

Diagnosing State-Of-The-Art Object Proposal Methods

Hongyuan Zhu¹
zhuh@i2r.a-star.edu.sg
Shijian Lu¹
slu@i2r.a-star.edu.sg
Jianfei Cai²
ASJFCAL@ntu.edu.sg
Guangqing Lee³
134174B@mymail.nyp.edu.sg

¹ Institute for Infocomm Research,
A*Star,
Singapore
² Nanyang Technological University,
Singapore
³ Nanyang Polytechnic,
Singapore

Recent top performing methods in PASCAL VOC [6] and ImageNet [13] make use of object proposal to replace exhaustive window search. Object proposal’s effectiveness is rooted in the assumption that there are general cues to differentiate objects from the background. Since the very first work by Alexe et al. [1], many object proposal methods have been proposed [2, 3, 4, 5, 7, 8, 10, 11, 12, 14, 15] and tested on various large scale datasets [6, 9, 13], and their overall detection rates versus different thresholds or window number have also been reported. Yet such partial performance summaries give us little idea of a method’s strengths and weaknesses for further improvement, and users are still facing difficulties in choosing methods for their applications. Therefore, more detailed analysis of existing state-of-the-arts is critical for future research and applications.

Our contributions can be summarized in three aspects. First, we investigate the influence of object-level characteristics over state-of-the-art object proposal methods for the first time. Although there are some similar works in categorical object detection, few research has been conducted on object proposal side to the best of our knowledge. Second, we introduce the concept of localization latency to evaluate a method’s localization efficiency and accuracy. Third, we create a fully annotated PASCAL VOC dataset with various object-level characteristics to facilitate our analysis. The annotations take us nearly one month’s time which will be released to facilitate further related research.

Our experiments are based on PASCAL VOC2007 test set, which has been widely used in evaluating object proposal methods. A proposed window B is treated as detected if its Intersection-over-Union (IoU) with a ground truth bounding box \bar{B} : $IoU(\bar{B}, B) = \frac{area(B \cap \bar{B})}{area(B \cup \bar{B})}$ is above a certain threshold T . We first study the localization accuracy of the existing methods. The region based methods have higher localization accuracy than window based methods. MCG and SelectiveSearch are the top performing region based methods, though window based EdgeBox shows comparable performance. The localization accuracy for region based methods are similar. One potential explanation is that all region based methods follow similar pipeline by grouping superpixels with either learned or hand-crafted edge measures.

A good object proposal method should not only produce candidates with high accuracy, but also use as less windows as possible. To summarize a method’s performance in terms of the accuracy and window number, we propose the **localization latency** metric:

$$LocalizationLatency = \frac{1}{|T|} \sum_T \log_2 \left(\frac{1}{M} \sum_K \sum_m OBJ(m, T, K) \right) \quad (1)$$

where K ($K \in \{100, 1000, 10000\}$) are the number of windows W_1^I, \dots, W_K^I in returned descending order. Let $OBJ(m, T, K)$ be the total number of sampled windows to localize all instances of the target category in image m under varying threshold $T \in \{0.5, 0.7, 0.85\}$ and K (i.e. $OBJ(m, T, K) = \sum_i \min\{k \in K : IOU(W_k^m, B_i^m) > T\}$). If using k windows still can’t localize the object, k is set to $K + 1$. M is the number of images contain the instances. From our experiments, the window based methods have lower localization latency than region based methods. Our interpretation for such result is that window based approach inherently assumes that objects are spatially compact.

We also examine the impact of additional properties: natureness, iconic view, object size, aspect ratio, color contrast, shape regularity and texture. Our study reveals that the cues of existing methods are generalized to both ‘nature’ and ‘man-made’ objects. While, existing methods are sensitive to the changes in viewpoint, object size, aspect ratio, color

contrast and shape regularity. While the impact of texture is pronounced.

- [1] Bogdan Alexe, Thomas Deselaers, and Vittorio Ferrari. What is an object? In *CVPR*, 2010.
- [2] Pablo Andrés Arbeláez, Jordi Pont-Tuset, Jonathan T. Barron, Ferran Marqués, and Jitendra Malik. Multiscale combinatorial grouping. In *CVPR*, 2014.
- [3] João Carreira and Cristian Sminchisescu. Constrained parametric min-cuts for automatic object segmentation. In *CVPR*, 2010.
- [4] Ming-Ming Cheng, Ziming Zhang, Wen-Yan Lin, and Philip H. S. Torr. BING: binarized normed gradients for objectness estimation at 300fps. In *CVPR*, 2014.
- [5] Ian Endres and Derek Hoiem. Category independent object proposals. In *ECCV*, 2010.
- [6] Mark Everingham, Luc J. Van Gool, Christopher K. I. Williams, John M. Winn, and Andrew Zisserman. The pascal visual object classes (VOC) challenge. *IJCV*, 88(2):303–338, 2010.
- [7] Ahmad Humayun, Fuxin Li, and James M. Rehg. RIGOR: reusing inference in graph cuts for generating object regions. In *CVPR*, 2014.
- [8] Philipp Krähenbühl and Vladlen Koltun. Geodesic object proposals. In *ECCV*, 2014.
- [9] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: common objects in context. In *ECCV*, 2014.
- [10] Santiago Manen, Matthieu Guillaumin, and Luc J. Van Gool. Prime object proposals with randomized prim’s algorithm. In *ICCV*, 2013.
- [11] Esa Rahtu, Juho Kannala, and Matthew B. Blaschko. Learning a category independent object detection cascade. In *ICCV*, 2011.
- [12] Pekka Rantalankila, Juho Kannala, and Esa Rahtu. Generating object segmentation proposals using global and local search. In *CVPR*, 2014.
- [13] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael S. Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *CoRR*, abs/1409.0575, 2014. URL <http://arxiv.org/abs/1409.0575>.
- [14] Koen E. A. van de Sande, Jasper R. R. Uijlings, Theo Gevers, and Arnold W. M. Smeulders. Segmentation as selective search for object recognition. In *ICCV*, 2011.
- [15] C. Lawrence Zitnick and Piotr Dollár. Edge boxes: Locating object proposals from edges. In *ECCV*, 2014.