Smoothness Metric

Overview

We add a metric to measure the smoothness or stability of objects detected across the frames in the video.

VLM Score vs Smoothness Score

While the VLM metric is good to filter out exploitative miners, the smoothness metric is a good metric to distinguish between high performing miners. We apply this metric after the VLM metrics and it is a complementary metric since the VLM-based metrics focus more on the spatial accuracy of the bbox predictions (on a subset of frames) whereas the smoothness focuses on the temporal accuracy of the bbox predictions (on all frames)

Intuition

We assume that smoother motion is equivalent to a higher inter-frame IoU. Smooth movements will overlap more with the previous frame than disconnected or sudden movements. High IoU ⇒ Smoother transition. Low IoU ⇒ Sudden Changes

Definitions

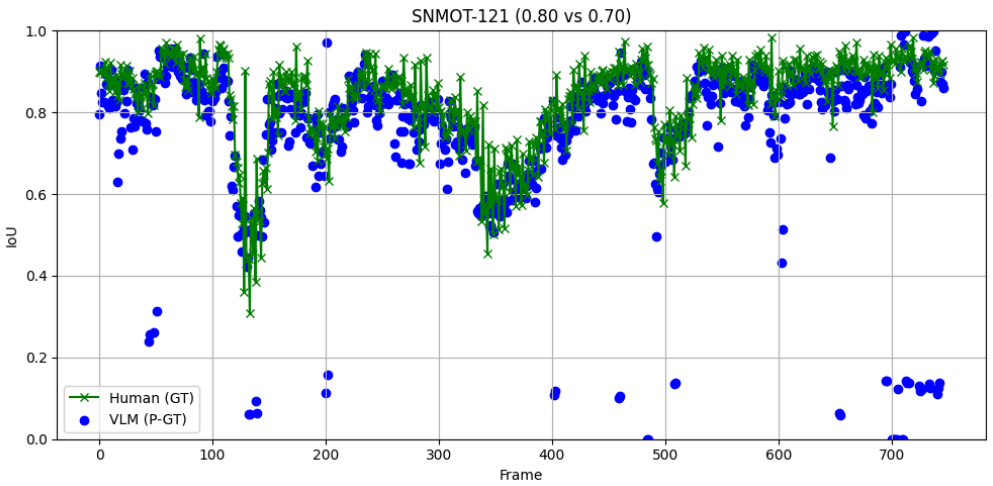
Smoothness = mean Inter-frame IoU / (1 + jerkiness)

Jerkiness = mean absolute inter-frame IoU

inter-frame IoU = IoU of frame@t vs frame@t+1

Smoothness of GT vs Pseudo-GT

The Human annotation averages a smoothness score of 80% while Pseudo-GT gets 70%. Random exploits get 30% smoothness. An example for one test video (SNMOT-121) is shown below


SNMOT-121 (0.80 vs 0.70)
IoU
Frame
Human (GT)
VLM (P-GT)
0.0
0.2
0.4
0.6
0.8
1.0
0
100
200
300
400
500
600
700