

Meta-Tracker: Fast and Robust Online Adaptation for Visual Object Trackers

Eunbyung Park Alexander C. Berg

Department of Computer Science,
University of North Carolina at Chapel Hill
{eunbyung, aberg}@cs.unc.edu

Abstract. This paper improves state-of-the-art visual object trackers that use online adaptation. Our core contribution is an online meta-learning-based method to adjust the initial deep networks used in online adaptation-based tracking. The meta learning is driven by the goal of deep networks that can quickly be adapted to robustly model a particular target in future frames. Ideally the resulting models focus on features that are useful for future frames, and avoid overfitting to background clutter, small parts of the target, or noise. By enforcing a small number of update iterations during meta-learning, the resulting networks train significantly faster. We demonstrate this approach on top of the high performance tracking approaches: tracking-by-detection based MDNet [1] and the correlation based CREST [2]. Experimental results on standard benchmarks, OTB2015 [3] and VOT2016 [4], show that our meta-learned versions of both trackers improve speed, accuracy, and robustness.

1 Introduction

Visual object tracking is a task that locates target objects precisely over a sequence of image frames given a target bounding box at the initial frame. In contrast to other object recognition tasks, such as object category classification and detection, in visual object tracking, instance-level discrimination is an important factor. For example, a target of interest could be one particular person in a crowd, or a specific product (e.g. coke can) in a broader category (e.g. soda cans). Therefore, an accurate object tracker should be capable of not only recognizing generic objects from background clutter and other categories of objects, but also discriminating a particular target among similar distractors that may be of the same category. Furthermore, the model learned during tracking should be flexible to account for appearance variations of the target due to viewpoint change, occlusion, and deformation.

One approach to these challenges is applying online adaptation. The model of the target during tracking, e.g. DCF (discriminative correlation filter) or binary classifier (the object vs backgrounds), is initialized at the first frame of a sequence, and then updated to be adapted to target appearance in subsequent

The code is available at https://github.com/silverbottlep/meta_trackers.

