Improving Object Proposals with Multi-Thresholding Straddling Expansion

Xiaozhi Chen Huimin Ma Xiang Wang Zhichen Zhao
Department of Electronic Engineering, Tsinghua University
{chenxz12,wangxiang14,zhaozc14}@mails.tsinghua.edu.cn, mhmpub@tsinghua.edu.cn

Abstract

Recent advances in object detection have exploited object proposals to speed up object searching. However, many of existing object proposal generators have strong localization bias or require computationally expensive diversification strategies. In this paper, we present an effective approach to address these issues. We first propose a simple and useful localization bias measure, called superpixel tightness. Based on the characteristics of superpixel tightness distribution, we propose an effective method, namely multi-thresholding straddling expansion (MTSE) to reduce localization bias via fast diversification. Our method is essentially a box refinement process, which is intuitive and beneficial, but seldom exploited before. The greatest benefit of our method is that it can be integrated into any existing model to achieve consistently high recall across various intersection over union thresholds. Experiments on PASCAL VOC dataset demonstrates that our approach improves numerous existing models significantly with little computational overhead.

1. Introduction

In recent years, object proposal generation [3, 4, 6, 7, 8, 15, 16, 17, 18, 19, 20, 24] has become a promising technique for many vision recognition tasks, especially for class-specific object detection. Instead of handling with tremendous amount of bounding boxes in sliding windows fashion, object proposal generation selects much fewer candidate bounding boxes (typically from hundreds to a few thousands per image) that cover most of the objects in the image. This technique benefits object detection from two aspects: speeding up the computation by reducing the candidate bounding boxes and improving the detection accuracy by allowing the usage of more sophisticated learning machinery. Recent detection models using object proposals [12, 22] have shown superior performance over sliding window based methods [10].

Object proposals are represented in the form of segment or bounding box. In this paper, we focus on generating bounding box proposals, which is particularly useful for class-specific object detection. We propose an effective approach to improve the quality of object proposals.

Our work is motivated by the following problems. First, most models [3, 7, 24] that directly generate bounding box proposals 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 [6, 8, 15, 18, 19, 20] are commonly 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 (Sect. 3.1).
- A box refinement method, namely Multi-Thresholding Straddling Expansion (MTSE), which effectively reduces localization bias via fast diversification (Sect. 3.2).

Figure 1 illustrates the overall pipeline of our method. 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.

We evaluate our method on PASCAL VOC2007. Experiments show that our method effectively improve existing models by a large margin. In particular, when using 2000 proposals, we achieve the highest recall at intersection over union threshold of 0.5 and 0.8 with 94.2% and 63.8%, respectively. In addition, the proposed MTSE takes only 0.15s, thus bringing little computational overhead to existing models. (Sect. 4).

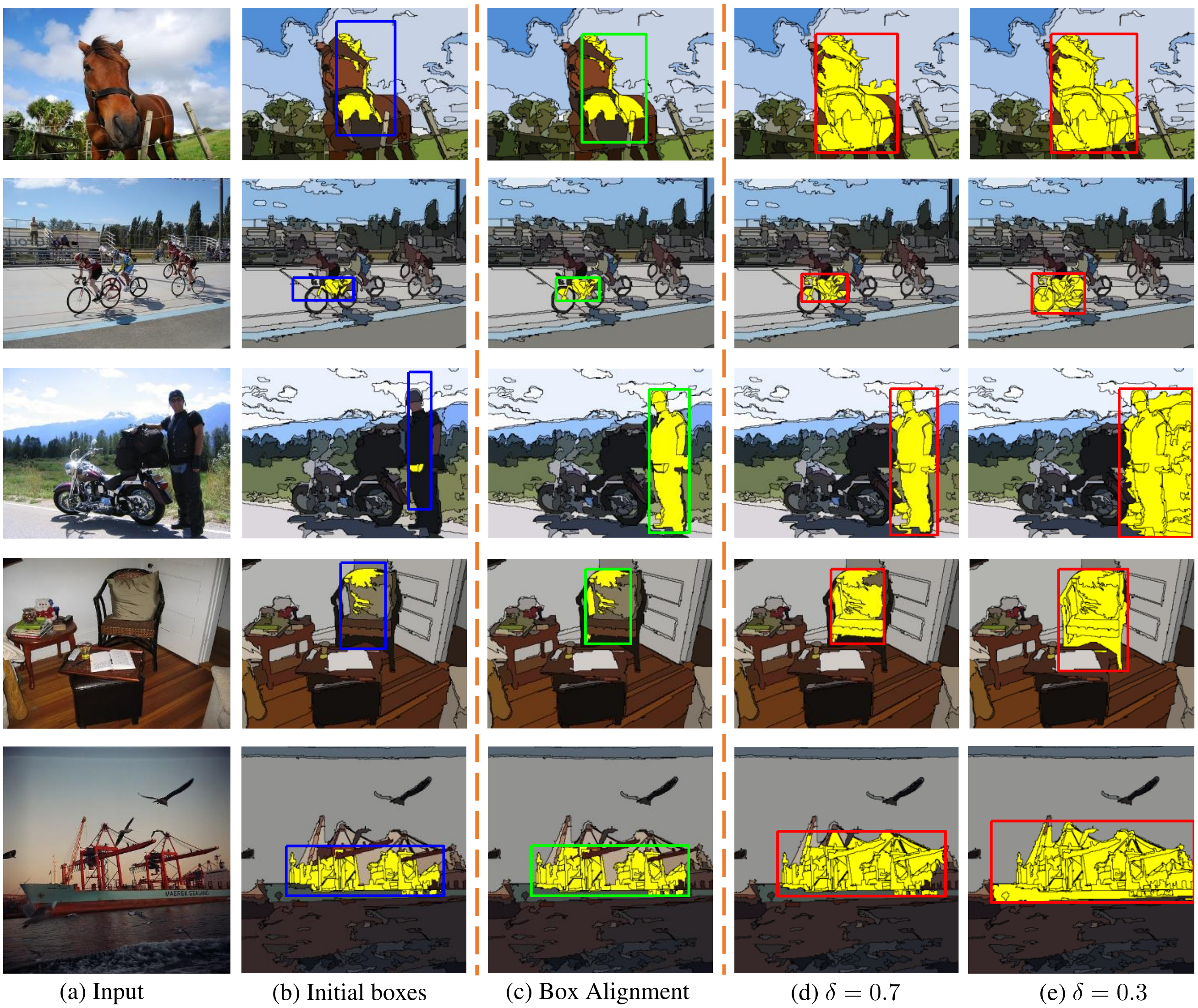


Figure 1. Illutration of our MTSE method using several examples. (a) Input images. (b) Initial bounding boxes. (c) Boxes after alignment. (d-e) Proposals after straddling expansion by setting the threshold δ to 0.7 and 0.3, respectively. Superpixels wholly enclosed by a bounding box are indicated in yellow. Best viewed in color.

2. Related Works

According to the pipeline for generating object proposal, most object proposal generators can be classified into two categories: *objectness-based* and *similarity-based*.

Objectness-based models [2, 3, 7, 24] focus on the designing of objectness measurement. Such methods try to directly distinguish objects from amorphous background stuff. To this end, a common pipeline is to first initialize a pool of candidate bounding boxes, then sort them with an objectness ranking model and output the top few proposals. To estimate the objectness score, various cues have been exploited, such as saliency, color contrast, edge density, superpixels straddling, location and size [3], binarized normed gradients [7], and edge maps [24]. Although these approaches are efficient in computation, we observe two drawbacks. The first is the lost in localization accuracy caused by the discrete sampling of initial bounding boxes. Thus most objectness-based methods have low recall at high intersection over union threshold. In this paper we will show that our MTSE method can significantly improve recall at high IoU thresholds for these methods. The second is the limited discriminant ability of exploited properties of generic objects. This can easily result in localization bias. For example, the recently proposed fast BING feature [7] has achieved high recall at low IoU threshold, but its localization is quite cursory. We owe this defect to the weak discriminativeness of simple gradient feature as it is inadequate to capture the essential properties of generic objects [23]. However, our method can lessen this bias by diversifying the proposals via multiple expansions. The localization accuracy can be significantly improved by our MTSE method, which will be shown in the experiments (Sect. 4).

Similarity-based models [4, 5, 6, 8, 15, 17, 18, 19, 21, 20] address the problem by merging similar regions based on diverse cues. Instead of directly designing/learning invariant features from generic objects, which is a very tough problem, these approaches are relatively easier and also effective. To this end, a common pipeline is to first initialize a set of seed regions (typically using superpixels [1, 11]), then merge similar regions to generate segment proposals. For regions similarity measurement, diverse and complementary cues including color, texture, location and size, are usually considered. For regions merging, Selective Search (SS) [20] performs a hierarchical grouping algorithm; Randomized Prim (RP) [18] generates random partial spanning trees using superpixel connectivity graph; Multiscale Combinatorial Grouping (MCG) [4] utilizes hierarchical segmentations and groups multiscale regions into proposals by exploring combinatorial space; Rantalankila et al. [19] propose a method combining locally superpixels merging and global graph cut to generate proposals; Geodesic Object Proposals (GOP) [17] computes a signed geodesic distance transform for each foreground-background mask and identifies certain critical level sets as object proposals. In addition, some other models [6, 8, 15] extract object proposals by solving a set of figure-ground segmentation problems. A common approach to improving proposal quality is to utilize multiple segmentations in different scales and colorspaces, which, however, requires more computations compared with objectness-based methods. In our work, we introduce multiple straddling expansions as opposed to multiple segmentations to diversify object proposals. Our method only requires one segmentation and the expansion algorithm is very fast, thus saving computational cost.

Our approach utilizes similar superpixels straddling feature as in [3], but we use it to guide box refinement instead of objectness measurement. In the work of [3], tightness is introduced to score bounding boxes. Boxes are likely to contain an object if they tightly enclose a set of superpixels. Similar idea was exploited in [24] which operates on edges. However, such measurement can easily result in localization bias. We prefer multiple degrees of superpixel tightness instead. We will show that diverse superpixel tightness can be achieved via multi-thresholding superpixel straddling expansion.

Box refinement is seldom explored in prior objectness-based methods [3, 7]. However, for these methods, bounding boxes are usually initialized using regular sampling for which it is hard to cover object precisely. Therefore, box refinement is indispensable to obtain accurate localization. Edge Boxes [24] has proposed to refine top-ranked bounding boxes using a greedy iterative local search after initial scoring. Such refinement is performed in a fixed searching step, whereas our approach utilizes superpixels to guide box refinement and requires no scoring function. In fact, we will show that our box refinement method can further improve Edge Boxes.

3. Methodology

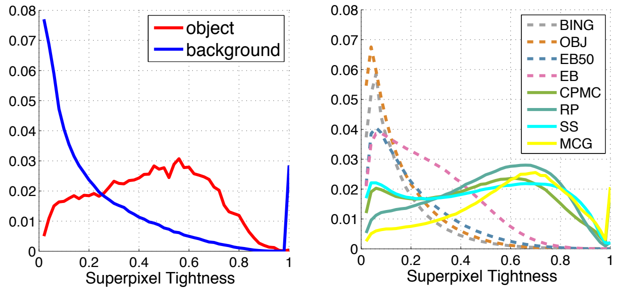
3.1. Superpixel Tightness: a Localization Bias Indicator

We observe that most objectness-based methods can hardly achieve stable recall for a wide range of intersection over union thresholds (*i.e.* strong localization bias). By contrast, similarity-based methods have better balance between recall and localization accuracy, as shown in Figure 5. To understand such bias, we introduce an indicator, *superpixel tightness (ST)*, which measures how tight a bounding box fits around an object. Given a set of superpixels $\mathcal{S}_\theta$ obtained using [11] with segmentation parameter θ, we define ST of a bounding box b as the proportion of the area of superpixels wholly enclosed by box b to the area of b. Formally,

$$ST(b) = \sum_{s \in \mathcal{S}_\theta} \frac{|s| \cdot \delta(|s| - |s \cap b|)}{|b|}, \tag{1}$$

where $\delta(x)$ is the Dirac delta function which is zero everywhere except at $x = 0$. For each superpixel s, we first compute the area $|s \cap b|$ of its intersection with box b, then sum up the number of pixels contained in the superpixels entirely inside b. $ST(b)$ is 0 if none of the superpixels is wholly enclosed by b. Such ST measure is similar to the superpixels straddling cue introduce by Alexe *et al.* [3]. The difference is that our ST measure doesn't consider superpixels straddling the box. Also, we utilize superpixels instead of contours as in Edge boxes [24] because superpixels provide a useful guidance in subsequent expansion process, to be introduced in next section.

The superpixel tightness measure is able to indicate the localization bias of object proposal generators. For demonstration, we first plot the distributions of superpixel tightness for ground truth objects and background regions respectively on PASCAL VOC2007 dataset in Figure 2(a). Background regions are randomly sampled in sliding window manner and have intersection over union overlap with ground truth objects less than 0.5. The figure clearly shows that objects possess diverse degrees of superpixel tightness while most background stuff incline to low tightness. Based on this observation, a good object proposal generator is supposed to produce bounding boxes with distribution of superpixel tightness similar to that of ground truth objects. However, currently most objectness-based methods fail to make it because of inadequate objectness hypotheses. To



(a) (b)

Figure 2. Distributions of superpixel tightness for (a) ground truth objects and background regions on PASCAL VOC2007, and (b) 1000 object proposals generated by several objectness-based models (in dashed lines) and similarity-based models (in solid lines). The values at $ST = 0$, which imply the proportion of bounding boxes that contain no superpixels entirely, are ignored in the figures for clarity. Best viewed in color.

show this, we plot the ST distributions for proposals generated by recent state-of-the-art methods in Figure 2(b). For objectness-based methods, we test OBJ [3], BING [7], EB50 (Edge Boxes 50) and EB (Edge Boxes 70) [24]; for similarity-based methods, we test CPMC [6], RP [18], SS [20] and MCG [4]. We found that all objectness-based methods have strong bias to low tightness while most similarity-based methods spread more evenly across various tightness. This accords with their bias in localization accuracy. In other words, we can use ST distribution to measure the bias in localization accuracy.

The characteristics of ST distribution imply a direction to improving the quality of object proposals. We should keep in mind that high-quality object proposals should also have low bias (high diversity) in ST distribution. To this end, a straightforward approach is to refine object proposals to obtain higher diversity.

3.2. Multi-Thresholding Straddling Expansion

To generate bounding box proposals with diverse degrees of superpixel tightness, we introduce a box refinement method using superpixels straddling. We utilize superpixel to guide box refinement because its key property is preserving object boundaries. We first define the *straddling degree* of a superpixel s with regard to a bounding box b as

$$SD(s,b) = \frac{|s \cap b|}{|s|}. \qquad (2)$$

It indicates the proportion of the superpixel's area $|s \cap b|$ inside b to the superpixel's area $|s|$. 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:

$$\mathcal{S}_\delta(b) = \mathcal{S}_{in}(b) \cup \{s \in \mathcal{S}_\theta | SD(s,b) \geq \delta\}, \qquad (3)$$

where $\mathcal{S}_{in}(b) = \{s \in \mathcal{S}_\theta | SD(s,b) = 1\}$ is the set of superpixels entirly inside b. A new box $\hat{b}$ is obtained by computing the minimum box enclosing $\mathcal{S}_\delta(b)$. Box b is possibly enlarged after the refinement. Figure 1 shows some examples of straddling expansion by setting δ to different values. Large value of δ produces a minor variant of b, thus possibly leading to a more precise location if b has a coarse overlap with an object. Small value of δ produces a distinct box, which can increase the possibility of jumping out of a "local minima" for inaccurate box.

Straddling expansion is able to reduce the bias of object proposals by diversifying superpixel tightness. We plot the ST distributions after applying straddling expansion to the bounding box proposals generated by three baseline models: BING [7], OBJ [3] and MCG [4], in Figure 3. It clearly shows that the ST distributions for small δ's are more distinct from the baseline's distribution than those for large δ's. For example, the ST distributions for BING and OBJ have major proportions in low values. After applying straddling expansion, we obtain larger proportions in high values of superpixel tightness. Similar results are observed for MCG, which is a similarity-based method. As MCG itself generates high-quality object proposals and has little bias in localization, straddling expansion on its proposals yields less distinct ST distribution.

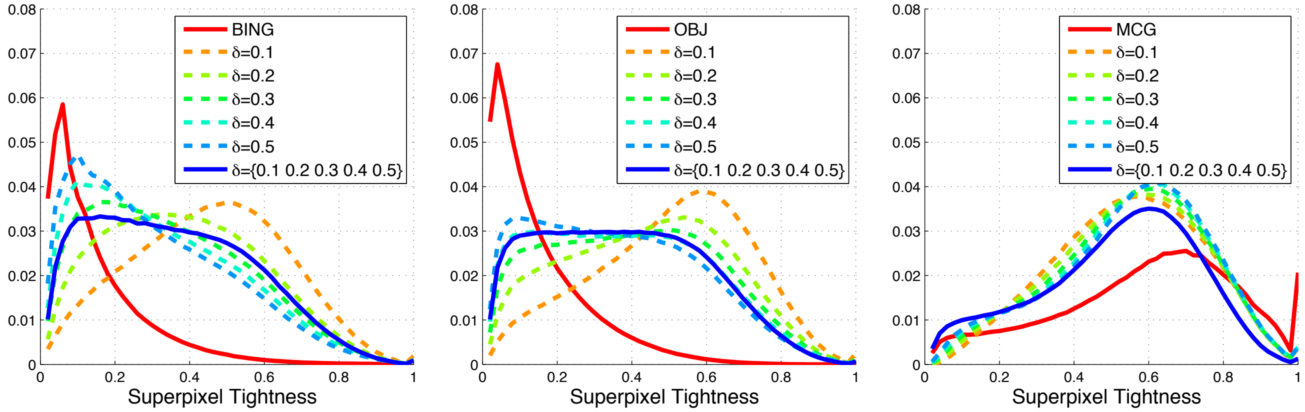
Instead of choosing a single value of δ, we use multiple δ's to perform straddling expansion, which we call multi-thresholding straddling expansion (MTSE). As shown in Figure 3, combining multiple thresholds can further obtain higher diversity. By this means, multiple bounding boxes are generated for each initial bounding box.

The unique benefit of our box refinement method is that it can naturally generate bounding boxes aligning with object boundaries preserved by superpixels. This property differentiate our method from Edge Boxes [24] which performs a fixed-step local search. Moreover, unlike similarity-based methods, straddling expansion doesn't require extracting low-level features (*e.g.* color, texture) to measure regions similarity. Only straddling degrees are computed for superpixels, thus the expansion process is very efficient.

3.3. Box Alignment

As a set of initial bounding boxes is required for MTSE, naturally we can directly feed the bounding boxes generated by an existing model into MTSE. However, the bounding boxes initialized in such way don't always align with the object boundaries. For example, the BING proposal [7] is essentially a subset of the sliding windows, which are uniformly distributed with fixed sizes and aspect ratios, thus such bounding boxes possibly have bad alignment with object boundaries. Therefore, we propose to align bounding boxes before straddling expansion.

Since the exact object boundaries are unknown, we adopt



(a) (b) (c)

Figure 3. Distributions of superpixel tightness before and after applying straddling expansion for three baseline models: (a) BING [7], (b) OBJ [3], and (c) MCG [4]. The threshold δ is set to five values individually. A combination of multiple δ's achieves the least bias in the ST distribution. Best viewed in color.

an approximation by aligning bounding boxes with potential object boundaries preserved by superpixels. We say a bounding box b aligns with superpixels S_θ if it is the minimum box enclosing a subset of the superpixels. Given an initial bounding box b, we first compute its inner set and straddling set, which are defined as

$$\begin{aligned} \mathcal{S}_{in} &= \{s \in \mathcal{S}_\theta | SD(s,b) = 1\}, \\ \mathcal{S}_{st} &= \{s \in \mathcal{S}_\theta | 0 < SD(s,b) < 1\}. \end{aligned} \tag{4}$$

Let $b(\mathcal{S})$ denote the minimum box enclosing the set of superpixels $\mathcal{S}$, and $O(b_i, b_j)$ denote the intersection over union overlap between b_i and b_j. Then we sort the straddling set $\mathcal{S}_{st}$ according to the intersection over union overlaps, so that its elements $\{s_1, ...s_K\}$ satisfy

$$O(b(\mathcal{S}_{in} \cup \{s_i\}), b) \geq O(b(\mathcal{S}_{in} \cup \{s_j\}), b), \forall i < j. \tag{5}$$

The box alignment process is to expand the bounding box from $b(S_{in})$, which is the one enclosing the inner set, to the one which is the closest to the given box b, by greedily adding in superpixels from the sorted straddling set. By this means, we obtain an aligned bounding box which has the highest overlap with the given bounding box. The specific procedure is summarized in pseudo-code in Algorithm 1.

Some examples are shown in the third column of Figure 1. Box alignment has the capability of "dragging" some coarse bounding boxes back to the main part of an object (*e.g*. Row 1-3 in Figure 1). In some cases (*e.g*. Row 3 in Figure 1), even the box alignment procedure is already sufficient to generate an accurate bounding box.

3.4. Implementation

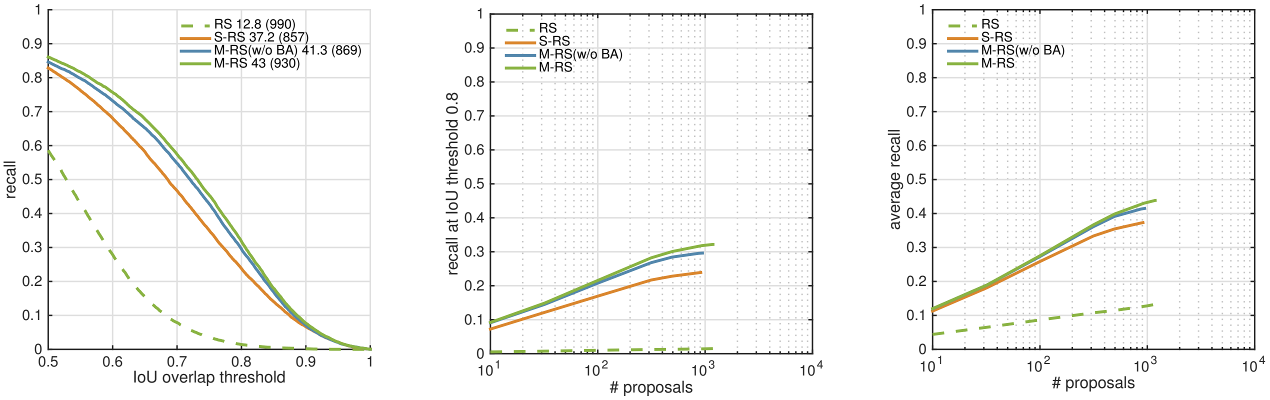
We compute a superpixel segmentation using [11] in Lab colorspace at a single scale. Straddling expansion is performed five times by setting the expansion threshold δ to $0.1 \times i, i = 1, 2, ..., 5$, which are determined based on the distribution of superpixel tightness.

Algorithm 1 Box Alignment

```
Input: initial box $b$, superpixels $\mathcal{S}_\theta$
Output: aligned box $b^\star$
compute inner set: $\mathcal{S} \leftarrow \mathcal{S}_{in}$
obtain sorted straddling set: $\{s_1, ...s_K\}$
$k \leftarrow 1$
$o \leftarrow O(b(\mathcal{S}), b)$
$\hat{o} \leftarrow O(b(\mathcal{S} \cup \{s_k\}), b)$
while $\hat{o} \geq o$ do
    $o \leftarrow \hat{o}$
    $\mathcal{S} \leftarrow \mathcal{S} \cup \{s_k\}$
    $k \leftarrow k + 1$
    $\hat{o} \leftarrow O(b(\mathcal{S} \cup \{s_k\}), b)$
end while
$b^\star \leftarrow b(\mathcal{S})$
```

As MTSE generates five sets of bounding boxes, to reduce redundancy, we rank each set by adding some randomness similar to [20]. Specifically, let $\hat{b}_i$ be the bounding box derived from the initial bounding box b_i using a certain value of δ, we score $\hat{b}_i$ with value $i \times R$, where R is a random number in range [0, 1]. A ranked list of bounding boxes is obtained by sorting all the bounding boxes in ascending order.

After ranking, we perform non-maximal suppression (NMS) to obtain the final proposals. We found that by setting the IoU threshold of NMS to 0.8 for objectness-based models, and 0.9 for similarity-based models, respectively, we can obtain high accuracy with a moderate budget of object proposals.



1000 proposals per image

80%-recall

Average recall

Figure 4. A Comparison of various variants of MTSE using regular sampling (RS) for box initialization. For the recall-overlap curves, numbers next to labels indicate average recall and average number of proposals per image. All variants achieve significant improvement over RS. The full MTSE (M-RS) performs slightly better than the single-thresholding approach (S-RS, by setting $\delta = 0.3$) and the model without box alignment (M-RS(w/o BA)). Best viewed in color.

4. Experiments

We evaluate our method on the PASCAL VOC2007 dataset [9]. The dataset contains 9,963 images from 20 categories with bounding box annotation for each object. All experimental results are reported on the test set, which consists of 4,952 images and 14,976 object instances.

Following [13, 14], we evaluate the quality of object proposals using the recall metric. Recall is computed as the fraction of ground truth objects covered above an IoU threshold. Typically, we use α-recall to denote recall at IoU threshold α. We use the recall-proposal curve to depict accuracy at different proposal budgets and recall-overlap curve to show the variation of recall over different localization precision. In addition, to measure the overall accuracy of proposals, we compute average recall (AR), which is the area under "recall-overlap" curve in IoU range of 0.5 to 1.0. AR is a comprehensive metric as Hosang et al. [13, 14] has shown that AR is highly correlated with the performance of class-specific object detectors.

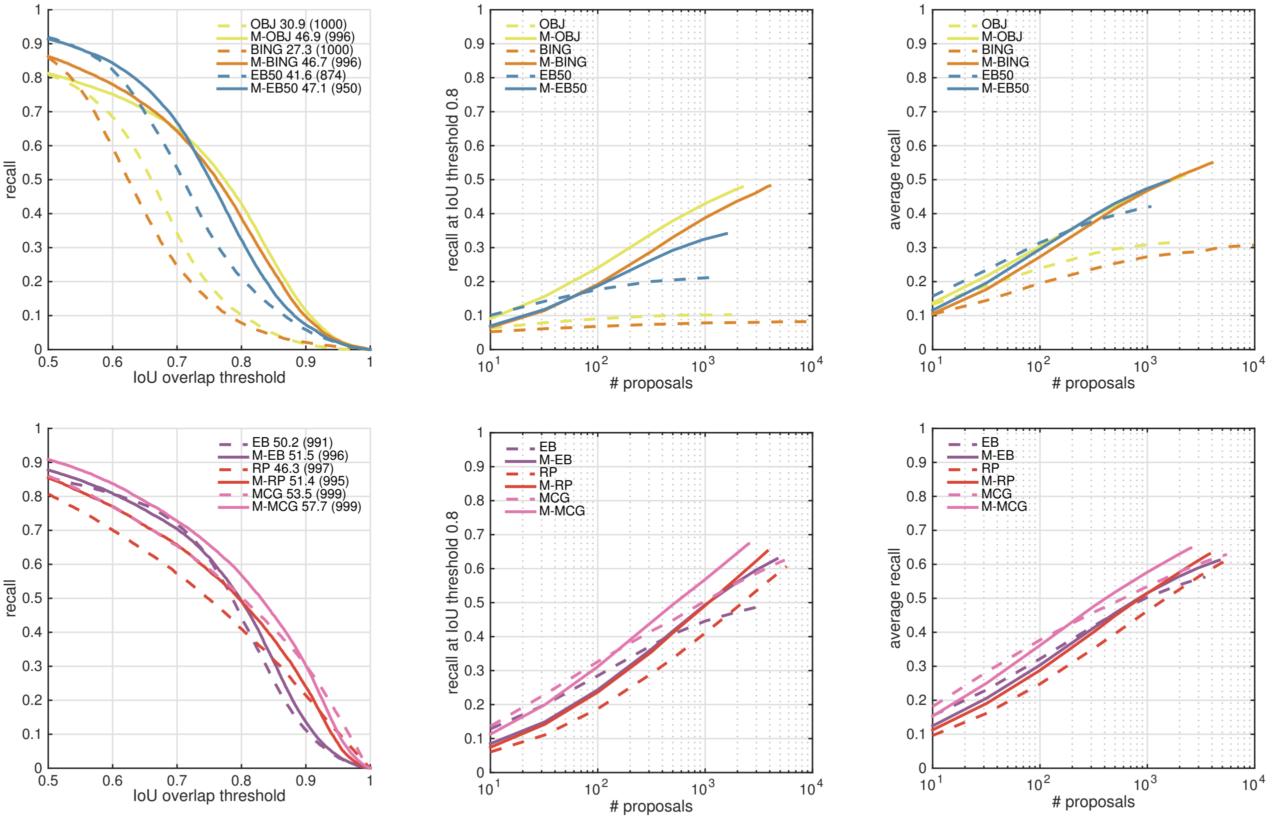
4.1. Validation of the Proposed Approach

We first verify the effectiveness of the two components of MTSE: box alignment and multi-thresholding expansion. We simply use regular sampling (RS), which is also used by [7, 24], to initialize a set of bounding boxes. Let M-RS denote the corresponding model integrated MTSE. As shown in Figure 4, single-thresholding straddling expansion (S-RS) already achieves a good performance in all metrics. By combining box alignment and extending to multi-thresholding approach, the accuracy is further improved.

As our method can be integrated into any object proposal generation model, we verify its effectiveness on numerous models. We test on objectness-based models including OBJ [3], BING [7], EB50 (Edge Boxes 50), EB (Edge Boxes 70) [24], and similarity-based models including RP [18] and MCG [4]. Correspondingly, we name their MTSE versions M-OBJ, M-BING, M-EB50, M-EB, M-RP, M-MCG, respectively. Figure 5 reports their performances using recall-overlap curve (using 1000 proposals), recall-proposal curve (at IoU threshold of 0.8) and AR-proposal curve.

Figure 5 clearly shows that MTSE successfully improves objectness-based models to a similar performance level with both high recall and accurate localization. In fact, for OBJ [3], BING [7], and EB50 [24], which are typically tuned for low overlap, our MTSE improves their recall at high overlap significantly while preserving high recall at low overlap. In particular, BING obtains the maximum boost in AR using 1000 proposals (from 0.273 to 0.467). Note that in our preliminary experiments, we have tried utilizing multiple colorspaces and denser sliding windows for BING but still obtained very low recall at high overlap threshold. Therefore our MTSE offers both higher accuracy and better diversity. For EB [24] which is tuned for IoU of 0.7, MTSE also obtains higher AR with little drop at IoU of 0.7.

Although most similarity-based models generate object proposals with less bias, MTSE further improves their performances. RP [18] is an expansion model based on superpixels similarity, which has achieved a good balance between accuracy and efficiency, whereas we further lead it to higher accuracy across almost all IoU thresholds. Similar improvement is achieved for MCG, which is a state-of-the-art model as reported in [13].



1000 proposals per image

80%-recall

Average recall

Figure 5. 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 and average number of proposals per image. Best viewed in color.

4.2. Overall Comparison

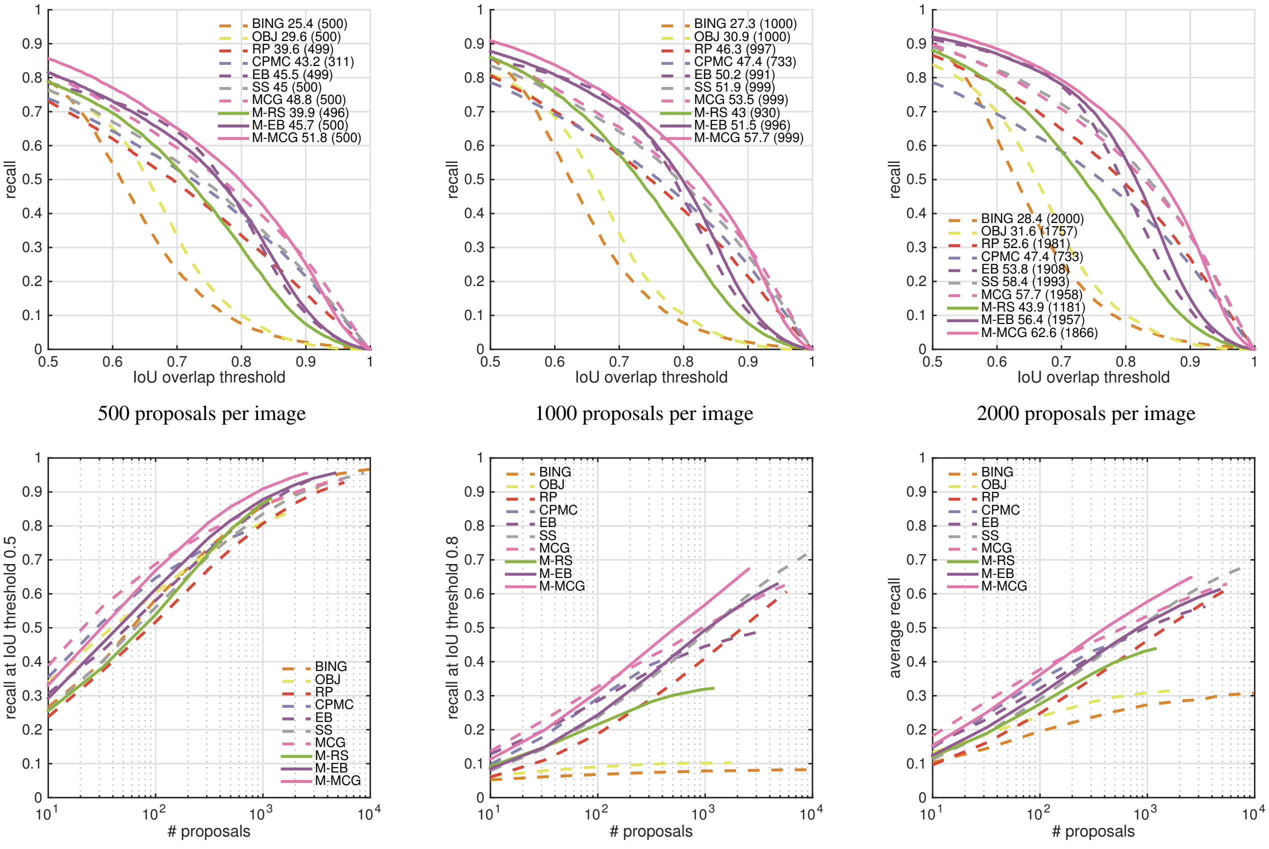
We extensively compare the improved models with more state-of-the-art methods, including CPMC [6] and SS [20], in Figure 6 and Table 1. Considering accuracy and efficiency, we recommend three variants of MTSE integrated models: M-RS, M-EB and M-MCG.

Accuracy. Figure 6 (top) presents recall-overlap curves for three proposal budgets: 500, 1000 and 2000, which are moderate for most object detection models and fair for methods that generate distinct number of proposals. Results show that M-MCG achieves the highest recall across almost the whole range of IoU thresholds (except threshold above 0.9). In particular, M-MCG even beats EB at $\alpha = 0.7$, for which EB is specially tuned. We also consider the 50%-recall, 80%-recall and average recall when varying the number of proposals. The statistics are summarized in Table 1. M-MCG stands out in all metrics followed by MCG, M-EB and SS. 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. The only flaw of M-MCG is that it has a slightly lower recall than MCG when using less than 100 proposals, because we use a very simple randomized approach to rank proposals.

Speed. Runtimes for all methods are presented in Table 1. Timings of our methods are evaluated on a 3.5 GHz i7 CPU. The runtime for MTSE is 0.15s, including 0.04s for colorspace conversion, 0.1s for superpixel segmentation, and 0.01s for proposal generation. Thus our MTSE brings little computational overhead to any existing model. M-RS is the second fastest model while being much more accurate than the fastest BING at high IoU threshold. For applications desiring both accuracy and speed, M-EB is also a very suitable alternative as it has comparable accuracy with M-MCG at IoU threshold less than 0.8 while requiring 0.45s only.

5. Conclusions

We propose a simple and effective approach to improve the quality of object proposals. The characteristics of superpixel tightness distribution shed light on ways to improving object proposals. The proposed multi-thresholding straddling expansion takes advantage of boundary-preserving su-



500 proposals per image

1000 proposals per image

2000 proposals per image

50%-recall

80%-recall

Average recall

Figure 6. An overall comparison of three MTSE integrated models (in solid lines) with numerous state-of-the-art models (in dashed lines). For the top row, numbers next to labels indicate average recall and average number of proposals per image. Best viewed in color.

Method	#prop = 500			#prop = 1000			#prop = 2000			Time(sec.)
	AR	50%-recall	80%-recall	AR	50%-recall	80%-recall	AR	50%-recall	80%-recall	
BING [7]	0.254	0.790	0.076	0.273	0.860	0.079	0.284	0.897	0.079	**0.06**
OBJ [3]	0.296	0.762	0.101	0.309	0.810	0.102	0.316	0.839	0.103	3
RP [18]	0.396	0.731	0.336	0.463	0.807	0.410	0.526	0.867	0.486	1
CPMC [6]	0.474	0.786	0.441	0.474	0.786	0.441	0.474	0.786	0.441	250
EB [24]	0.455	0.785	0.407	0.502	0.856	0.445	0.538	0.913	0.471	0.3
SS [20]	0.450	0.766	0.403	0.519	0.835	0.485	0.584	0.890	0.571	10
MCG [4]	0.488	0.816	0.448	0.535	0.861	0.505	0.577	0.898	0.557	30
M-RS	0.399	0.788	0.301	0.430	0.861	0.318	0.439	0.881	0.322	0.15
M-EB	0.457	0.816	0.416	0.515	0.878	0.493	0.564	0.920	0.559	0.45
M-MCG	**0.518**	**0.856**	**0.494**	**0.577**	**0.909**	**0.568**	**0.626**	**0.942**	**0.638**	30.2

Table 1. Results of MTSE integrated models compared to results of numerous state-of-the-art models for three budgets of object proposals: 500, 1000, 2000. Runtimes are taken from [13]. M-MCG achieves the highest accuracy in all metrics. BING takes 0.06s on our platform without using multithreading and image preloading, the same evaluation condition with our methods. Hosang *et al*. [13] reports a slower 0.2s runtime probably due to some unoptimized disk I/O, according to our communication with J. Hosang.

perpixel, generating object proposals with both high diversity and accurate localization. By integrating our method into existing models, we achieve state-of-the-art results in all metrics. Learning adaptive threshold for bounding box refinement is included in our future works, which could further increase the accuracy with fewer proposals.

Acknowledgements This work was supported by the National Natural Science Foundation of China (No. 60502013 and No. 61171113)

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE TPAMI*, 34(11):2274–2282, Nov. 2012.

[2] B. Alexe, T. Deselaers, and V. Ferrari. What is an object? In *IEEE CVPR*, pages 73–80. IEEE, 2010.

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

[4] P. Arbelaez, J. Pont-Tuset, J. Barron, F. Marqus, and J. Malik. Multiscale combinatorial grouping. In *IEEE CVPR*, 2014.

[5] J. Carreira and C. Sminchisescu. Constrained parametric min-cuts for automatic object segmentation. In *IEEE CVPR*, pages 3241–3248, 2010.

[6] J. Carreira and C. Sminchisescu. Cpmc: Automatic object segmentation using constrained parametric min-cuts. *IEEE TPAMI*, 34(7):1312–1328, July 2012.

[7] M.-M. Cheng, Z. Zhang, W.-Y. Lin, and P. H. S. Torr. BING: Binarized normed gradients for objectness estimation at 300fps. In *IEEE CVPR*, 2014.

[8] I. Endres and D. Hoiem. Category-independent object proposals with diverse ranking. *IEEE TPAMI*, 36(2):222–234, Feb 2014.

[9] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html.

[10] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE TPAMI*, 32(9):1627–1645, Sept 2010.

[11] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *IJCV*, 59(2):167–181, 2004.

[12] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *IEEE CVPR*, 2014.

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

[14] J. Hosang, R. Benenson, and B. Schiele. How good are detection proposals, really? In *BMVC*, 2014.

[15] A. Humayun, F. Li, and J. M. Rehg. Rigor: Recycling inference in graph cuts for generating object regions. In *IEEE CVPR*, 2014.

[16] H. Kang, M. Hebert, A. Efros, and T. Kanade. Data-driven objectness. *IEEE TPAMI*, PP(99):1–1, 2014.

[17] P. Krähenbühl and V. Koltun. Geodesic object proposals. In *ECCV*, pages 725–739. 2014.

[18] S. Manen, M. Guillaumin, and L. Van Gool. Prime object proposals with randomized prim's algorithm. In *IEEE ICCV*, pages 2536–2543, Dec 2013.

[19] P. Rantalankila, J. Kannala, and E. Rahtu. Generating object segmentation proposals using global and local search. In *IEEE CVPR*, 2014.

[20] J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders. Selective search for object recognition. *IJCV*, 104(2):154–171, 2013.

[21] K. E. Van de Sande, J. R. Uijlings, T. Gevers, and A. W. Smeulders. Segmentation as selective search for object recognition. In *IEEE ICCV*, pages 1879–1886. IEEE, 2011.

[22] X. Wang, M. Yang, S. Zhu, and Y. Lin. Regionlets for generic object detection. In *IEEE ICCV*, pages 17 – 24. 2013.

[23] Q. Zhao, Z. Liu, and B. Yin. Cracking bing and beyond. In *BMVC*, 2014.

[24] C. L. Zitnick and P. Dollár. Edge boxes: Locating object proposals from edges. In *ECCV*, 2014.