

- 1 selects a template out of a set of N templates ($N=4$) and uses it as input to an existing deep CNN based tracker
- 2 The template set is updated at regular intervals but its size remains same since an old one is removed whenever a new one is added
- 3 two part network \rightarrow matching network that produces a prediction heatmap for each template and a policy network that produces a normalized score for each map.
- 4 MN takes a crop of the input image and a template as input \rightarrow crop is generated based on the last known location of the tracker
- 5 A variant of policy based RL is used for training the PN \rightarrow input prediction map is treated as the state s and action a is the choice of one of the templates; reward is given only at the end of the episode based on the success of the tracker which is defined as having an overlap error less than a threshold
- 6 a variant of the REINFORCE algorithm is used for training \rightarrow for each frame/decision in an episode, gradients are obtained by BP assuming that the decision was optimal and these are accumulated for a single update based on the overall outcome of the episode \rightarrow +ve update for success and -ve for failure
- 7 There is also something called stochastically sampled action roll-outs for better exploration of the state space but this is not explained in detail
- 8 There is also an experience replay memory maintained from previous episodes from which a certain fraction of gradients are sampled randomly while applying the update. This is supposed to 1) remove correlation in memory data and 2) reduce variance of the update
- 9 Adam optimizer is used for MN and adagrad one for the PN
- 10 50K episodes of lengths varying from 30 to 300 are used and template is updated every 50 frames. Tracker success is determined by the mean IoU in the last 20 frames of each episode.