PhD

Deep Affinity Network for Multiple Object Tracking ax1810.11780 tpami19

2020-02-08 1:00:40 PM

Methods for preprocessing pairs of frames for MOT inspired by a similar method for detectors

Main idea is joint appearance modeling and association metrics prediction in a single end to end trainable network

A pair of frames along with the the centers up all the objects in the two frames used as input – a maximum number of objects set to 80 used to achieve fixed sized input and output

dummy rows and columns with all zeros used to fill up the Association matrix in cases of fewer objects

2020-02-17 6:36:56 PM

Feature extraction is based on the rather dubious idea that the pixel at the center of each object encodes important information about it – extended to select feature values from nine empirically selected layers and concatenated together to create a 520 dimensional feature vector for each object

The two 520 x 80 matrices are combined together in such a way that each feature vector in the first one is concatenated with every single vector in the second one to allow all possible permutations and thus create a 80 x 80 x 1040 dimensional tensor

Finally this one is subjected to a compression network with 5 depth wise convolutional layers to convert it into a single 80 x 80 Association matrix

Two different association matrices of size 81 x 80 and 80 x 81 are constructed by adding an extra row and column to the base 80 x 80 matrix to account for objects that disappear or appear in the second frame and carry out both forward and backward association

this extra row and column seem to be filled with some sort of hyper parameter (=10) rather than zeros

composite loss is made up of four different losses – one each for the two association matrices, another to maintain consistency between them and an assemble loss whose purpose is not quite clear but seems to have something to do with non-maximum Association suppression

During inference, only a single frame is passed through the network and its features are compared with stored features from several previous frames to compute the final association matrices which are then combined together in some sort of accumulator matrix and then subjected to the Hungarian algorithm

this last process seems to be riddled with heuristics including a certain number of frames for which trajectories are allowed to disappear before being removed from the stored trajectories as well as a maximum number of previous frames with which comparisons are made in each frame

Performance on DETRAC and MOT seems comparable to the state of the art but not significantly better