2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops

Texture Complexity based Redundant Regions Ranking for Object Proposal

Wei Ke*[1,2], Tianliang Zhang.[1], Jie Chen[2], Fang Wan[1], Qixiang Ye†[1] and Zhenjun Han[1]

[1]University of Chinese Academy of Sciences
[2]CMV, University of Oulu, Finland
{kewei11, zhangtianliang13, wanfang13}@mails.ucas.ac.cn, jiechen@ee.oulu.fi,{qxye, hanzhj}@ucas.ac.cn

Abstract

Object proposal has been successfully applied in recent visual object detection approaches and shown improved computational efficiency. The purpose of object proposal is to use as few as regions to cover as many as objects. In this paper, we propose a strategy named Texture Complexity based Redundant Regions Ranking (TCR) for object proposal. Our approach first produces rich but redundant regions using a color segmentation approach, i.e. Selective Search. It then uses Texture Complexity (TC) based on complete contour number and Local Binary Pattern (LBP) entropy to measure the objectness score of each region. By ranking based on the TC, it is expected that as many as true object regions are preserved, while the number of the regions is significantly reduced. Experimental results on the PASCAL VOC 2007 dataset show that the proposed TCR significantly improves the baseline approach by increasing AUC (area under recall curve) from 0.39 to 0.48. It also outperforms the state-of-the-art with AUC and uses fewer detection proposals to achieve comparable recall rates.

1. Introduction

For visual object detection, object localization is the first and most important step. Conventional approaches that use a sliding-window strategy [6, 7, 11] to localize objects, tending to generate millions of candidate windows. The classification of so many windows in the following detection step is computationally expensive, in particular, when complex features and/or classification methods are used. Recently, an alternative way, i.e. object proposal, has been investigated to improve the efficiency of object localization. Object proposal tends to produce much fewer (2 magnitude fewer) object candidate windows than the sliding-window strategy, which, with no doubts, will contribute to improve computation efficiency, as well as preserving the object detection rate.

Over the recent detection proposal approaches, Hosang et al. [13] carry out a performance comparison with respect to their repeatability and recall. Selective Search [21, 22] and EdgeBoxes [24]are regarded as two state-of-the-art approaches with high recall rates and efficiency. On one hand, Selective Search is a typical superpixel merging method, which generates criterion homogeneous regions with high localization accuracy. It tends to produce tens of thousands of windows and leads to serious redundancy problem. To alleviate such a problem, a pseudo-random algorithm is often utilized to select the final proposals since there are no objectness for each region to determine the confidence to be an object. On the other hand, EdgeBoxes, a typical objectness method, takes the hypothesis that objects usually possess more complete contours, and a sliding window strategy is used to localize regions with complete contours as object proposals. The using of complete contours helps it reduce the redundancy of proposals, however, the sparse sliding windows results in loss of aspect-ratio and localization accuracy. Taking the complementarity of these two kinds of methods inspires us to integrate them together to generate object proposal with high accuracy and confidence.

We propose a strategy named Texture Complexity based Redundant Regions Ranking (TCR) for object proposal. Our approach first produces redundant regions using Selective Search. It then calculates a Texture Complexity (TC) score for each region using complete contour number and Local Binary Pattern (LBP) entropy. As the redundant regions are generated mainly on the color cue, we propose using texture complexity cue as complementary. By TC score based ranking, the recall of the region is preserved, and the number of proposal regions is reduced significantly. The novelty of our approach are summarized as follows:

- Proposing a robust objectness measurement named TC score, which incorporates complete contour number

*This work was supported in part by the CSC, china.
†Corresponding author

978-1-5090-1437-8/16 $31.00 © 2016 IEEE
DOI 10.1109/CVPRW.2016.139

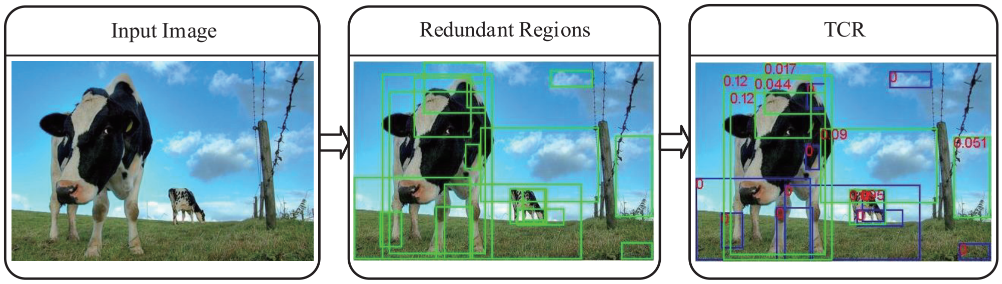


Figure 1: Flowchart of the proposed TCR approach. In the second image, two boxes that best cover the objects (cows) and 15 randomly chosen boxes are shown among the results of Selective Search. Scored by TC score, the regions are shown in the third image. A number of boxes (zero-scored ones with blue rectangles) could be removed by the ranking procedure and the larger scored boxes (green rectangles) are brought forward.

and LBP entropy.

- Integrating superpixel merging and objectness methods for object proposal. We use Selective Search to generate redundant regions, and rank the regions using the TC score.

- Combining with other superpixel merging methods easily without chaning any parameters.

The remainder of this paper has been organized as follows: we describe the related works in section 2. Then the proposed TCR based approach is presented in section 3. Experimental results are given in section 4, and we finally conclude the approach in section 5.

2. Related works

Object proposal is defined as a procedure to discovery a small set of bounding boxes that can precisely cover objects as many as possible. Existing object proposal approaches are coarsely categorized into superpixle merging based and objectness based.

By solving a sequence of Constrained Parametric Min-Cuts (CPMC), Carreira and Sminchisescu [3] propose generating figure-ground segmentations to indicate objects. An image could generate up to 10,000 redundant regions, which are subsequently ranked by a trained regressor. Uijlings et al. [21, 22] propose a hierarchical strategy (Selective Search) to merge color homogeneous regions and generate object proposals. They propose using multiple low-level features and merging functions to generate redundant regions so that as many as objects are covered. Such an approach has been successfully applied in the R-CNN object detection research [12]. Similar to the strategy of Selective Search, Manen et al. [17] propose using learned weights as a function to merge superpixels. By taking advantages of both CPMC and Selective Search, Rantalankila et al. [20] propose using a merging process with a large pool of features, and generating segmentations using a CPMC-like process. One of the most recent methods, MCG [2], combines multi-scale hierarchical segmentation regions into highly-accurate object candidates. MCG achieves a high recall rate, but does not concern the importance of computation efficiency. Xiao et al. [23] propose a complexity-adaptive metric distance used for superpixel merging, which achieves improved grouping in different levels of complexity. Most of these superpixel merging based methods produce lots of bounding boxes without assigning object confidence to them.

An early work on objectness based object proposal is [1]. Alexe et al. propose using objectness as a score of how likely a detection window contains an object. The score is estimated based on a combination of multi-cues including saliency, color contrast, edge density, location and size statistics. Chen et al. [5] propose Binarized Normed Gradients (BING) using the sliding window method for object proposal, which is based on an efficient weak classifier trained using binarized normed gradients. With delicate design of binary computations, a low computation cost of BING is guaranteed, which, as reported, can reach 300 fps on a PC platform. EdgeBoxes [24] also operates in a sliding window manner of multiple scales and multiple aspect-ratios. The scores of objects are estimated by the detected complete contour. Karianakis et al. [15] combine the lower convolutional layers of Convolutional Neural Networks (CNN) with the fast boosting decision forests to produce robust object proposals. Different with the conventional sliding window methods using image pyramid varying scales, the sliding window used for object proposal need to consider the length-width ratio which varies for different kinds of objects. However, the image cannot be slid densely in these methods in order to decrease computational efficiency, and the object proposals produced by objectness based methods are often localized inaccurately.

Most recently, Chen et al. [4] focus on the object proposal localization bias and propose Multi-Thresholding

Straddling Expansion (MTSE) to reduce localization bias using superpixel tightness. The tightness is just a property of a region, so it also produces orderless object proposals. In [16], a deep score is learned by CNN and is used for update the confidence of object proposal. However, this deep learning based objectness is data-driven, which needs to be well trained when combining with superpixel merging methods.

3. Approach

To fully excavate the complementarity of superpixel merging and objectness, we propose a strategy, i.e., Texture Complexity based Redundant Regions Ranking for object proposal. The outline of TCR algorithm is shown in Figure 1. For an input image, tens of thousands of redundant regions are generated using a color segmentation approach and a hierarchical superpixel merging procedure (section 3.1). Texture Complexity (TC) score is then calculated to measure the confidence of each region to be an object (section 3.2). TC score based ranking is used to reduce the redundancy of regions (section 3.3).

3.1. Redundant regions generation

Given an image, redundant regions are generated using an image segmentation approach and a hierarchical region merging procedure, i.e., Selective Search [21, 22]. The image segmentation procedure first uses a graph-cut algorithm to initialize the image into color homogeneous blobs, i.e. superpixels. A blob merging process is then performed until the whole image becomes a single region, as shown in Figure 2. Regions are released when every two blobs are merged from the hierarchical merging structure. To get a high recall rate, it uses a variety of color spaces, different similarity measures combination and initial blob sizes. Specifically, five color spaces are used: HSV, Lab, the RG channels of normalized RGB plus intensity, the Hue channel H from HSV, and intensity. The four similarity measures are color, texture, size and area.

The regions generated by the above procedure are extremely redundant because it incorporates a diversity of merging procedures. Tens of thousands of region proposals are generated for a natural scene image of resolution. Some boxes are highly overlapped with each other, and one object could be covered many times. To the best of our knowledge, such a superpixel merging based approach is unable tomeasure the objectness confidence of the regions. In order to reduce the region number, a pseudo-random algorithm is deployed, which, without any doubt, hurts the object recall rate.

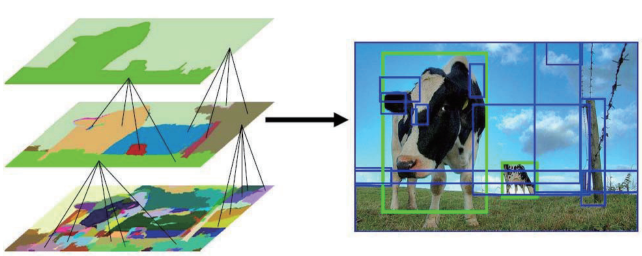


Figure 2: The hierarchical structure of the Selective Search. Blue boxes in right image are randomly selected false positives and the green boxes are true positives.

Figure 3: An edge map computed based on structured forests.

3.2. Texture complexity

Considering the redundancy of region generation with superpixle merging based approaches, we propose a new algorithm by adding the objectness measurement to rank the generated regions. As the redundant regions are produced mainly on the color cue, we utilize the texture complexity, which consists of complete contour number and LBP entropy, to supplement the color cue.

3.2.1 Complete contour number

For an edge map computed based on structured forests [8] as shown in Figure 3, a complete contour is a group of adjacent edge points with an affinity orientation. The number of complete contours that are wholly contained in a bounding box is an explicitly powerful indication of the existence of an object. Let $X = \{x_n = (m_n, o_n)\}_{n=1}^{W \times H}$ denote the edge map of an image with size $W \times H$, where m_n and o_n are edge magnitude and orientation of the pixel, respectively. Supposing the set of edge groups in an image is $S = \{s_i\}$, the set of edge groups in a bounding box b is $S_b \subset S$. The complete contour number in a bounding box is computed as

$$w_e = \frac{\sum\limits_{s_i \in S_b} f(s_i) m_i}{2(h_b + w_b)^{\kappa}}, \tag{1}$$

where m_i is the magnitude of s_i, h_b and w_b are the width and height of the bounding box. The perimeter of the bounding box serves as its normalization factor, and κ is a parameter used to offset the bias of larger windows having more contours on average. The weight function $f(s_i)$ is computed by

$$f(s_i) = \begin{cases} 1 - \max_{P} \prod_{j=1}^{|P|-1} a(\bar{o}_j, \bar{o}_{j+1}) & \text{if} s_i \in S_b \\ 0 & \text{else} \end{cases}, \quad (2)$$

where $a(\bar{o}_j, \bar{o}_{j+1})$ is the affinity of the mean orientations of contour t_j and contour t_{j+1}, P is an ordered path with length $|P|$, which is from t_1 overlapping b to s_i. Once it overlaps the bounding box b, one complete contour has little chance to be a part of the object in this bounding box. Its weight is certainly set to 0. Furthermore, those closely associated with it may have the consistent orientations even though they are completely located in b and they could contribute trivially to the object in b with smaller weights.

On the pre-computed edge map, the confidence score for each region is calculated with equation (1). The larger the score is, the more compete contours exist in the box, which tends higher confidence that the box is an object.

3.2.2 LBP entropy

We take the hypothesis that an object usually has dense texture, while the background is always textureless or diverse, as shown in Figure 5. The LBP entropy captures these differences, which verifies the hypothesis accordingly. Firstly, LBP [18] is calculated on a 3×3 patch as shown in Figure 4. The central pixel is set to the base pixel. And the neighbourhoods of a centre pixel are checked for evaluating the occurrences of equal/higher grey level values to the base. The larger one is encoded with 1, and the smaller is set to 0. An eight-bin code (256 patterns) is obtained according to an ordered connection of the binary values. Specifically, LBP entropy of a region is computed as

$$w_t = -\sum_{i=0}^{255} p_i \log p_i, \quad (3)$$

where p_i is the probability of the pattern, which can be derived by

$$p_i = \frac{N_i}{\sum N_i} = \frac{N_i}{h_b \times w_b}, \quad (4)$$

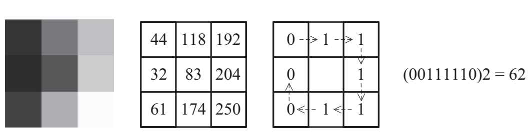


Figure 4: Local Binary Pattern.

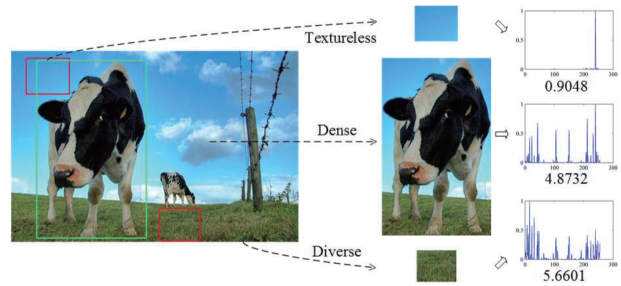


Figure 5: LBP histogram and entropy.

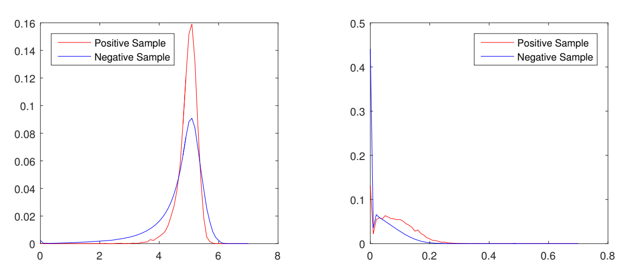
where N_i is the number of the $i-\text{th}$ pattern, h_b and w_b are the width and height of a region, respectively. As shown in Figure 5, the LBP entropy of an object region is neither very large nor very small.

Measuring objectness with LBP entropy is efficient, because the LBP feature for each region is easily obtained from the pre-computed image feature map and LBP entropy is computed without size alignment.

3.3. TCR

The distributions of complete contour number and LBP entropy on PASCAL VOC 2007 train dataset are shown in Figure 6. According to Figure 6(a), object proposal is a trade-off between the recall rate and the window number (redundancy). Following increase the threshold of complete contour number, the redundancy decreases but the recall of positive samples is hurt. In order to achieve a high recall rate, it requires to use a small threshold value. However, it means more false positives. We combine LBP entropy to complete contour number to reduce the false positives. According to Figure 6(b), the LBP entropy variance of positives is less than negatives. Therefore, we propose a gate function for LBP entropy as

$$g(w_t) = \begin{cases} 1 & \text{if } w_t \in (T_{ml}, T_{mr}) \\ 0.5 & \text{if } w_t \in (T_l, T_{ml}] \cup [T_{mr}, T_r), \\ 0 & \text{else} \end{cases} \quad (5)$$



(a) Complete contour number (b) LBP entropy

Figure 6: The distributions of positive and negative samples on the PASCAL VOC 2007 train dataset.

where T_l, T_{ml}, T_{mr}, and T_r are four thresholds. The box is brought forward if its LBP entropy is close to the peak value of the distribution. It means it is more likely to be an object. We remove the boxes with too large or too small LBP entropy value. These boxes are usually textureless or background regions. With the complete contour number in equation (1) and LBP entropy in equation (3), the combined TC score for a region is defined as

$$o = w_e \cdot g(w_r). \tag{6}$$

We rank the redundant regions using TC scores to reduce the redundancy, i.e. object proposal number.

4. Experimental results

4.1. Metrics

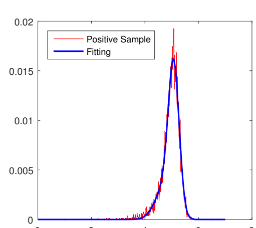
Dataset. Following [1, 5, 19, 24], we evaluate our approach on the PASCAL VOC 2007 dataset [10]. The dataset consists of train (4501 images), validation (2510 images) and test subsets (4952 images). We acquire the parameters of our TC score on the train dataset, illustrate the efficiency of incorporation approach TCR on the validation dataset, and compare our approach with the state-of-the-art on the test dataset.

Evaluation procedures. We follow the same evaluation procedure as [24], using recall, proposal number, and proposal-object overlap (Intersection over Union, IoU).

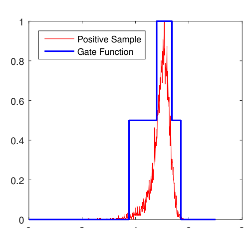
- Recall: With higher recall, the following classifier is more potential to get high detection accuracy. If one object is missed in the object proposal stage, the classifier can no longer detect the object.
- Proposal number: Less proposal number is the efficiency guarantee of the following classifier.
- IoU: Larger IoU means more accuracy localization, so that the following feature extraction methods can extract more rich information features.

Better approaches are recognized by fewer proposals and larger IoU, while keeping the recall rate. There are three commonly used experimental setups: recall vs. window number with given IoU, recall vs. IoU with given proposal number, and the minimum proposal number with given recall and IoU.

Parameters. In the contour number term of TC score, is set to be larger than 1 to reject large windows. Following the setting in [24], we set . In the LBP entropy term, we get the parameter from PASCAL VOC 2007 training set. Since the two sides of the positive samples distribution have the different slope, two-term Gaussian is used to fit the distribution, as shown in Figure 7(a). With the experimentally determined Gaussian parameters $\mu_1 = 5.080, \delta_1 = 0.285, \mu_2 = 4.728, \delta_1 = 0.486$, we set $T_L = \mu_2 - 2\delta_2, T_{M1} = \mu_1 - \delta_1, T_{M2} = \mu_1 + \delta_1, T_R = \mu_2 + 2\delta_2$. The gate function is shown is Figure 7(b).



(a) Fitting

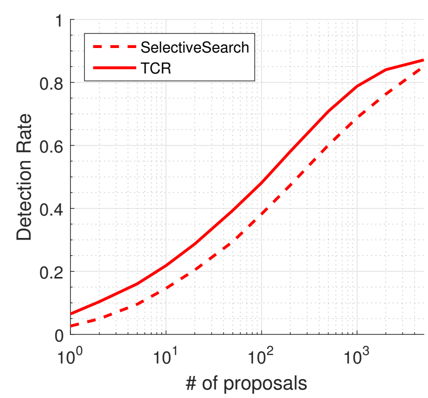


(b) Gate function

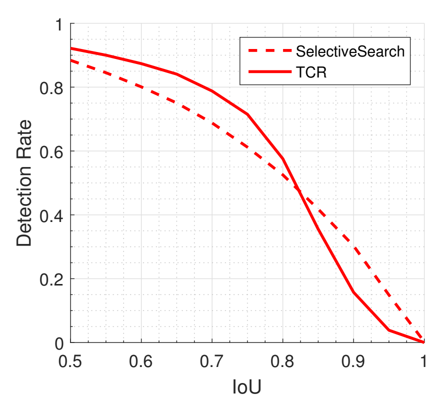
Figure 7: Compute the parameters of the LBP entropy term.

4.2. Comparison with baseline

Selective Search [21, 22]is used as the baseline of our TCR approach. Given a box, TC score is calculated using equation (6). The efficiency of TCR is evaluated on PASCAL VOC validation datasets. Figure 8 illustrates the comparison between the TCR approach and the baseline.



(a) Recall vs. number of proposals



(b) Recall vs. IoU

Figure 8: Comparison with the baseline approach.

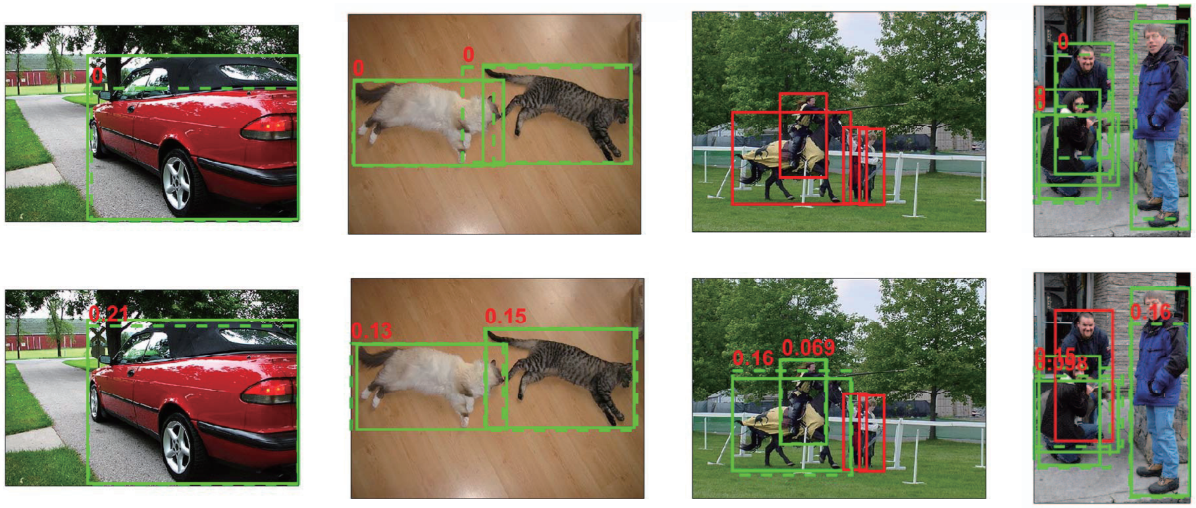


Figure 9: Object proposal examples of the PASCAL VOC 2007 validation dataset. In each image, ground-truth boxes are annotated in solid lines, for which a highly overlapping proposal exists (with the corresponding proposals shown as green dash rectangles) and red rectangles are ground-truth boxes that are missed. The first row is from Selective Search and the bottom row is from our TCR approach.

Recall vs. the number of proposal is shown in Figure 8(a) with IoU 0.7. It can be seen that TCR significantly improves the recall rate more than 10% when using 100 or 1000 proposals. Recall is improved to 0.87 while Selectives Search is 0.85. In addition, TCR needs only 720 detection proposals to reach 75% recall rate, while Selective Search needs 1777.

Recall vs. IoU is shown in Figure 8(b) when using 1000 detection proposals. In Figure 8(b), it can be seen that TCR outperforms Selective Search when IoU ranges from 0.5 to 0.81. When IoU is larger than 0.81, TCR reports a lower rate than Selective Search, which may due to the fact that TCR further employs a Non-Maximum Suppression (NMS) procedure. However, the IoU larger than 0.5 is usually good enough for the detection and classification tasks, one can conclude from Figure 7(b) that TCR outperforms the baseline with respect to recall vs. IoU.

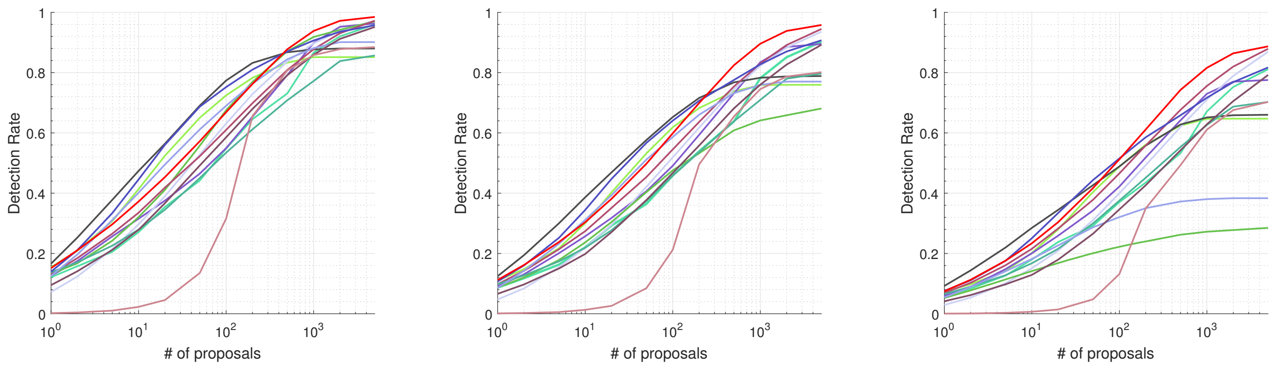
Examples of object proposals are shown in Figure 9. With IoU = 0.7, we choose 1000 detection proposals to illustrate the ground-truth matching results. It indicates that our TCR can bring forward the more accuracy location proposals from the first two columns, and the boxes which is more potential to be an object from the third column. The last column illustrates a missed object (the black-dressed man). The reason of such missing detection is that the distribution of LBP is regular, so that it is invalid as the entropy is too small.

4.3. Comparison with the State-of-the-art

We compare TCR against recent approaches including Objectness [1], MCG [2], CPMC [3], BING [5], Endres [9], Rigor [14], RandomizedPrims [17], Rahtu [19], Rantalankila [20], Selective Search [21, 22], complexity-adaptive (CA) [23] and EdgeBoxes [24]. Results of compared approaches are provided by Hosang et al. [13], and curves are generated using the Structured Edge Detection Toolbox V3.0 [24].

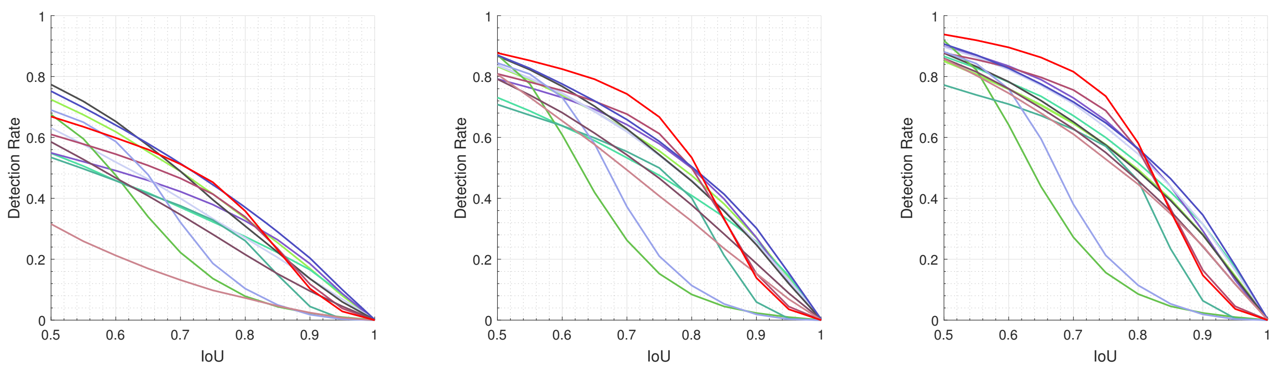
Recall versus number of proposal is illustrated in Figure 10, and we compare recent approaches using IoU thresholds of 0.5, 0.6 and 0.7. The red curves show the recall performance of TCR. It can be seen in Figure 10 that the recall of TCR approach outperforms the state-of-the-art, in particular, when IoU = 0.7. Recall versus IoU is shown in Figure 11. The number of proposals is set to 100, 500 and 1000, respectively. Varying IoU from 0.5 to 0.75, Endres, CPMC and MCG perform slightly better than TCR with 100 proposals. Nevertheless, TCR achieves the highest recall rate given 500 and 1000 proposals.

In table 1, we compare the numbers of proposals required by each approach with 25%, 50% and 75% recall rates and IoU 0.7. It can be seen that the TCR approach has the highest recall rate of 0.89 when using thousands of bounding boxes. TCR needs 535 detection proposals to achieve 75% recall rate. It's the fewest among all compared approaches. TCR uses only 12 and 91 detection proposals to achieve 25% and 50% recall rates, respectively, which small enough with all compared approaches. In table 1, it can be seen that the AUC of TCR is 0.48, which is the highest among all compared approaches. All the comparisons in table 1 confirm that TCR improves the state-of-the-art.



(a) IoU = 0.5 (b) IoU = 0.6 (c) IoU =0.7

Figure 10: Comparing results with the state-of-the-art using recall versus number of proposals. (Best viewed in color).



(a) 100 proposals per image (b) 500 proposals per image (c) 1000 proposals per image

Figure 11: Comparing result with state-of-the-art using Recall versus IoU. (Best viewed in color).

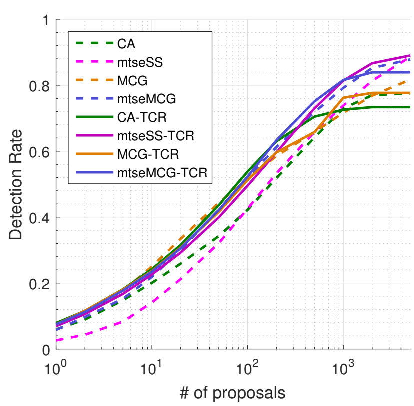
Method	AUC	N@25%	N@50%	N@75%	Recall
BING [5]	0.20	302	-	-	0.28
Rantalankila [20]	0.25	146	520	-	0.70
Objectness [1]	0.27	28	-	-	0.38
RandomizedPrims [17]	0.35	42	358	3204	0.79
Rahtu [19]	0.36	29	310	-	0.70
Rigor [14]	0.38	25	367	1961	0.81
Selective Search [22, 21]	0.39	29	210	1416	0.87
CPMC [3]	0.41	17	112	-	0.65
Endres [9]	0.44	**07**	112	-	0.66
MCG [2]	0.46	10	**86**	1562	0.82
EdgeBoxes [24]	0.47	12	96	6558	0.88
TCR (our approach)	0.48	12	91	**535**	0.89

Table 1: Comparison results of the performance using TCR with relevant methods.

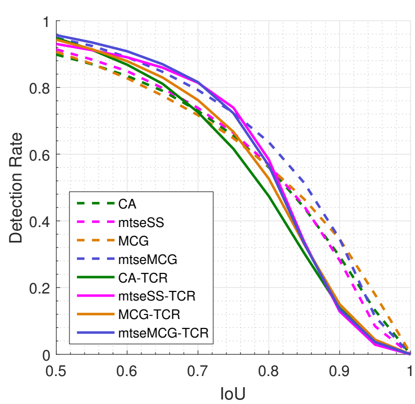
4.4. Extending to other proposal generation method

As discussed in section 4.1, we statistic the TC score directly on the ground-truth of training dataset, getting the thresholding parameters of the gate function . It means the TCR approach can be added to other region generation methods without re-training like DeepBox [16]. We evaluation TCR approach with some other recently region generation method, i.e. MCG [2], the method CA [23], mtseMCG and mtseSS [4].

The pairwise comparison results are shown in Figure 12 and table 2. Mostly, the AUC is increased when adding the TCR to either of the region generation methods. The number of proposal decreases significantly compared with to the original methods. With CA methods, the recall cannot get to 75% as there is a Nox-Maximum Suppression (NMS) operation after superpixel merging. As the same reason of CA, the NMS in our approach also gets some bad influences to the recall results for MCG, CA, mtseMCG and mtseSS. In mtseSS, the recall keeps 0.89 and gets the highest recall of these compared methods in table 2.



(a) Recall vs. number of proposals



(b) Recall vs. IoU

Figure 12: Comparison with the baseline approach.

Method	AUC	N@75%	Recall
MCG [2]	0.46	1562	0.82
MCG-TCR	0.46	922	0.78
CA [23]	0.42	1418	0.78
CA-TCR	0.46	-	0.73
mtseMCG [4]	0.47	672	0.89
mtseMCG-TCR	0.48	496	0.84
mtseSS [4]	0.41	1112	0.89
mtseSS-TCR	0.47	589	0.89

Table 2: Comparison results of the performance using TCR with relevant methods.

It's interesting to see that Selective Search is the most stable performance method, either integrated with the tightness to prevent the bias of localization, or integrated with our proposed objectness TC score to reduce the number of bounding boxes.

5. Conclusions

Object proposal methods reduce object candidate windows from millions of sliding windows to thousands of regions. Our motivation is that the proper integration of color, texture complexity cues contributes to better object proposal. To fully integrate accuracy localization of color based superpixel merging methods and the ranking confidence of the objectness based methods, we propose a strategy named Texture Complexity based Redundant Regions Ranking (TCR) to further improve the performance of object proposal. Selective Search is employed to output redundant regions, and TC score is computed to measure the confidence of a region to be an object. TC score consists of two terms, complete contour number and LBP entropy, which can be efficiently calculated with the pre-computed edge and LBP maps. Through a gate function, the two terms are fused together to reduce the redundant of regions, which essentially improves the accuracy and efficiency of overall object detection system. In addition. TCR is easy to extend to some other region generation methods and reduce the object proposal number.

Acknowledgement

This work was supported in part by the National Science Foundation of China under Grant 61271433, Grant 61202323 and Beijing Municipal Science and Technology Commission.

References

[1] B. Alexe, T. Deselaers, and V. Ferrari. Measuring the objectness of image windows. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(11):2189–2202, 2012.

[2] P. Arbeláez, J. Pont-Tuset, J. Barron, F. Marques, and J. Malik. Multiscale combinatorial grouping. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 328–335, 2014.

[3] J. Carreira and C. Sminchisescu. Cpmc: Automatic object segmentation using constrained parametric min-cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(7):1312–1328, 2012.

[4] X. Chen, H. Ma, X. Wang, and Z. Zhao. Improving object proposals with multi-thresholding straddling expansion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2587–2595, 2015.

[5] M.-M. Cheng, Z. Zhang, W.-Y. Lin, and P. Torr. Bing: Binarized normed gradients for objectness estimation at 300fps. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3286–3293, 2014.

[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893, 2005.

[7] P. Dollár, R. Appel, S. Belongie, and P. Perona. Fast feature pyramids for object detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(8):1532–1545, 2014.

[8] P. Dollár and C. Zitnick. Structured forests for fast edge detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1841–1848, 2013.

[9] I. Endres and D. Hoiem. Category-independent object proposals with diverse ranking. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(2):222–234, 2014.

[10] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136, 2015.

[11] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1627–1645, 2010.

[12] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[13] J. Hosang, R. Benenson, and B. Schiele. How good are detection proposals, really? *arXiv preprint arXiv:1406.6962*, 2014.

[14] A. Humayun, F. Li, and J. Rehg. Rigor: reusing inference in graph cuts for generating object regions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 336–343, 2014.

[15] N. Karianakis, T. J. Fuchs, and S. Soatto. Boosting convolutional features for robust object proposals. *arXiv preprint arXiv:1503.06350*, 2015.

[16] W. Kuo, B. Hariharan, and J. Malik. Deepbox: Learning objectness with convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2479–2487, 2015.

[17] S. Manen, M. Guillaumin, and L. Gool. Prime object proposals with randomized prim's algorithm. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2536–2543, 2013.

[18] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, 2002.

[19] E. Rahtu, J. Kannala, and M. Blaschko. Learning a category independent object detection cascade. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1052–1059, 2011.

[20] P. Rantalankila, J. Kannala, and E. Rahtu. Generating object segmentation proposals using global and local search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2417–2424, 2014.

[21] J. R. Uijlings, K. E. van de Sande, T. Gevers, and A. W. Smeulders. Selective search for object recognition. *International journal of computer vision*, 104(2):154–171, 2013.

[22] K. E. Van de Sande, J. R. Uijlings, T. Gevers, and A. W. Smeulders. Segmentation as selective search for object recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1879–1886, 2011.

[23] Y. Xiao, C. Lu, E. Tsougenis, Y. Lu, and C.-K. Tang. Complexity-adaptive distance metric for object proposals generation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 778–786, 2015.

[24] C. L. Zitnick and P. Dollár. Edge boxes: Locating object proposals from edges. In *Computer Vision–ECCV 2014*, pages 391–405. Springer, 2014.