Introduction

What is Proto-Object?
Proto-objects can be seen as pre-attentive structures coherent in limited space and time. Proto-objects can bind various low-level features over a small region of space and a short period of time, becoming "highest-level output of low-level vision".

Contrary to precise object recognition after the deployment of attention, the notion of proto-object is more like object-level gist that can be computed rapidly in parallel over the entire visual field (as illustrated in Fig. 1).

Why Low Resolution?
Proto-objects are computed in parallel over the entire visual field where most regions are in lower resolution than the fovea area [1]. Human can perceive objects well even they are in low a resolution of 16x16 [2]. Fixations from lower resolution images can predict well fixations on corresponding higher resolution images [3].

Data Preparation
Large-scale attention data from SALICON dataset [4]: Salient patches: multi-scale patches in low resolution sampled from top five local maxima in the blurred ground truth maps.
Non-salient patches: randomly sampled from the positions where saliency values are less than the mean of the blurred ground truth maps.

The Model

Convolutional Neural Network (CNN): Model saliency prediction as a binary classification problem on salient and non-salient patches in low resolution. Multiple scales in low resolution are concatenated and linear integrated at the final stage.

Two CNN structures are used for each single scale:
2-layer model: Input Size 16x16, C(5,64)-MP(2)-C(5,512)-MP(2)
3-layer model: Input Size 36x36, C(5,64)-MP(2)-C(5,128)-MP(2)-C(5,512)-MP(2)
where C(\(f\),\(n\)) indicates \(n\) convolution kernels in size of \(f \times f\), MP(\(f\)) indicates non-overlap max pooling in \(f \times f\).

Results

By training on salient and non-salient patches in low resolution, proto-object representations can be learned out in a deep architecture similar to the conceptual schematic described in [1]. The proposed models are competitive in predicting eye fixations in natural scenes compared with state-of-the-art saliency models.

Conclusion

This poster can be downloaded at: http://bit.ly/1P7OJJS

Reference