PhD

Context Matters Refining Object Detection in Video with Recurrent Neural Networks bmvc16

2018/05/18 11:12 AM

generate psedo labels - preliminary low quality detections - using standard still image detectors
train RNN with GRU to improve these

2018/10/10 9:22 AM

- uses a two-step training process
- first, it fine-tunes the yolo object detector using the YouTube video data set where apparently only the last layer is trained but the the loss itself seems to be pretty much the same as yolo;
- in the second step, the output of this fine tuned Yolo detector, known as pseudo labels for some reason, is further used to train an RNN using a composite loss which is supposed to encourage accuracy in individual frames as well as consistency across time;
- the RNN and has two layers with 150 nodes per layer and is of the type gated recurrent unit or GRU;
- the four losses include detection loss which is similar to the use or loss, the category loss vid share only penalizes incorrect classification and ignores localization errors, the consistency loss which minimizes the difference between the object locations in consecutive frames and finally theHello hello hello
- similarity loss which tries to ensure that the predictions of the RNN are similar to the pseudo labels at each time step;
- based on the design of these losses, it seems that the YouTube data set does not provide the ground truth for all of the frames but only at the end of each sequence so it seems that the ground truth available on the final frame of the sequence is used for some of the losses while other heuristic stuff is used for per frame losses;