

Joint Detection and Recounting of Abnormal Events by Learning Deep Generic Knowledge Supplementary Material

1. Feature space visualization

Figure 1 shows the features embedded into two-dimensions using a t-SNE algorithm to compare the discriminability of the feature qualitatively. We compare three features: HOG, SDAE, and our multi-task Fast R-CNN features on Avenue and UCSD Ped2 datasets. Multi-task Fast R-CNN can discriminate truck, bike, and person while the HOG feature cannot. Although the SDAE feature discriminates truck, it cannot discriminate person and bike, which are difficult to distinguish without learning these concepts.

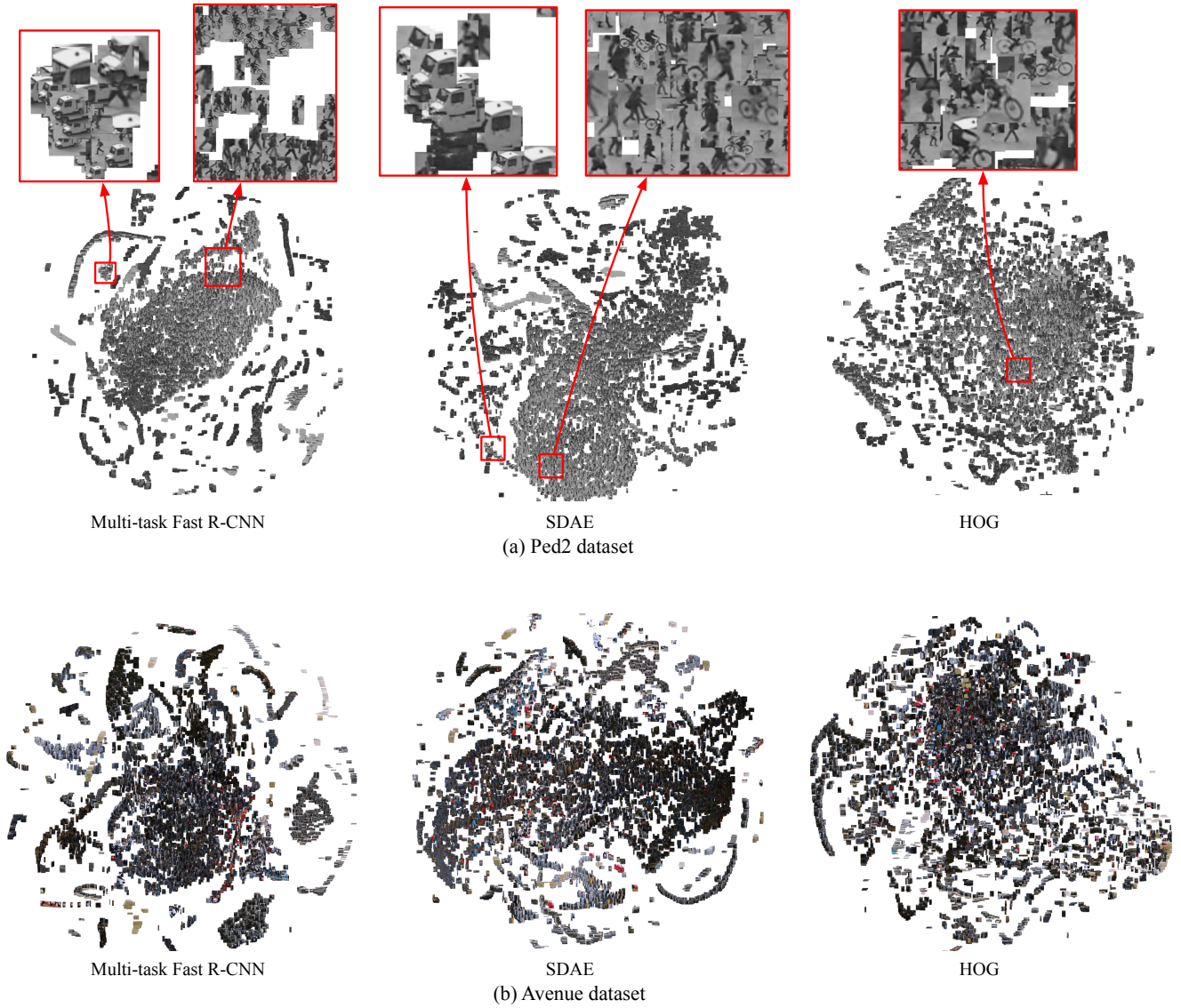


Figure 1. Visualization of the feature spaces using t-SNE on (a) Avenue and (b) UCSD Ped2 datasets.

2. Avenue17 dataset

We evaluated Avenue datasets using the subset (i.e., Avenue17 dataset) excluding five clips out of 22 clips that contained static but abnormal objects, viz., a red bag on the grass and a person standing on the grass, which is regarded as normal in the Avenue dataset. Figure 2 (a) and (b) show the static abnormal objects in the excluded 5 clips and Fig. 2 (c) and (d) show normal (training) and abnormal scenes respectively in the original Avenue dataset. While no objects are put on the grass in all frames in the training data, there is a red bag on the grass in the test data. Besides, while many people walking in the wrong direction are annotated as abnormal objects in (d), the person standing on the grass in (b) is not annotated even if it is rarer and more prominent than others. Therefore, we remove the five clips that contain the red bag (clip 01 and 02) and the person standing on the grass (clip 08, 09, and 10).

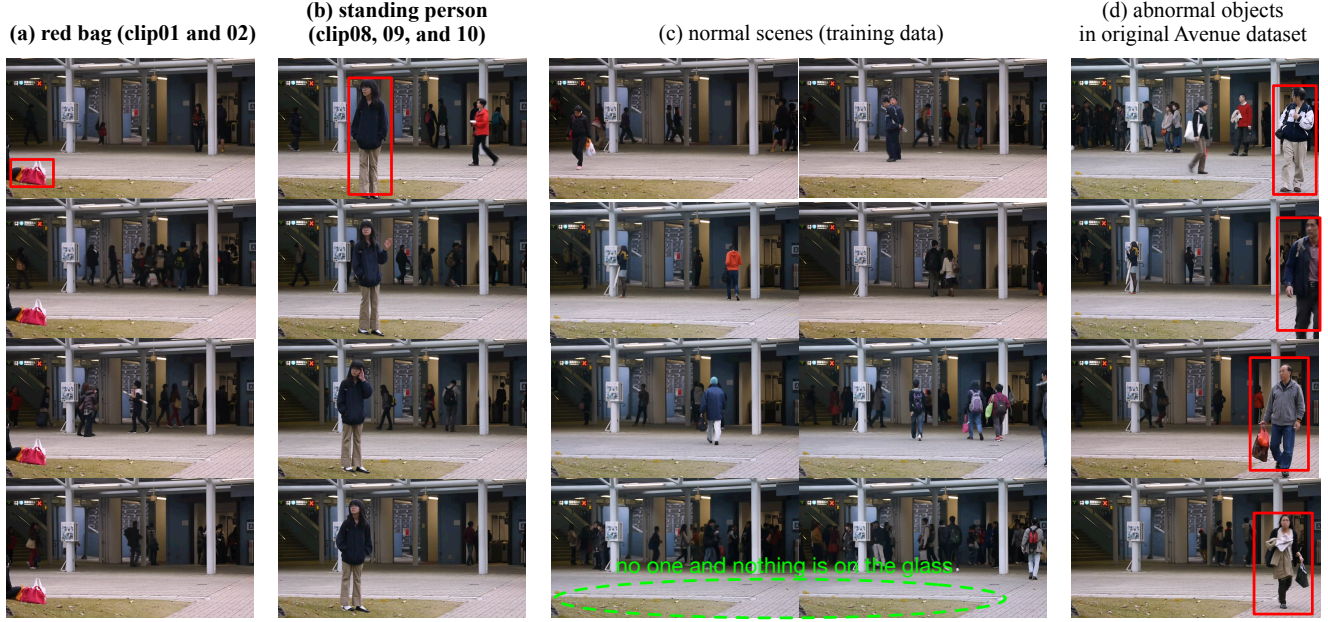


Figure 2. (a), (b), and (c) are normal and (d) is abnormal of Avenue dataset. We consider two objects in (a) and (b) are abnormal and excluded the five clips that contain them.