SeqRank: Sequential Ranking of Salient Objects

Huankang Guan and Rynson W.H. Lau

Department of Computer Science, City University of Hong Kong

Introduction

We propose SeqRank, a novel salient object ranking model, to predict the order of human observer's attention to various scene objects. We propose to detect and rank salient objects in a sequential manner, which aligns with human viewing behavior, and also enables us to explore temporal information. SeqRank achieves the new SOTA on SOR benchmarks, and particularly it improves the SA-SOR/MAE scores by +6.1%/-13% on IRSR benchmark.

Contribution.

1. We propose SeqRank, a novel approach for salient object ranking through modeling human foveal and peripheral vision. It also enables us to explore temporal relationships for a more reliable SOR.
2. We propose two innovative modules, *i.e.* Fovea Module and Sequential Ranking Module, These two modules work coherently to mimic human viewing behavior.
3. We conduct extensive experiments to confirm the effectiveness of our approach, and our model achieves new SOTA results on SOR benchmarks.

Method

SeqRank is composed of a backbone network, a pixel decoder, a Fovea Module (FOM), and a Sequential Ranking Module (SRM). The FOM learns to progressively refine the learnable object queries from image features for an accurate salient object segmentation, while the SRM predicts the next object that is likely to be visited, conditioning on the previous visiting history. By continuously updating the visiting history and invoking the SRM, all salient objects can be detected in a sequential order that reflects how human attention shifts among them.

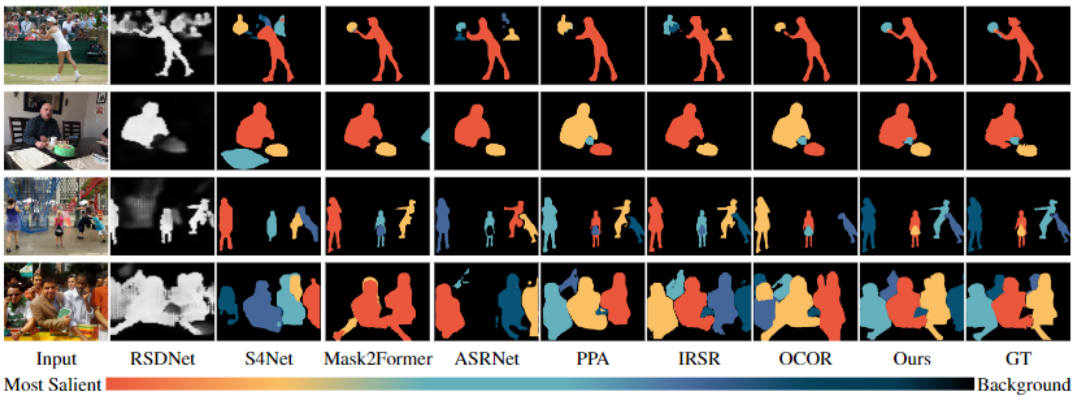


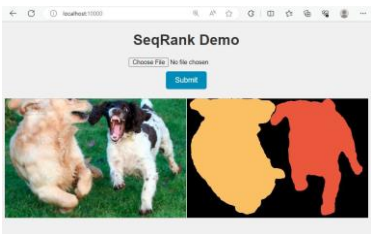
Figure 5: Visual Comparison. Salient instances are colorized using varying color temperatures, ranging from warm to cool, indicating the order in which they are visited. In general, our model produces more favorable results compared to other methods.

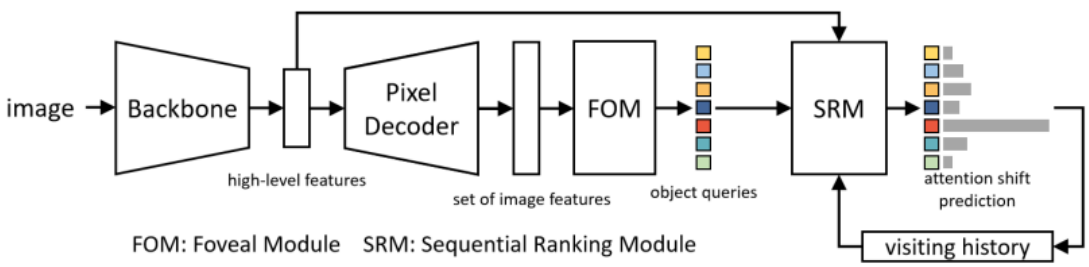
Method	Task	ASSR Test Set (2418)			IRSR Test Set (2929)		
		SA-SOR↑	SOR↑	MAE↓	SA-SOR↑	SOR↑	MAE↓
VST (Liu et al. 2021b)	SOD	0.422	0.643	9.99	0.183	0.571	8.75
MENet (Wang et al. 2023)	SOD	0.369	0.627	9.60	0.162	0.558	8.25
S4Net (Fan et al. 2019)	SID	0.451	0.649	14.4	0.224	0.611	12.1
QueryInst (Fang et al. 2021b)	IS	0.596	0.865	8.52	0.538	0.816	7.13
Mask2Former (Cheng et al. 2022)	IS	0.635	0.867	7.31	0.521	0.799	7.14
RSDNet (Islam, Kalash, and Bruce 2018)	SOR	0.386	0.692	18.2	0.326	0.663	18.5
ASRNet (Siris et al. 2020)	SOR	0.590	0.770	9.39	0.346	0.681	9.44
PPA (Fang et al. 2021a)	SOR	0.635	0.863	8.52	0.521	0.797	8.08
IRSR (Liu et al. 2021a)	SOR	0.650	0.854	9.73	0.543	0.815	7.79
OCOR (Tian et al. 2022a)	SOR	0.541	**0.873**	10.2	0.504	0.820	8.45
Ours	SOR	**0.685**	0.870	**7.22**	**0.576**	**0.822**	**6.20**

Table 1: Quantitative Comparison. SOD: Salient Object Detection. SID: Salient Instance Detection. IS: Instance Segmentation. SOR: Salient Object Ranking. The best is marked in bold and the second-best is marked with an underline.

We further evaluate SeqRank with newly collected examples from Internet. SeqRank works well!

Code is Available!
Code package also includes a web demo!





Conclusion.

In this work, we propose SeqRank, which detects and ranks salient objects in a sequential manner, aligning with human viewing behavior and yielding a favorble results. We hope this work can facilitate various applications, that require understand human attention.