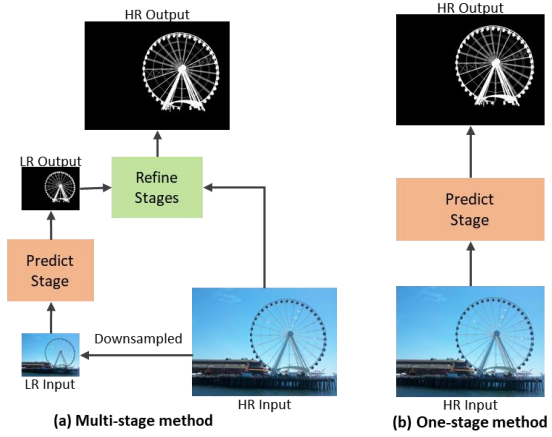
One-stage Progressive Dichotomous Segmentation

Jing Zhu Karim Ahmed Wenbo Li Yilin Shen Hongxia Jin

Samsung Research America AI Center

Problem

Current existing methods can be classified into two categories. Multi-stage methods have the potential to generate superior results, but they often come with time and memory expenses. On the contrary, one-stage methods offer a more straightforward approach with lower computation costs. However, they typically yield inferior performance because many effective high-complexity networks (e.g., transformer), cannot directly handle high-resolution images due to resource limit.



Contribution

- We propose a one-stage framework with an efficient yet effective convolutional attention module that could directly work on high-resolution images for dichotomous segmentation.
- We design the progressive prediction schema into the decoder of the model which enables the gradual refinement of the segmentation map level by level. A multi-scale supervision loss is introduced to enhance the model learning.

Approach

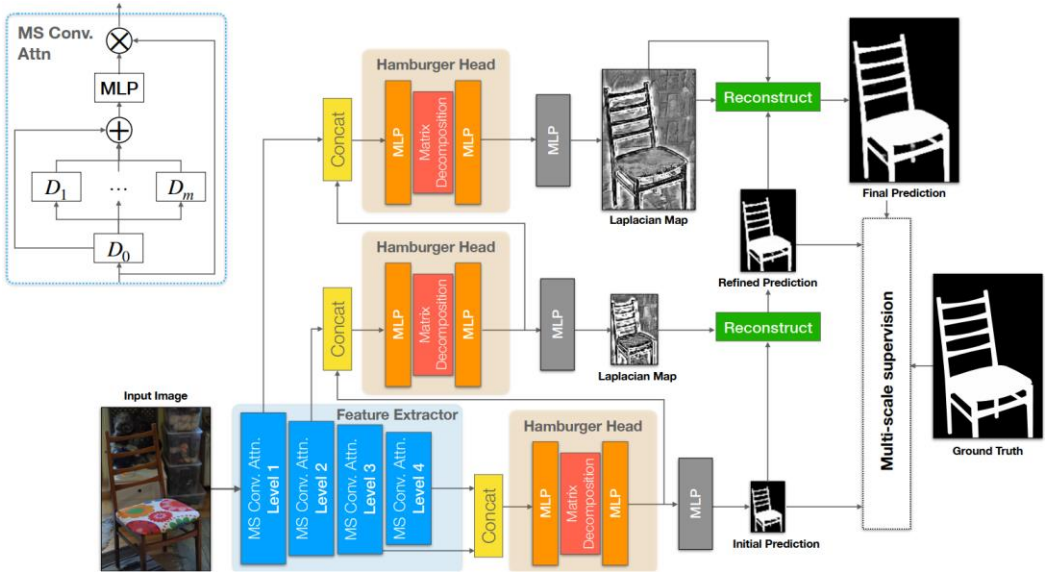


Figure 2: The framework of our proposed model, which consists of a feature extractor with multi-scale convolutioanl attentions to generate effective multi-scale features, and a progressive decoder with hamburger heads to gradually utilize the multi-scale features to achieve high-resolution results. Multi-scale supervision is introduced to enhance the model's training capabilities. An overview of the multi-scale convolution attention is depicted at the upper left, where D_0~D_m denote depth-wise convolutions.

- **Novel Point 1:** a one-stage framework with an efficient yet effective multi-scale convolutional attention feature extractor, enabling direct processing of high-resolution images for dichotomous segmentation.
- **Novel Point 2:** a progressive decoder with a specifically designed progressive prediction mechanism, which generates an initial segmentation map using the lowest-resolution features and progressively refines the map's resolution level by level, leveraging the extracted multi-scale features from the feature extractor.

Experimental Results

Table 1: Comparison of dichotomous segmentation on four DIS5K [28] testing subsets with alternative approaches. Overall performance is computed by taking the mean value of the scores from the four subsets. The best performance has been **bolded** and the second best result is marked in **blue**. Higher F^{mx}, F^{w}, S, E scores and lower M, HCE values indicate the better performance. Our method outperforms all the the single-stage methods with fewer model parameters and computational operations (FLOPs).

Dataset	Metric	Multi-stage		Single-stage					
		HRSOD [36]	InSpyReNet [20]	PFNet [25]	HRNet [30]	SegNeXt [16]	IS-Net [28]	PGNet [32]	Ours
	Params(M)	32.4	90.7	46.5	63.6	**27.6**	44.0	72.7	**28.0**
	Time (ms)	425.7	733.0	**70.5**	172.6	275.7	**80.5**	127.8	287.0
	FLOPs (G)	315.7	461.2	**59.9**	373.8	137.6	159.8	160.3	**129.2**
	Input Size	1024^2	1024^2	416^2	1024^2	1024^2	1024^2	1024^2	1024^2
DIS-TE1	$F^{mx}\uparrow$	0.726	**0.854**	0.646	0.668	0.771	0.740	0.821	**0.822**
	$F^{w}\uparrow$	0.658	**0.792**	0.552	0.579	0.681	0.662	0.728	**0.745**
	$M\downarrow$	0.079	**0.044**	0.094	0.088	0.076	0.074	0.070	**0.069**
	$S\uparrow$	0.766	**0.873**	0.722	0.742	0.789	0.787	0.834	**0.845**
	$E\uparrow$	0.803	**0.893**	0.786	0.797	0.820	0.820	0.846	**0.859**
	$HCE\downarrow$	198	**110**	253	262	177	149	173	**147**
DIS-TE2	$F^{mx}\uparrow$	0.781	**0.895**	0.720	0.747	0.826	0.799	0.841	**0.857**
	$F^{w}\uparrow$	0.714	**0.846**	0.633	0.664	0.741	0.728	0.782	**0.802**
	$M\downarrow$	0.074	**0.038**	0.096	0.087	0.068	0.070	**0.066**	**0.066**
	$S\uparrow$	0.795	**0.905**	0.761	0.784	0.828	0.823	0.842	**0.867**
	$E\uparrow$	0.832	**0.925**	0.829	0.840	0.879	0.858	0.888	**0.903**
	$HCE\downarrow$	467	**255**	567	555	427	**340**	405	346
DIS-TE3	$F^{mx}\uparrow$	0.806	**0.912**	0.751	0.784	0.843	0.830	0.877	**0.882**
	$F^{w}\uparrow$	0.732	**0.868**	0.664	0.700	0.765	0.758	0.803	**0.831**
	$M\downarrow$	0.069	**0.038**	0.092	0.080	0.062	0.064	0.059	**0.058**
	$S\uparrow$	0.819	**0.915**	0.777	0.805	0.841	0.836	0.857	**0.873**
	$E\uparrow$	0.863	**0.942**	0.854	0.869	0.899	0.883	0.906	**0.912**
	$HCE\downarrow$	1007	**523**	1082	1049	871	687	838	**680**
DIS-TE4	$F^{mx}\uparrow$	0.789	**0.902**	0.731	0.772	0.834	0.827	0.859	**0.863**
	$F^{w}\uparrow$	0.726	**0.847**	0.647	0.687	0.755	0.753	0.798	**0.819**
	$M\downarrow$	0.072	**0.046**	0.107	0.092	0.069	0.072	0.067	**0.066**
	$S\uparrow$	0.804	**0.902**	0.763	0.792	0.823	0.830	0.844	**0.870**
	$E\uparrow$	0.848	**0.927**	0.838	0.854	0.883	0.870	0.895	**0.908**
	$HCE\downarrow$	3720	**2336**	3803	3864	3679	2888	3449	**2768**
Overall	$F^{mx}\uparrow$	0.776	**0.890**	0.712	0.743	0.818	0.799	0.849	**0.856**
	$F^{w}\uparrow$	0.708	**0.838**	0.624	0.658	0.735	0.726	0.778	**0.799**
	$M\downarrow$	0.074	**0.042**	0.097	0.087	0.069	0.070	0.065	**0.064**
	$S\uparrow$	0.796	**0.898**	0.756	0.781	0.820	0.819	0.844	**0.864**
	$E\uparrow$	0.837	**0.922**	0.827	0.840	0.870	0.858	0.884	**0.896**
	$HCE\downarrow$	1348	**806**	1427	1432	1289	1016	1216	**986**

Table 2: The ablation studies on DIS-TE1 to verify the effectiveness of each module. We can observe the progressive schema has a greater impact on the model's performance than the multi-scale supervision, while they together yield the best performance.

Decoder		Supervision		$F^{mx}\uparrow$	$F^{w}\uparrow$	$M\downarrow$	$S\uparrow$	$E\uparrow$	$HCE\downarrow$
Single Head	Progressive	Single-scale	Multi-Scale						
✓		✓		0.758	0.676	0.079	0.772	0.813	211
	✓	✓		0.806	0.729	0.073	0.817	0.834	169
	✓		✓	**0.822**	**0.745**	**0.069**	**0.845**	**0.859**	**147**