Co-saliency Detection via Looking Deep and Wide

Dingwen Zhang[1], Junwei Han[1]*, Chao Li[1] and Jingdong Wang [2]
[1] Northwestern Polytechnical University; [2] Microsoft Research, P.R. China

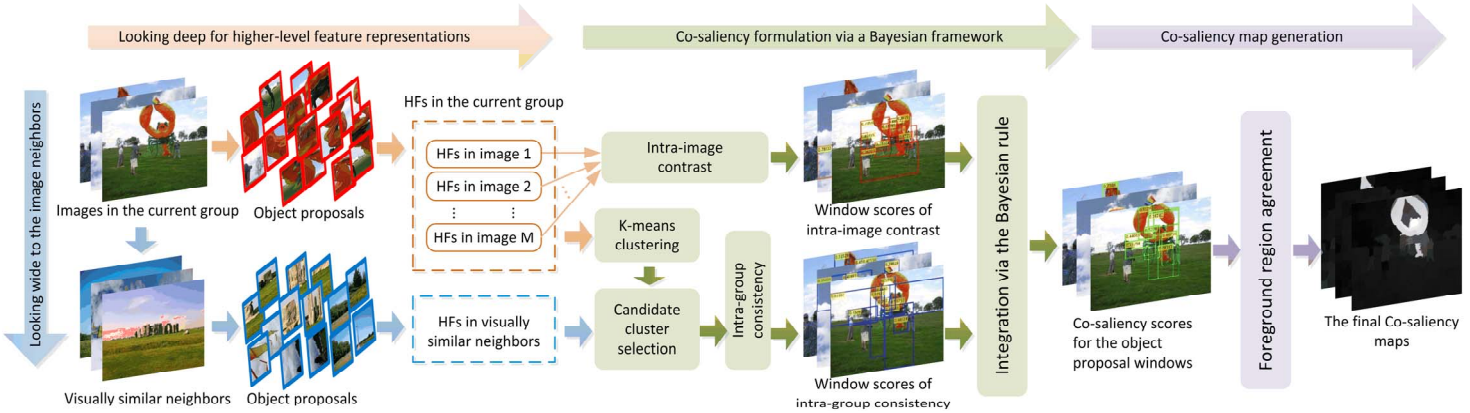

Figure 1: The paradigm of the proposed co-saliency detection approach, where HF denotes the higher-level feature.

With the goal of effectively identifying common and salient objects in a group of relevant images, co-saliency detection has become essential for many applications such as video foreground extraction, surveillance, image retrieval, and image annotation. In this paper, we propose a unified co-saliency detection framework by introducing two novel insights: 1) looking deep to transfer higher-level representations by using the convolutional neural network with additional adaptive layers could better reflect the properties of the co-salient objects, especially their consistency among the image group; 2) looking wide to take advantage of the visually similar neighbors beyond a certain image group could effectively suppress the influence of the common background regions when formulating the intra-group consistency.

To explore these useful information (i.e. the deep information and wide information) for co-saliency detection, we propose a unified framework as shown in Figure 1. Firstly, the wide and deep information are explored for the object proposal windows extracted in each image. Then the co-saliency scores are calculated by integrating the intra -image contrast and intra-group consistency via a principled Bayesian formulation. Finally the window-level co-saliency scores are converted to the superpixel-level co-saliency maps through a foreground region agreement strategy. In summary, the novelties of this paper are threefold:

1) We propose to explore the properties of the co-salient objects by using CNN with additional transfer layers, which brings deep information for discovering the higher-level properties of the co-salient objects.
2) We introduce the idea to make use of the visually similar neighbors from the other image groups, which brings wide information for suppressing the common background regions in co-saliency detection.
3) We propose a simple but effective framework which uniformly embeds the deep and wide information in co-saliency detection through a Bayesian formulation and a foreground region agreement strategy.

In order to represent image regions with deep information, we adopt BING [1] to extract object proposal windows (OPs) in each image firstly, and then build higher-level features for them via a CNN [2] with additional transfer RBM [3] layer. Due to the scarce of training data, directly learning a whole CNN from the co-saliency dataset is problematic. To solve this problem, we pre-train a CNN on the source data (i.e. the images in the ImageNet dataset) firstly. Then, we stack two additional transfer layers formed by two fully connected RBMs. The objective of a RBM is to minimize an energy function defined as the joint distribution over the visible units and hidden units, thus maximizing probability for the training data in an unsupervised manner. As the two RBM layers are trained on the target domain, they can adapt the whole network to extract domain-specific higher-level features in the co-saliency dataset.

In order to estimate the probability of each OP to be the co-salient region, we extend the Bayesian framework proposed by [4] to formulate this problem. To be specific, let $\{x_m^p\}$ denote the OPs in image I_m, and $\{y_m^p\}$ denote whether or not a certain OP belongs to a co-salient region. Then, we define the co-saliency of $\{x_m^p\}$ as:

$$\begin{aligned} \mathrm{Cosal}(x_m^p) &= \Pr(y_m^p = 1 \mid x_m^p) \\ &= \frac{\Pr(x_m^p \mid y_m^p = 1)\Pr(y_m^p = 1)}{\Pr(x_m^p)} \\ &\propto \underbrace{\frac{1}{\Pr(x_m^p)}}_{\substack{\text{Intra-image} \\ \text{contrast}}} \underbrace{\Pr(x_m^p \mid y_m^p = 1)}_{\substack{\text{Intra-group} \\ \text{consistency}}} \end{aligned} \quad (1)$$

where the intra-image contrast is used to explore the discrimination of each OP from the image context, and the intra-group consistency is used to explore the similar characteristics of the co-salient regions.

In order to capture the co-salient objects with finely defined boundaries, we convert the window-level co-saliency scores to the superpixel-level saliency maps. Inspired by [5] we propose a foreground region agreement (FRA) strategy to tackle this problem. The proposed FRA strategy takes into consideration the overall agreement between salient regions over the whole image and the whole group with containing two phases, i.e. the intra-image FRA and the intra-group FRA. The intra-image FPA is implemented by first selecting some estimated foreground queries in each image, and then using the Manifold Ranking algorithm [35] to compute the agreement of each image region to those queries. The intra-group FRA is implemented by modifying the co-saliency of each superpixel based on the agreement to its similar regions among other images in the image group.

Finally, we conducted comprehensive experiments to evaluate the proposed approach on two benchmark datasets. Qualitative and quantitative comparisons with other state-of-the-art co-saliency methods demonstrated the effectiveness of the proposed approach. In addition, we also demonstrate that competitive performance can be obtained when directly applying our method in image co-segmentation.

[1] M.-M. Cheng. BING: Binarized normed gradients for objectness estimation at 300fps. In *Proc. CVPR*, 2014.
[2] P. Sermanet. Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229, 2013.
[3] Y. Bengio. Learning deep architectures for AI. *Foundations and Trends in Machine Learning,* 2(1):1-127, 2009.
[4] L. Zhang. SUN: A Bayesian framework for saliency using natural statistics. *JOV*, 8(7):32, 2008.
[5] Y. Jia, and M. Han. Category-Independent Object-level Saliency Detection. In *Proc. ICCV*, 2013.