836	0.859	0.074	0.797	0.844	0.657	0.861	0.108	0.748	0.786	0.613
U <sup>2</sup> -Net <sup>†</sup> (Ours)	RSU	4.7	0.943	0.041	0.885	0.918	0.808	0.849	0.086	0.768	0.831	0.627	0.841	0.124	0.697	0.759	0.559