上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

Saliency Prediction on Omnidirectional Images with Brain-Like Shallow Neural Network

Dandan Zhu, Yongqing Chen, Xiongkuo Min, Defang Zhao, Yucheng Zhu, Qiangqiang Zhou, Tian Han, Guangtao Zhai and Xiaokang Yang

Artificial Intelligence Institute, Shanghai Jiao Tong University
Hainan Air Traffic Management Sub-Bureau
Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University
School of Software Engineering, Tongji University
Department of Computer Science, Stevens Institute of Technology

1. Introduction

- ODIs have higher resolution, making it difficult for streaming and rendering;
- Although these deep feedforward CNNs perform well in saliency prediction task, they have the following limitations:
  - ✓ Deep feedforward CNNs are too complex in design and contain a vast number of layers, which is difficult to map to the ventral stream structure of the brain visual system.
  - ✓ They lack biologically-important brain structures (i.e. recurrenc connectivity), which is difficult to match the complex neurons states in the brain.

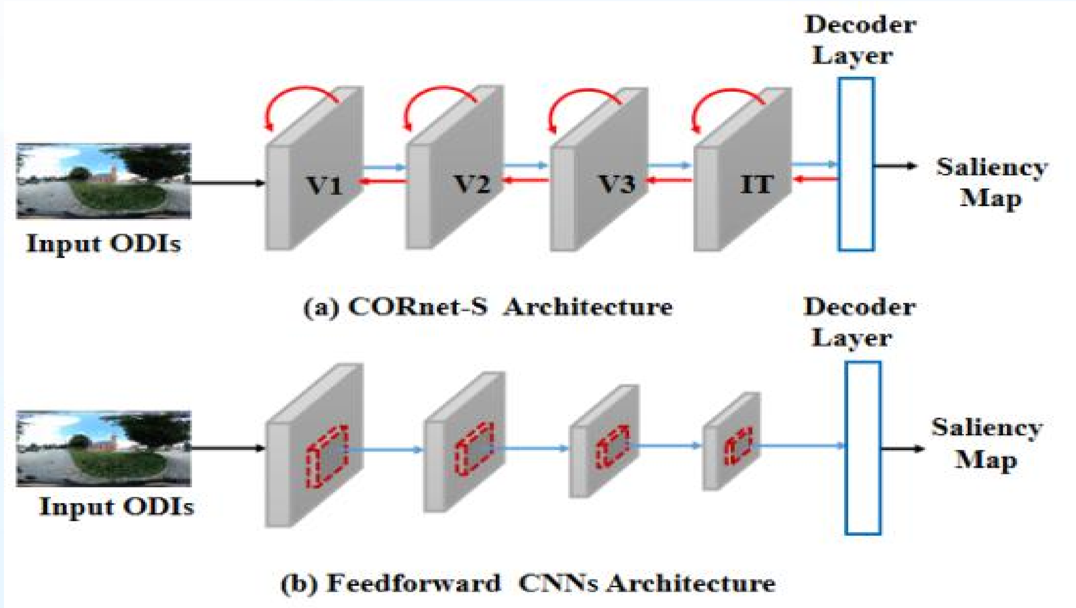

Fig.1 Architecture overview of deep recurrent CORnet-S and deep feedforward CNNs.

2. Proposed Model

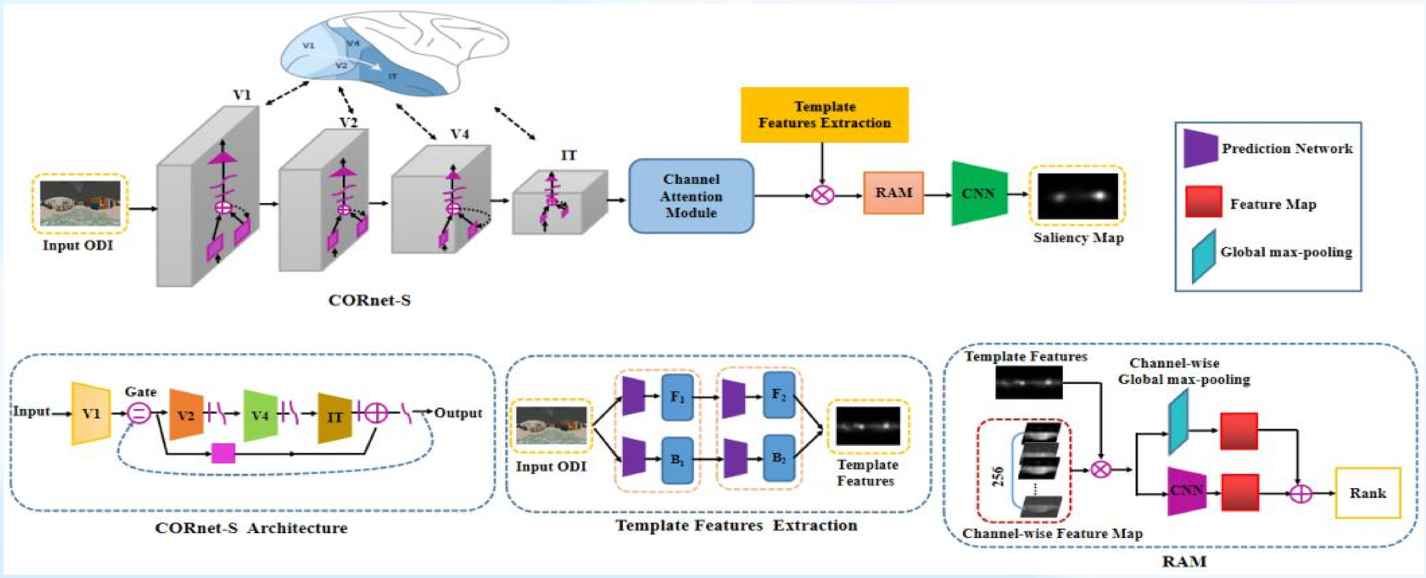

Fig.2 Architecture overview of proposed brain-like saliency prediction model.

2.1 CORnet-S module

The CORnet-S module is a lightweight ANN with four computational areas, conceptualized as analogous to the ventral visual areas (V1, V2, V4 and IT) and recurrent connections. We modify the original CORnet-S structure and add a channel attention module behind the IT area and the channel attention maps are calculated as follows:

$$F_c = \sigma(MLP(Avgpool(f)) + MLP(Maxpool(f)))$$
$$= \sigma(w_1(w_0(f^c_{avg})) + w_1(w_0(f^c_{\max}))),$$

where σ is the sigmoid function, w_0 and w_1 are the MLP weights.

2.2 Template feature extraction module

Specifically, we employ a two-stage network to learn part attention maps. The first stage individually predicts foreground attention F^1 and background attention B^1 by two independent prediction networks:

$$F^1 = \varphi^1(F^{TF}), \; B^1 = \phi^1(F^{TF}),$$

where F^{TF} is the feature map obtained by vgg16 network, φ^1 and ϕ^1 denote two prediction networks. In second stage, the attention maps obtained by the first stage are further refined and the specific equations are expressed as follows:

$$F^2 = \varphi^2(F \mid F^1, B^1), \;\; B^2 = \phi^2(F \mid F^1, B^1),$$

where F^2 and B^2 are the foreground attention map and background attention map, respectively.

2.3 Ranking attention module

To calculate the ranking scores of the channel-wise feature maps, we utilize a two-layer network f_n refinement by summing with the channel-wise global max-pooling of the tensor $f_{\max}$ in an element-wise manner:

$$r_i = f_n(S_i) + f_{\max}(S_i),$$

For the ranking scores of channel-wise feature maps in S_i, we need to rank these channel-wise feature maps according to the ranking score r_i:

$$S_i' = rank(S_i \mid r_i),$$

where S_i' represents the ordered channel-wise feature maps afeter rank. Then we need to select important features for the final fine-grained saliency prediction and discard redundant features.

3. Experiments

3.1 Qualitative comparison

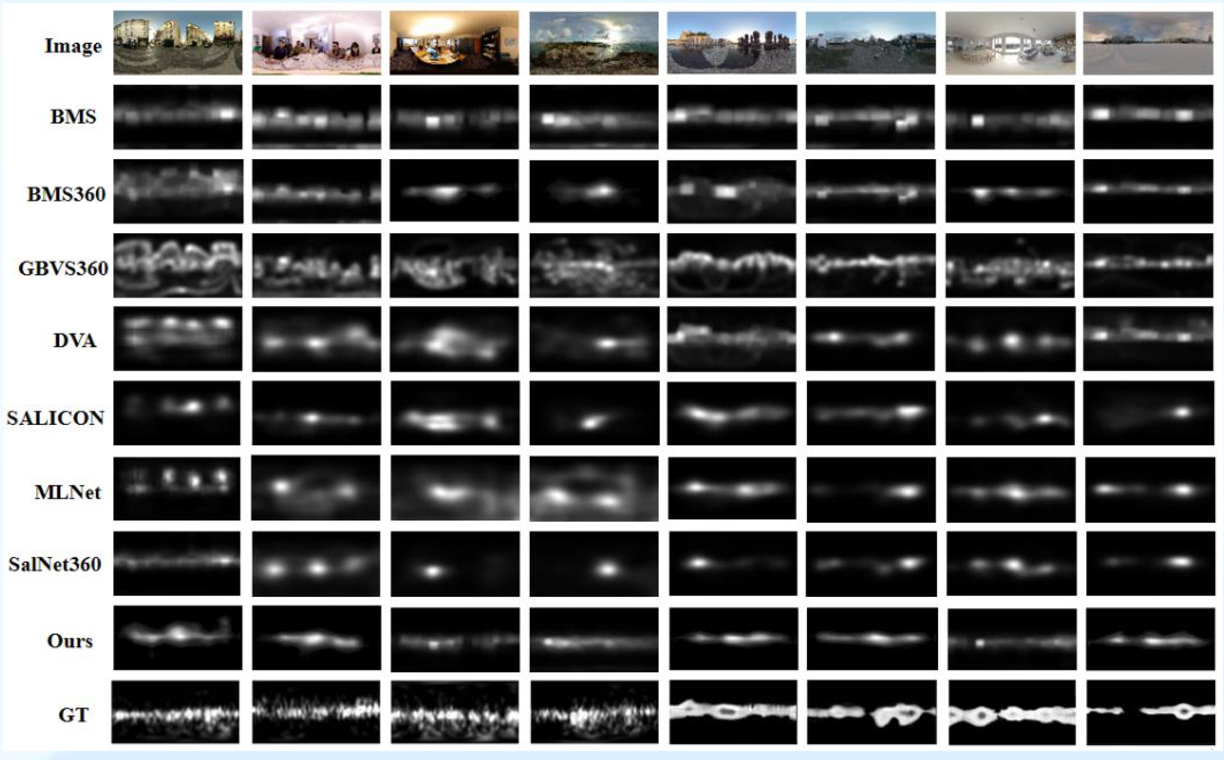

Fig.3 Visual comparison of our results with other approaches for predicting saliency maps of head fixations on the Salient360! dataset and the ODS dataset.

3.2 Quantitative comparison

Evaluation metrics: NSS, CC, AUC and KL divergence.

Methods	Salient360!				ODS			
	CC	AUC	NSS	KL divergence	CC	AUC	NSS	KL divergence
BMS [9]	0.562	0.721	0.963	0.589	0.545	0.687	0.942	0.634
BMS360 [4]	0.716	0.754	1.372	0.583	0.648	0.724	1.224	0.615
GBVS360 [4]	0.587	0.836	0.994	0.562	0.569	0.696	0.975	0.571
DVA [16]	0.728	0.772	1.394	0.594	0.612	0.765	1.327	0.541
SALICON [13]	0.745	0.781	0.998	0.554	0.724	0.769	0.987	0.538
MLNet [17]	0.764	0.812	1.012	0.713	0.745	0.797	1.081	0.686
SalNet360 [11]	0.795	0.843	1.581	0.514	0.776	0.821	1.565	0.534
Ours	**0.913**	**0.922**	**2.020**	**0.498**	**0.892**	**0.878**	**2.015**	**0.512**

Table 1: Quantitative comparison of our model with other methods over Salient360 and ODS datasets.