PhD

Tracking by Animation Unsupervised Learning of Multi-Object Attentive Trackers cvpr19 ax1809.03137

2019-11-28 6:57:54 PM

The main idea is to learn multi object tracking in an end to end fashion and in a completely unsupervised way by using the tracking results to reconstruct the input frames and then constructing a loss on the difference between the input and the reconstructed frames

Uses background subtraction to extract all of the non-tracked areas in the frames which only works for static backgrounds which limits its applicability

Variable number of tracked objects are dealt by using a maximum on the number of objects that the system can track and something called adaptive computation time is used to reduce all the extra computation caused by the fact that number of objects is usually much smaller than this maximum

The latter seems to consists mostly of processing trackers in decreasing order of confidence and terminating iterations for trackers after the confidence falls below a threshold, 0.5 in this case

A variation on RNN is proposed for tracking/data association called reprioritized attentive tracking or RAT where the main ideas seems to be to use a kind of filtering or attentive mechanism to preprocess the raw feature maps before passing them through the RNN so that each tracker is sort of able to limit its processing only to its corresponding object without interfering too much with other trackers

some sort of neural Turing machine inspired mechanism is used for sequential reading and writing of shared memory type layers so that trackers are able to communicate with each other and avoid conflicts

Occlusion is modeled using a Photoshop style multi-layer model for an image where each tracker works on one of several overlapping layers so that the final reconstruction is done by a combination of layer compositing followed by frame compositing

the former cannot deal with conclusion since outputs from multiple trackers in the same layer are simply added but it can be parallelized while the latter can handle occlusion but cannot be parallelized so combination is supposed to provide a balance between the two

This renderer is both differentiable and parameter less – it is deterministic to supposedly allow better tracker training while also allowing back propagation through it

Testing is only done on a couple of synthetic toy data sets on which the performance seems to be very good as well as on a relatively easy real word data set with static backgrounds where the performance seems to be comparable to modern state of the art trackers