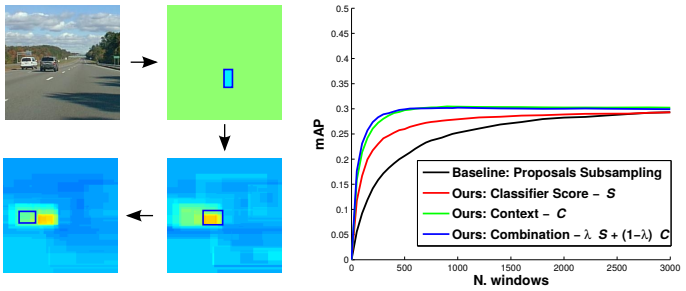
An Active Search Strategy for Efficient Object Class Detection

Abel Gonzalez-Garcia, Alexander Vezhnevets, Vittorio Ferrari
University of Edinburgh

Figure 1: Example of a car search using our method.



(a) (b)

Figure 2: (a) Our search method in action. (b) Results of our method tested on SUN2012 using R-CNN as window classifier. Baseline: evaluation of proposals in arbitrary order.

Modern object class detectors [2, 4, 5, 8] partition the image into a set of windows and then score each window with a classifier to determine whether it contains an instance of the object class. In the classical sliding window approach [2, 4], the window set is very large, containing hundred of thousands of windows on a regular grid at multiple scales. This approach is prohibitively expensive for slow, powerful window classifiers which are state-of-the-art nowadays [5, 8], such as Convolutional Neural Networks (CNN) [5]. For this reason, these detectors are based instead on object proposal generators [1, 7, 8], which provide a smaller set of a few thousand windows likely to cover all objects. Hence, this reduces the number of window classifier evaluations required. However, in both approaches the window classifier evaluates *all* windows in the set, effectively assuming that they are independent.

In this work, we propose an *active search strategy* that sequentially chooses the next window to evaluate based on previously observed windows, rather than going through the whole window set in an arbitrary order. Observing a window not only provides information about the presence of the object in that particular window, but also about its surroundings and even distant areas of the image. Our search method extracts this information and integrates it into the search, effectively guiding future observations to interesting areas, likely to contain objects. This results in a more natural and elegant way of searching, avoiding wasteful computation in uninteresting areas and focusing on the promising ones. As a consequence, our method is able to find the objects while evaluating much fewer windows, typically only a few hundreds.

We use two guiding forces in our method: context and window classifier score. Context exploits the statistical relation between the appearance and location of a window and its location relative to the objects, as observed in the training set. For example, the method can learn that cars tend to be on roads below the sky. Therefore, observing a window in the sky in a test image suggests the car is likely to be far below, whereas a window on the road suggests making a smaller horizontal move. We learn the context force, in a Random Forest framework that provides great computational efficiency as well as accurate results. The classifier score of an observed window provides information about the score of nearby windows, due to the smoothness of the classifier function. It guides the search to areas where we have observed a window with high score, while pushing away from windows with low score. Observing a window with part of a car, for example, will attract the search to its surroundings. Our method effectively combines these two forces.

Fig. 1 shows the intuition of our method on detecting cars. It starts at window w_0 and it moves away immediately, since w_0 contains a piece of building, not a car. Context determines the direction of the move, as cars tend to be on streets below buildings. Hence, the next visited location is on the road. After observing w_1, the method continues searching along the road, as indicated by context. For window w_2, however, the score of the classifier is rather high, as it contains a piece of car. Therefore, the search focuses around this area until it finds a tight window on the car.

Experiments on the challenging SUN2012 dataset [9] and PASCAL VOC10 [3] demonstrate that our method explores the image in an intelligent way, effectively detecting objects in only a few hundred iterations. Fig. 2(a) presents an example of our search method in action. Our method can be applied on top of any classifier, we use the state-of-the-art R-CNN [5] and the popular UvA Bag-of-Words model of [8], both using of object proposals [8]. For R-CNN on SUN2012, our search strategy matches the detection accuracy of evaluating all proposals independently, while evaluating 9× fewer proposals on average, as we can see in fig. 2 (b). As our method adds little overhead, this translates into an actual wall-clock speedup. When computing CNN features on the CPU [6], the processing time for one test image reduces from 320s to 36s (9× speed-up). When using a GPU, it reduces from 14.4s to 2.5s (6× speed-up). Hence, our method opens the door to using expensive classifiers by considerably reducing the number of evaluations while adding little overhead. For the UvA window classifier, our search strategy only needs 35 proposals to match the performance of evaluating all of them (a reduction of 85×). By letting the search run for longer, we even *improve* accuracy while evaluating 30× fewer proposals, as it avoids evaluating some cluttered image areas that might lead to false-positives.

[1] B. Alexe, T. Deselaers, and V. Ferrari. Measuring the objectness of image windows. *IEEE Trans. on PAMI*, 2012.

[2] N. Dalal and B. Triggs. Histogram of Oriented Gradients for human detection. In *CVPR*, 2005.

[3] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *IJCV*, 2010.

[4] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *IEEE Trans. on PAMI*, 32(9), 2010.

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

[6] Y. Jia. Caffe: An open source convolutional architecture for fast feature embedding. http://caffe.berkeleyvision.org/, 2013.

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

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

[9] J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba. SUN database: Large-scale scene recognition from Abbey to Zoo. In *CVPR*, 2010.

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