Saliency Detection with Recurrent Fully Convolutional Networks

TIWAKI

Linzhao Wang[1], Lijun Wang[1], Huchuan Lu[1], Pingping Zhang[1], Xiang Ruan[2]

[1] Dalian University of Technology [2] TIWAKI Corporation

1. Introduction

1.1 Problem

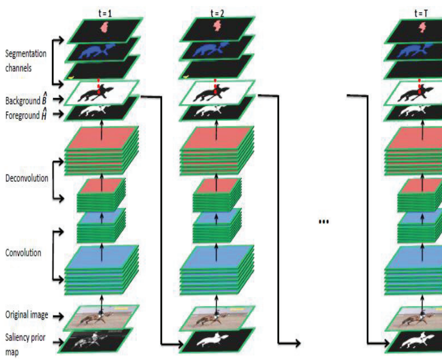
- **Saliency priors,** which are shown to be effective in previous work, are completely discarded by most CNN based methods.
- **A limited size** of local image patches is considered by most CNNs methods.
- **Binary classification** problems, which represent saliency detection, have relatively weak supervision information.

1.2 Our Solution

- **Incorporation** for the saliency priors into the network to facilitate training and inference.
- **The recurrent structure** to refine the coarse inference from previous time steps.
- **A RFCN pre-training method** for saliency detection using semantic segmentation data to both leverage strong supervision from multiple object categories and capture the intrinsic representation of generic objects.

2. Overview

- Employs three kind of low-level contrast features, including color, intensity and orientation, and the center prior knowledge to introduce **saliency prior maps**.
- Train the RFCN with two stage training strategy, pre-training on the segmentation data set and fine-tuning on the saliency data set. The recurrent structure can incorporate the saliency prior maps into the CNNs with an end-to-end training method.
- Refine the saliency maps with a post-processing method which can improve the performance of RFCN. Edge-preserving maps are produced with the computation of color confidence and spatial confidence.



3. Algorithm

3.1 Saliency Prior Maps

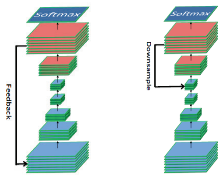
- We encode prior knowledge into a saliency prior map which serves as the input to the network. The priors include color, intensity and orientation feature contrast. Then we integrate these priors together and filter the result with a gaussian function, which proves center prior information.

.

$$P(s_i) = \mathcal{U}(s_i) \times (\mathcal{G}(s_i) + \mathcal{I}(s_i) + \mathcal{O}(s_i))$$

3.2 Recurrent Architecture

- Architecture 1: Forward propagation of the whole network is conducted in every time step, which is very expensive in terms of both computation and memory. We adopt this architecture for accuracy in our paper.
- Architecture 2: In the t-th time step, the predicted foreground map in the last time step serves as saliency prior map. The deconvolution part takes the convolution feature map as well as the foreground map to refine the saliency prediction.



Architecture 1 Architecture 2

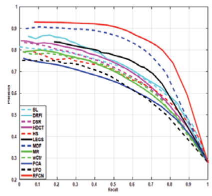
3.3 Training RFCN

- Pre-training is conducted on the PASCAL VOC 2010 semantic segmentation data set. Saliency detection and semantic segmentation are highly correlated but essentially different in that saliency detection aims at separating generic salient objects from background, whereas semantic segmentation focuses on distinguishing objects of different categories.
- After pre-training, we modify the RFCN network architecture by removing the first C+1 channels of the last feature map and only maintaining the last two channels.

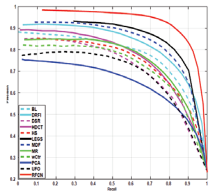
3.4 Post-Processing

- Given the final saliency score map predicted by the RFCN, we first segment the image into foreground and background regions by thresholding it with its mean saliency score.
- We then compute a spatial confidence and a color confidence score for each pixel
- Finally, we weight the predicted saliency scores by spatial and color confidences to dilate the foreground region.

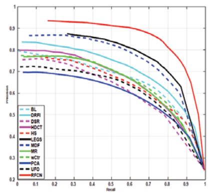
4. Results



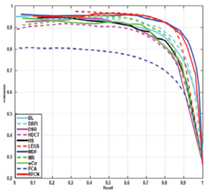
PR-curves on SOD



PR-curves on ECCSD



PR-curves on BSD



PR-curves on SED1

*	SOD		ECSSD		PASCAL-S		SED1	
	F-measure	AUC	F-measure	AUC	F-measure	AUC	F-measure	AUC
RFCN	0.7426	0.9053	0.8340	0.9714	0.7468	0.9453	0.8502	0.9640
MTDS	0.6978	0.9233	0.7589	0.9009	0.7310	0.9287	-	-
LEGS	0.6492	0.8117	0.7887	0.9230	0.6951	0.8857	0.8414	0.9328
MDF	0.6966	0.8532	0.7557	0.9180	0.6562	0.8806	0.8194	0.9710
BL	0.5723	0.8503	0.6825	0.9147	0.5668	0.8633	0.7675	0.9528
DRFI	0.6031	0.8464	0.7337	0.9391	0.6159	0.8913	0.8024	0.9528
wCtr	0.5978	0.8014	0.6774	0.8779	0.5972	0.8433	0.7889	0.9159
DSR	0.5968	0.8210	0.6636	0.8604	0.5513	0.8079	0.7877	0.9086
MR	0.5697	0.7899	0.6932	0.8820	0.5881	0.8205	0.8255	0.9223
HS	0.5210	0.8145	0.6363	0.8821	0.5278	0.8330	0.7426	0.9161
PCA	0.5370	0.8212	0.5796	0.8737	0.5298	0.8371	0.6256	0.9030
UFO	0.5480	0.7840	0.6442	0.8587	0.5502	0.8088	-	-

(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k)

Comparisons of saliency maps. Top, middle and bottom two rows are images from the SOD, ECSSD, PASCAL-S and SED1 data sets, respectively.(a) Original images, (b)ground truth, (c)Our RFCN method, (d)LEGS, (e)MDF, (f)DRFI, (g)wCtr, (h)HDCT, (i)DSR, (j)MR, (k)HS.

5. Conclusions

- We propose a recurrent fully convolutional network based saliency detection.
- Our method integrates low level saliency prior knowledge and fully convolutional neural networks with a recurrent structure.
- Experimental results on five benchmark data sets show that the proposed algorithm achieves favorable results against the state-of-the-art methods.