

Fast CNN-Based Object Tracking Using Localization Layers and Deep Features Interpolation

Object trackers based on Convolution Neural Network (CNN) have achieved state-of-the-art performance on recent tracking benchmarks, while they suffer from slow computational speed. The high computational load arises from the extraction of the feature maps of the candidate and training patches in every video frame. The candidate and training patches are typically placed randomly around the previous target location and the estimated target location respectively. In this paper, we propose novel schemes to speed-up the processing of the CNN-based trackers. We input the whole region-of-interest once to the CNN to eliminate the redundant computations of the random candidate patches. In addition to classifying each candidate patch as an object or background, we adapt the CNN to classify the target location inside the object patches as a coarse localization step, and we employ bilinear interpolation for the CNN feature maps as a fine localization step. Moreover, bilinear interpolation is exploited to generate CNN feature maps of the training patches without actually forwarding the training patches through the network which achieves a significant reduction of the required computations. Our tracker does not rely on offline video training. It achieves competitive performance results on the OTB benchmark with 8x speed improvements compared to the equivalent tracker.