# The Role of Context Selection in Object Detection

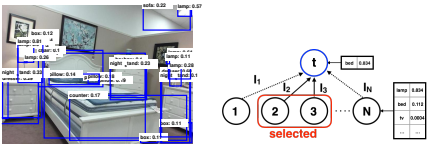
Ruichi (Rich) Yu[1]
richyu@cs.umd.edu
Xi (Stephen) Chen [2]
chnxi@microsoft.com
Vlad I. Morariu[1]
morariu@umiacs.umd.edu
Larry S. Davis[1]
lsd@umiacs.umd.edu

[1] University of Maryland
College Park, MD. USA.

[2] Microsoft Corporation
One Microsoft Way,
Redmond, WA. USA.

We investigate why the utility of context information in object detection is limited through the evaluation of the effect of different pure context cues. We analyze the predictive potential of context in an idealized case where the labels of all contextual objects are known, and only these labels and their relationships to a target object are used to predict the target object label. These experiments reveal that, despite ignoring the appearance of the target object, pure context is effective at predicting the target object class. Not surprisingly, different categories vary in their ability to predict certain target objects. Based on this study, we propose a region-based context re-scoring method with dynamic context selection, illustrated in figure 1(b), which tries to eliminate false positive contextual regions while emphasizing likely true positive and informative ones. Specifically, we introduce a latent variable for each contextual region that determines if that region will be selected to provide context information. In practice, it is intractable to select the optimal set of contextual regions that provide the most trustworthy information when contradictory evidence exists, *for* and *against* the target object being in a certain class. Instead, we decompose the problem by selecting informative regions providing the strongest supporting and refuting evidence independently to compute a *For upper-bound* (FUB) and an *Against upper-bound* (AUB) of the confidence score, and then re-score the confidence for that object being in that class with the difference between the two upper-bounds. The model for computing the two upper-bounds is trained by latent-SVM [1].

The proposed method is evaluated on the SUN RGB-D dataset and achieves 48.25% mean average precision (mAP), an improvement of $\sim$ 2.8% over using object detections without context (45.47%). We also conduct experiments to study the performance of the selection model. Both the simulations on pure context and the real-world experiments using the proposed selection method demonstrate the importance of object-to-object context and the gain attributed to the context selection scheme.




(a) Detection Results (b) Context Selection

Figure 1: (a) Imperfect detections from the Fast R-CNN detector; (b) The proposed context selection method.

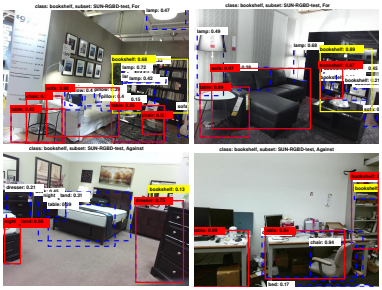


Figure 2: 1st row: FUB model. 2nd row AUB model. The yellow boxes are the target objects, the red boxes are the selected contextual regions, and the blue dashed boxes are the ones that are not selected.

[1] Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(9): 1627–1645, September 2010.