Improving Object Proposals with Multi-Thresholding Straddling Expansion

Xiaozhi Chen, Huimin Ma, Xiang Wang, Zhichen Zhao
Department of Electronic Engineering, Tsinghua University

Recent advances in object detection have exploited object proposals to speed up object searching. Most object proposal generators can be classified into two categories: *objectness-based* models [1, 4, 8] which focus on designing an objectness measure to directly distinguish objects from amorphous background stuff, and *similarity-based* models [2, 3, 6, 7] which merge similar regions based on diverse cues or do multiple figure-ground segmentations.

Our work is motivated by the following problems. First, most objectness-based models suffer from strong localization bias, which means they can hardly achieve high recall consistently across various intersection over union (IoU) thresholds. Second, diversification strategies required by most models are usually computationally expensive. To achieve high accuracy, many models have to utilize multiple segmentations to diversify object proposals, at the cost of much more computations. Our solution for these issues is based on two main contributions:

- A measurement for localization bias, which enlightens a direction to improving the quality of object proposals.
- A box refinement method, namely Multi-Thresholding Straddling Expansion (MTSE), which effectively reduces localization bias via fast diversification.

Our key idea is to utilize superpixels straddling to refine bounding boxes. Given an image and a set of initial bounding boxes, we first align bounding boxes with potential boundaries preserved by superpixels. Then we perform multi-thresholding expansion guided by superpixels straddling for each bounding box. Such a simple procedure benefits object proposals from numerous aspects: 1) significant reduction in localization bias, 2) fast diversification effect requiring only one segmentation, and 3) seamless integration into any existing model to improve their accuracy with little computation overhead.

Superpixel Tightness. To understand localization bias, we introduce an indicator, *superpixel tightness (ST)*, which measures how tight a bounding box fits around an object. The superpixel tightness measure is indicative of the localization bias of object proposal generators. Figure 1 (a) shows that objects and background regions yield quite different distribution of superpixel tightness. Figure 1 (b) shows that existing objectness-based models have strong bias to low superpixel tightness while while most similarity-based methods spread more evenly across various tightness. This accords with their bias in localization accuracy, which means we can use *ST* distribution to measure the localization bias.

MTSE. We first define the *straddling degree* of a superpixel s with regard to a bounding box b as $SD(s,b) = |s \cap b|/|s|$. Given an initial bounding box b, we expand it according to the straddling degrees of superpixels. Formally, we define *straddling expansion* with a threshold δ as the following refinement:

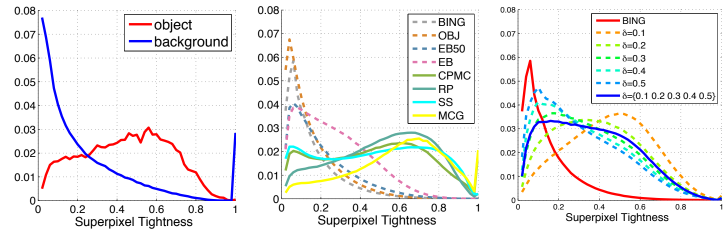
$$\mathcal{S}_{\delta}(b) = \mathcal{S}_{in}(b) \cup \{s \in \mathcal{S}_{\theta} | SD(s,b) \geq \delta\}, \tag{1}$$

where $\mathcal{S}_{in}(b) = \{s \in \mathcal{S}_{\theta} | SD(s,b) = 1\}$ is the set of superpixels entirly inside b. To achieve higher diversity, we use multiple δ's to perform straddling expansion. We take BING [4] to demonstrate the effect of MTSE in Figure 1 (c). It shows that MTSE is able to reduce its bias in *ST* distribution, which indicates an improvement of localization accuracy (Figure 2).

Box Alignment. To reduce the lost of localization precision caused by regular sampling for bounding boxes initialization, which is used by most objectness-based models, we propose an efficient method to align bounding boxes with potential object boundaries preserved by superpixels. Bounding boxes after alignment are fed into MTSE.

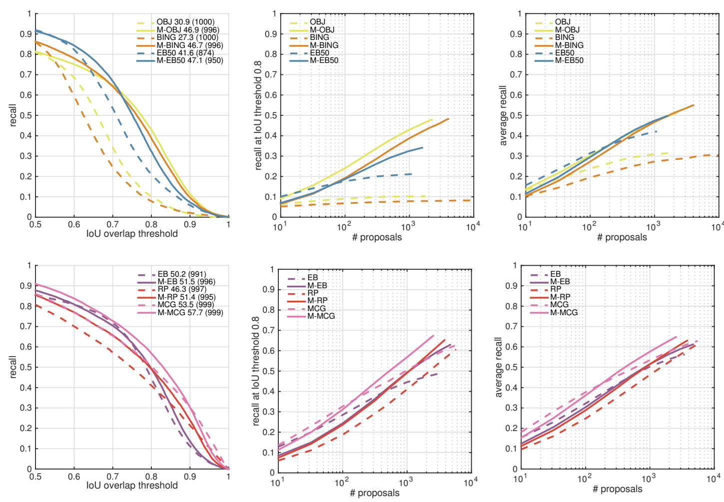
Results. As shown in Figure 2. The proposed MTSE effectively improves existing models in terms of both recall and localization accuracy. Specifically, when using less than 2000 proposals, M-MCG achieves **94.2%** recall at IoU of 0.5. For the strict 0.8 IoU threshold, M-MCG still has **63.8%** recall. In addition, the runtime for MTSE is only 0.15s, thus bring little computation overhead to existing models.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.



(a) (b) (c)

Figure 1: Distributions of superpixel tightness for (a) ground truth objects and background regions on PASCAL VOC2007, (b) 1000 object proposals generated by several models: BING [4], OBJ [1], EB50, EB [8], CPMC [3], RP [6], SS [7] and MCG [2], and (c) BING after applying MTSE.



1000 proposals per image 80%-recall Average recall

Figure 2: Performances of numerous models (in dashed lines) and their improved versions (in solid lines) using MTSE. For the recall-overlap curves, numbers next to labels indicate average recall [5] and average number of proposals per image.

[1] B. Alexe, T. Deselaers, and V. Ferrari. Measuring the objectness of image windows. *IEEE TPAMI*, 34(11):2189–2202, Nov 2012.

[2] Pablo Arbelaez, J. Pont-Tuset, Jon Barron, F. MarquÃl's, and Jitendra Malik. Multiscale combinatorial grouping. In *IEEE CVPR*, 2014.

[3] Joao Carreira and Cristian Sminchisescu. Constrained parametric mincuts for automatic object segmentation. In *IEEE 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 *IEEE CVPR*, 2014.

[5] J. Hosang, R. Benenson, P. Dollár, and B. Schiele. What makes for effective detection proposals? *arXiv:1502.05082*, 2015.

[6] S. Manen, M. Guillaumin, and L. Van Gool. Prime object proposals with randomized prim's algorithm. In *IEEE ICCV*, 2013.

[7] K.E.A. Van de Sande, J.R.R. Uijlings, T. Gevers, and A.W.M. Smeulders. Segmentation as selective search for object recognition. In *IEEE ICCV*, 2011.

[8] C. Lawrence Zitnick and Piotr Dollár. Edge boxes: Locating object proposals from edges. In *ECCV*, 2014.