Abstract:

The recurring context in which objects appear holds valuable information that can be employed to predict their existence. This intuitive observation indeed led many researchers to endow appearance-based detectors with explicit reasoning about context. The underlying thesis suggests that stronger contextual relations would facilitate greater improvements in detection capacity. In practice, however, the observed improvement in many cases is modest at best, and often only marginal. In this work we seek to improve our understanding of this phenomenon, in part by pursuing an opposite approach. Instead of attempting to improve detection scores by employing context, we treat the utility of context as an optimization problem: to what extent can detection scores be improved by considering context or any other kind of additional information? With this approach we explore the bounds of improvement by using contextual relations between objects and provide a tool for identifying the most helpful ones. We show that simple co-occurrence relations can often provide large gains, while in other cases a significant improvement is simply impossible or impractical with either co-occurrence or more precise spatial relations. To better understand these results we then analyze the ability of context to handle different types of false detections, revealing that tested contextual information cannot ameliorate localization errors, severely limiting its gains. These and additional insights further our understanding on where and why utilization of context for object detection succeeds and fails.