Pattern Recognition 69 (2017) 299–309

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Object proposal with kernelized partial ranking

Jing Wang[a,b,*], Jie Shen[b], Ping Li[a,b]

[a] Department of Statistics and Biostatistics, Rutgers University, Piscataway, NJ 08854, USA
[b] Department of Computer Science, Rutgers University, Piscataway, NJ 08854, USA

ARTICLE INFO

Article history:
Received 14 August 2016
Revised 29 January 2017
Accepted 19 March 2017
Available online 25 April 2017

Keywords:
Object proposal
Partial ranking
Consistent weighted sampling

ABSTRACT

Object proposals are an ensemble of bounding boxes with high potential to contain objects. In order to determine a small set of proposals with a high recall, a common scheme is extracting multiple features followed by a ranking algorithm which however, incurs two major challenges: **1)** The ranking model often imposes pairwise constraints between each proposal, rendering the problem away from an efficient training/testing phase; **2)** Linear kernels are utilized due to the computational and memory bottleneck of training a kernelized model.

In this paper, we remedy these two issues by suggesting a *kernelized partial ranking model*. In particular, we demonstrate that **i)** our partial ranking model reduces the number of constraints from $O(n^2)$ to $O(nk)$ where n is the number of all potential proposals for an image but we are only interested in top-k of them that has the largest overlap with the ground truth; **ii)** we permit non-linear kernels in our model which is often superior to the linear classifier in terms of accuracy. For the sake of mitigating the computational and memory issues, we introduce a consistent weighted sampling (CWS) paradigm that approximates the non-linear kernel as well as facilitates an efficient learning. In fact, as we will show, training a linear CWS model amounts to learning a kernelized model. Extensive experiments demonstrate that equipped with the non-linear kernel and the partial ranking algorithm, recall at top-k proposals can be substantially improved.

1. Introduction

Objectness is an emerging topic in the computer vision community proposed by Alexe et al. [1], which aims to produce an ensemble of regions (i.e., object proposals) that have high probability to contain objects. The main advantage of object proposal is that it can dramatically reduce the search space from millions of positions, scales and aspect ratios to hundreds of suggested candidates while ensuring a high recall. Therefore, it is an important technique for further vision tasks such as object recognition, detection and scene understanding [2–5].

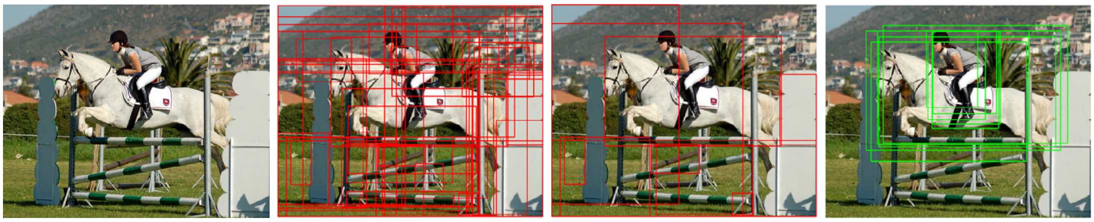
Since in most scenarios, object proposal actually serves as a preprocessing step, several important ingredients should be considered for a successful proposal algorithm. First, the algorithm should be fast enough. Otherwise, its superiority to the sliding window paradigm will be degraded. Second, it should produce a manageable number of proposals with a high recall.

To this end, a large body of works are devoted to effective features and fast grouping strategies. For example, in the work of [6], Cheng et al. designed a binary feature descriptor termed "Bing" and trained a linear model to estimate the locations of objects. Their algorithm is computationally efficient without loss of much accuracy. In [3,7,8], they utilized deep convolutional networks to extract features. The deep network mainly used GPU to speed up the process. Specifically, Ren et al. introduced a region proposal network which could be combined with the Zeiler and Fergus model [9] and the Simonyan and Zisserman model [4,5]. In [10,11], Uijlings et al. started with the low level super-pixels and carefully designed some simple yet effective features that could deal with a variety of image conditions. Then proposals were generated by grouping the super-pixels according to the handcrafted features. As there is not much computational cost in the grouping process, their algorithm is efficient. Notably, their model is fully unsupervised and hence no parameter will be learned or tuned. In [12–14], various visual cues, such as segmentation and saliency were utilized to describe a candidate region. Subsequently, based on the similarity of region features, a hierarchical grouping strategy was adopted to form the final object proposals. Arbeláez et al. [14] proposed a multi-scale hierarchical segmentation and grouped multiscale regions by features of size/location, shape and contours.

Usually, a proposal algorithm tends to produce a large number of candidates. Hence, existing algorithms always provide a confidence score for each candidate which indicates the probability of

* Corresponding author.
E-mail address: jing.julia.wang@gmail.com (J. Wang).

http://dx.doi.org/10.1016/j.patcog.2017.03.022
0031-3203/



(a) Original image (b) Proposals (c) SVM^{rank} (d) Top-k by ours

Fig. 1. Illustration of the importance of accurate estimation for top-k object proposals. Given an image shown as (a), existing object proposal methods (e.g., Selective Search) generate a set of proposals as illustrated in (b). Typically, *only* the top-k candidates are used to feed further vision tasks like object detection. In (c) and (d), we visualize the top-k results produced by SVM^{rank} and our model respectively. Clearly, our ranking model is superior to (c) since there is fewer inaccurate proposals within the top-k candidates.

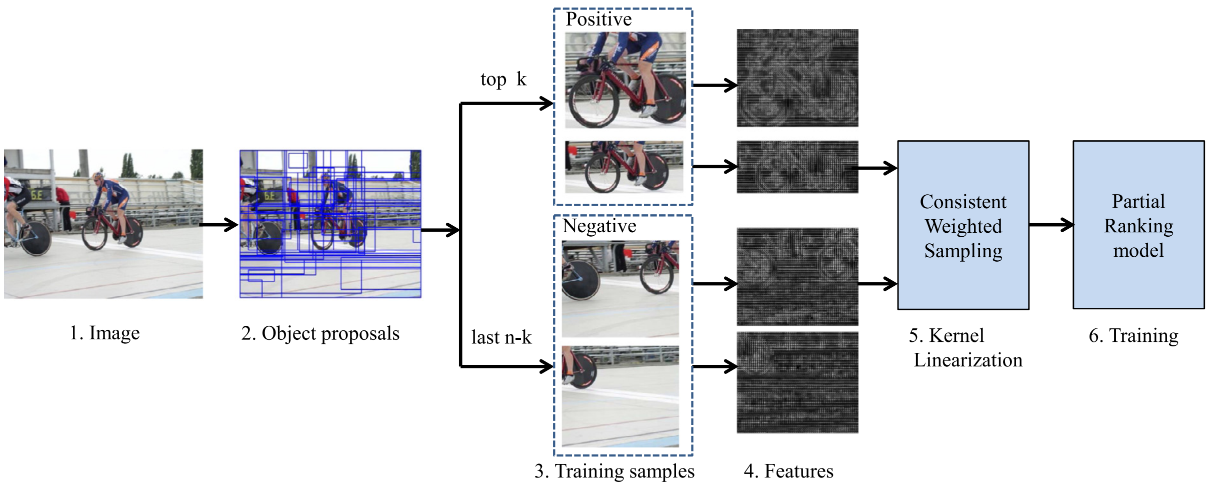


Fig. 2. Overview of the learning procedure. Our system (1) takes an input image, (2) obtains proposals which are produced by some previous proposal algorithm such as Selective Search, (3) splits the set of candidates into the top-k subset and the last $n-k$ subset according to the IoU to the ground truth, (4) computes features for each proposal followed by (5) consistent weighted sampling, and then (6) learns the partial ranking model based on the output of CWS.

containing an object. Commonly used schemes for the objectness scoring are summarized in [15,16]. Among them, the large margin based SVM^{rank}, or its variant is a popular solution [6,10,13,14,17]. Given all the candidates of an image, SVM^{rank} considers the pairwise ranks as constraints. However, imposing such full rankings for each candidate is possibly not necessary, and sometimes over constrained. To see this, consider the case that we have two candidates with Intersection Over Union (IoU) 0.01 and 0.001. Actually, they both can be treated as incorrect proposals. In this case, constraining the first candidate to have a higher rank than the other does not help much for the model construction. As we only care about the top-k candidates, a full ranking algorithm such as SVM^{rank} is not suitable for object proposals location. In Fig. 1, we give an example showing that an accurate prediction for the top-k candidates is more important than obtaining the rank for all candidates.

Related to the ranking algorithm, previous works usually devise hand-crafted features and feed them to a linear predictor. Yet, as has been known, non-linear kernels are often superior to the linear one in terms of prediction accuracy. One possible shortcoming of non-linear kernel is the memory and computation bottleneck. Fortunately, recent progress demonstrates that a class of popular kernels can be approximated by linear functions, such as shift-invariant kernels [18] and generalized min-max (GMM) kernels [19].

In this paper, henceforth, we propose a new partial ranking algorithm with support of non-linear kernel. The overview of the procedure is illustrated in Fig. 2. Given the ground truth and an ensemble of object candidates which are produced by existing methods, we compute the IoU for each candidate and then split these potential objects into two subsets, one of which consists of the top-k candidates and the remaining forms another group. The feature used here is the popular HOG [20], which will be described in Section 3.1. Yet, one can also replace it with other popular descriptors, such as SIFT or CNN features. Then we perform (0-bit) consistent weighted sampling (CWS) [19,21–23] on the features followed by learning our partial ranking model. In this way, learning a ranking model with non-linear kernel amounts to learning a linear hyperplane, hence efficient. The definition of CWS is deferred to Section 3.2. The derivation of our model and the learning algorithm are elaborated in Section 3.4.

The main difference of our model and other ranking methods is that, when training the model, we split the candidates of each image into two groups: one with top-k rankings and the other consisting of the remaining candidates. We only compare the candidate from the first subset and the one from the second subset, instead of comparing all pairs of candidates. On account of such constraints, our model can focus on obtaining a reliable prediction for only the top-k candidates rather than learning to rank all the candidates. Also note that our partial ranking model is different from top-k ranking models in information retrieval, which aims to provide an accurate ranking for each top-k retrieval [24]. In our case, it is not necessary to provide an accurate ranking within the top-k candidates in that, when utilizing the k proposals for further processing, like recognition, we typically do not care about the orders of proposals.

1.1. Contribution

We make two technical contributions in the work. First, by observing that the broadly studied SVM^{rank} usually over constrains the object proposal problem, we suggest a partial ranking