Learning a distance function with a Siamese network to localize anomalies in videos

Bharathkumar Ramachandra
North Carolina State University
Raleigh, NC 27695
bramach2@ncsu.edu

Michael J. Jones
Mitsubishi Electric Research Labs (MERL)
201 Broadway, 8th floor; Cambridge, MA 02478
mjones@merl.com

Ranga Raju Vatsavai
North Carolina State University
Raleigh, NC 27695
rrvatsav@ncsu.edu

1. Understanding the distance function learned

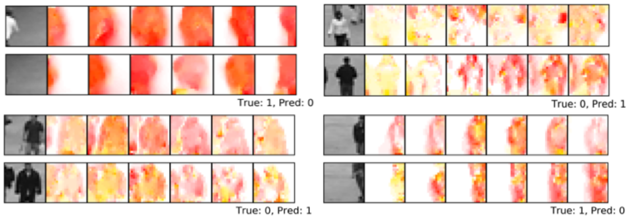

Figure 1. Examples of large prediction errors made by our model on UCSD Ped1. Classes 0 and 1 refer to similar and dissimilar pairs respectively. Best viewed in color.

We also tried to gain some insight into what properties the distance function learned by the CNN possesses. To this end, we recorded the video patch pairs on which the CNN makes large errors, that is, either classifying similar pairs as dissimilar or vice versa, with high predicted probability. Figure 1 is a visualization of 4 such video patch pairs when the target dataset is UCSD Ped1. Remarkably, the CNN seems to find it hard to correctly classify examples that are conceivably hard for humans. Specifically, the dissimilar pairs that have been misclassified seem to contain a skateboarder moving only slightly faster than a pedestrian would, and the similar pairs that have been misclassified exhibit some distinct differences in their flow fields.

2. Track and region based ROC curves

Figures 2 through 7 show the ROC curves for our CNN approach (denoted "CNN distance") as well as that of [1]'s FG masks (denoted "FG L2 distance") and flow (denoted "Flow L1 distance") methods on all 3 datasets. Overall, it appears that our approach of using a learned representation and learned distance function is able to achieve better detection performance, demonstrated by higher true positive rates at low false positive rates.

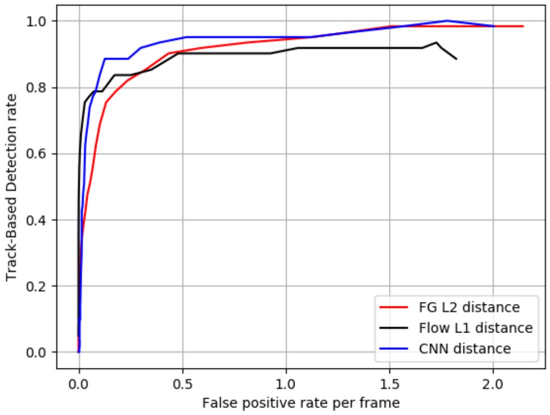

Figure 2. Track-based ROC curves on UCSD Ped1.

3. More detection result visualizations

Figures 8 through 25 present additional true positive, false positive and false negative detection results from our approach for all 3 datasets. As in the submission document, the green bounding boxes refer to ground truth anomalies and the red regions our detections at a fixed threshold on anomaly scores.

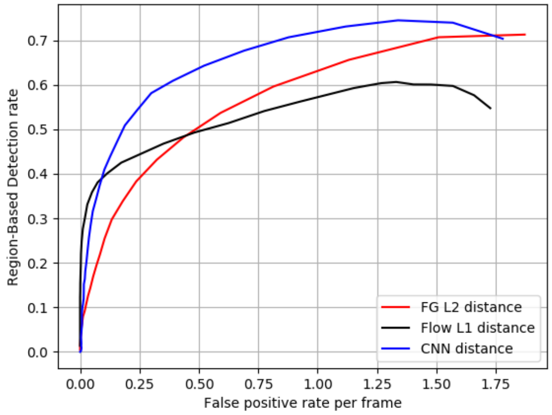

Figure 3. Region-based ROC curves on UCSD Ped1.

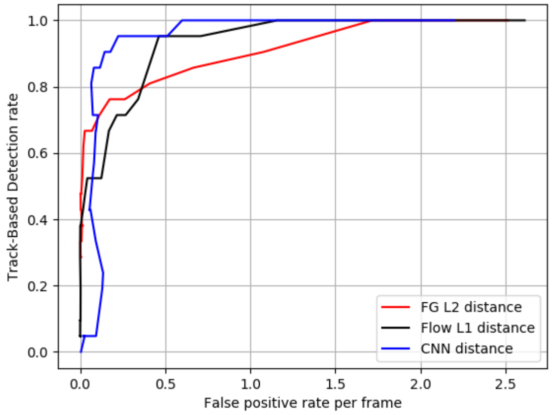

Figure 4. Track-based ROC curves on UCSD Ped2.

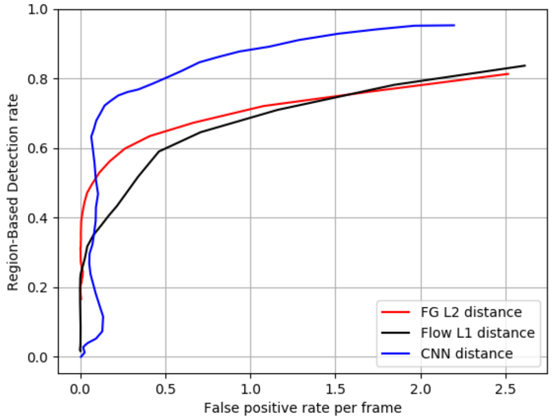

Figure 5. Region-based ROC curves on UCSD Ped2.

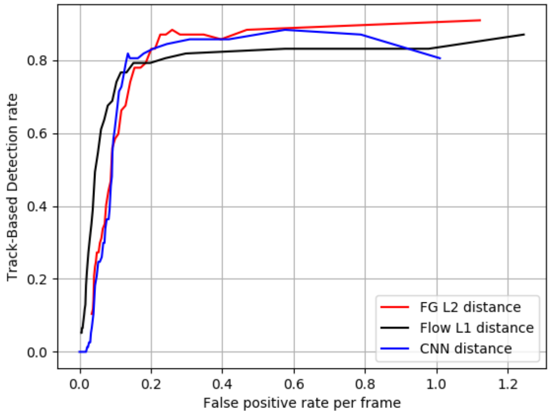

Figure 6. Track-based ROC curves on CUHK Avenue.

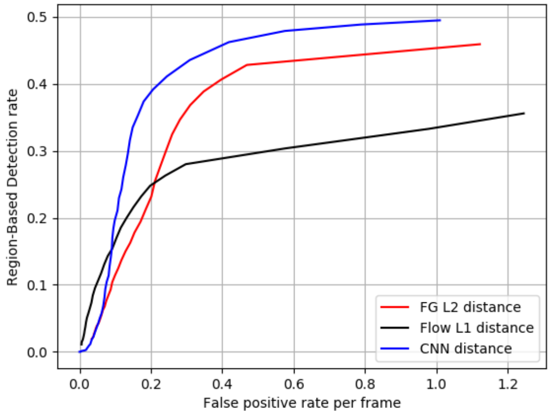

Figure 7. Region-based ROC curves on CUHK Avenue.

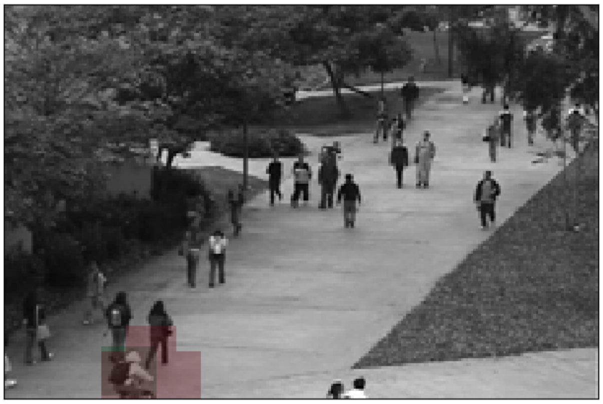

Figure 8. True positive in UCSD Ped1 - a biker.

4. More frame-level anomaly score visualizations

Figures 26 through 31 provide additional frame-level anomaly score visualizations for some test sequences us-

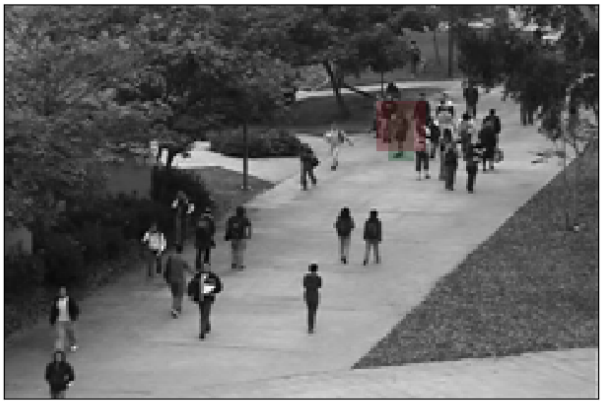

Figure 9. True positive in UCSD Ped1 - a skateboarder.

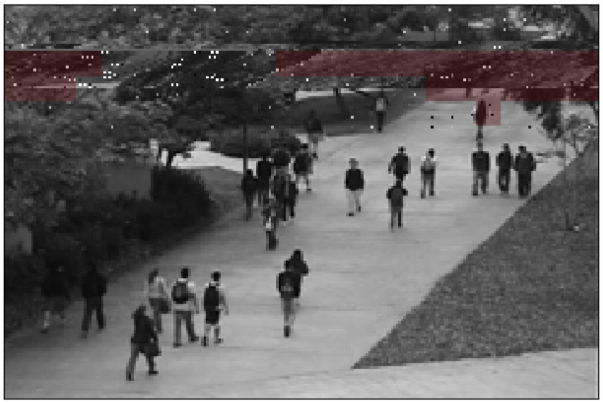

Figure 10. False positive in UCSD Ped1 - camera fault.

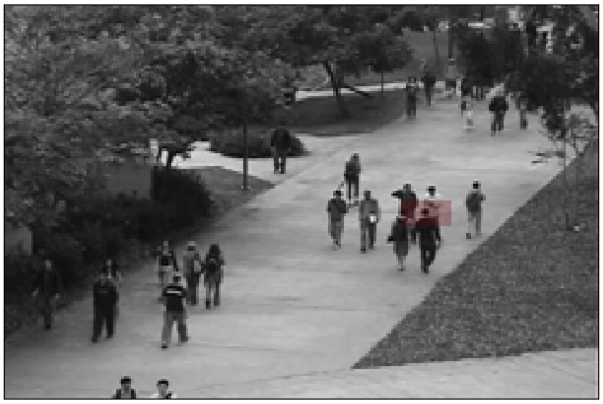

Figure 11. False positive in UCSD Ped1 - seemingly random.

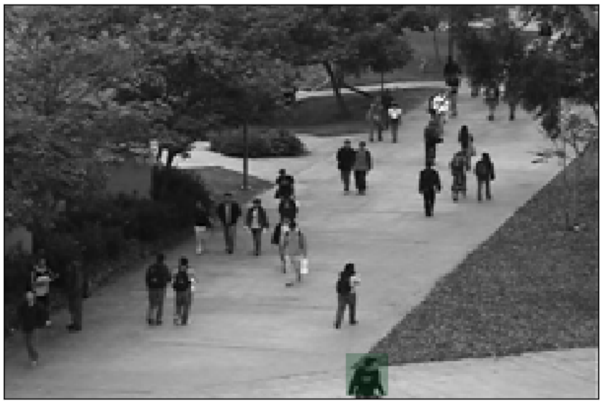

Figure 12. False negative in UCSD Ped1 - biker not yet fully in the camera frame.

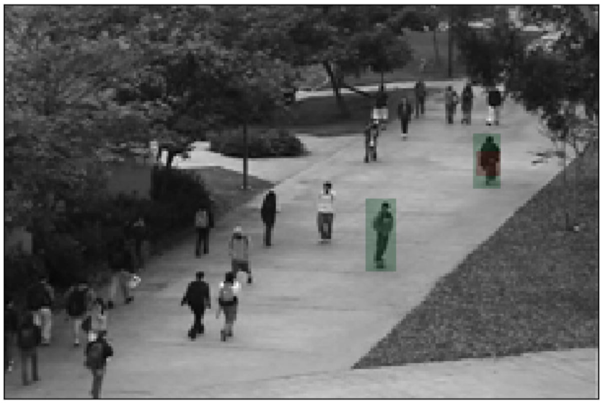

Figure 13. False negative in UCSD Ped1 - skateboarder moving slowly.

ing our approach from all 3 datasets. As in the submission document, green shading on the plot indicates ground truth anomalous frames and we also show detection visualizations at select frames.

5. Visualizations of learned representations for video patch pairs

Figures 32 through 36 show select video patch pairs from UCSD Ped2, their learned representations and the distance measured between them by our CNN. To generate this set of figures, we used the CNN corresponding to the scenario where the target dataset was UCSD Ped2 to give a realistic idea of distance measurement at 'test time'. Each group of 3 rows is a visualization of the feature maps of the first video patch before element-wise subtraction (1st row), the second video patch before element-wise subtraction (2nd row), and the element-wise subtraction layer's output (3rd row). All 128 feature maps are shown on columns, wrapping around

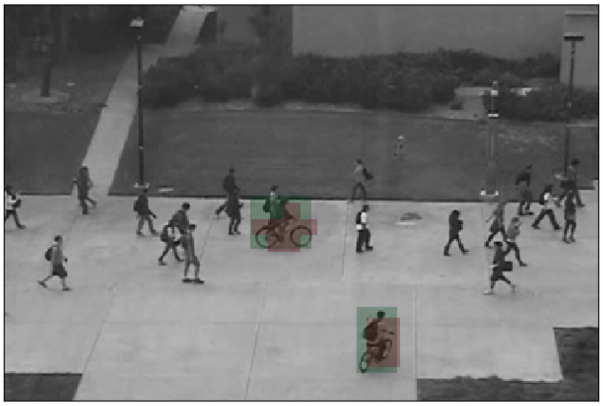

Figure 14. True positive in UCSD Ped2 - 2 bikers.

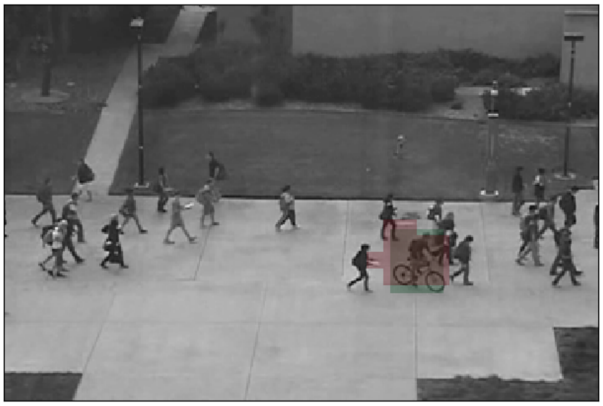

Figure 15. True positive in UCSD Ped2 - a biker.

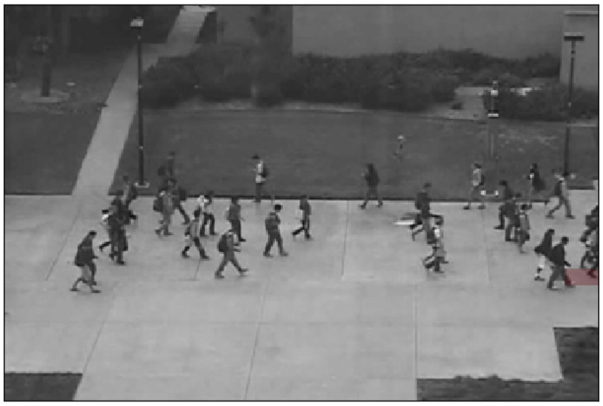

Figure 16. False positive in UCSD Ped2 - seemingly random.

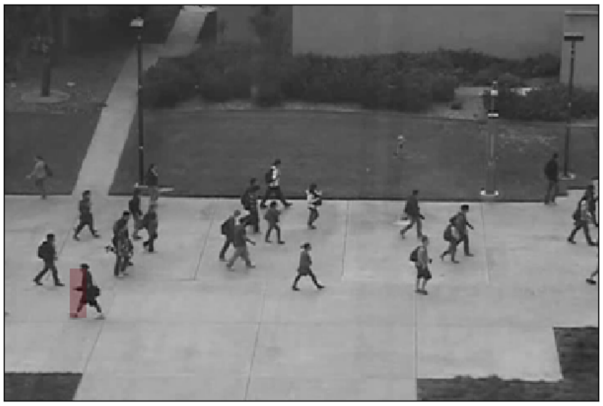

Figure 17. False positive in UCSD Ped2 - unusual movement in this region of the camera frame.

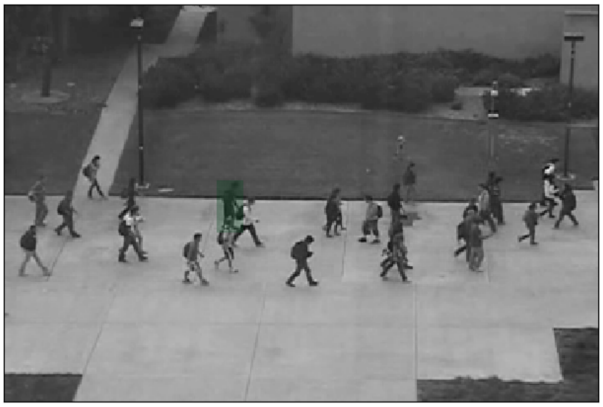

Figure 18. False negative in UCSD Ped2 - occluded, slow-moving skateboarder.

to the next row when necessary. Specific feature maps could exhibit high activations for features such as speed, direction, velocity, shape, texture and illumination among others.

References

[1] B. Ramachandra and M. Jones. Street Scene: A new dataset and evaluation protocol for video anomaly detection. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2020. 1

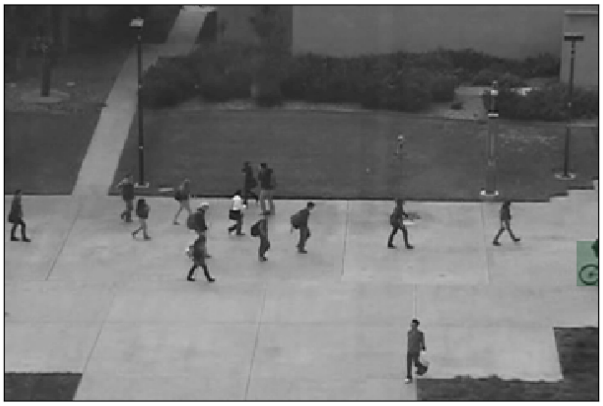

Figure 19. False negative in UCSD Ped2 - biker partially left the camera frame.

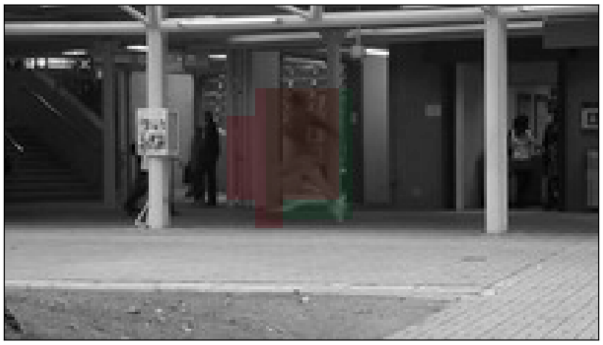

Figure 20. True positive in CUHK Avenue - person running.

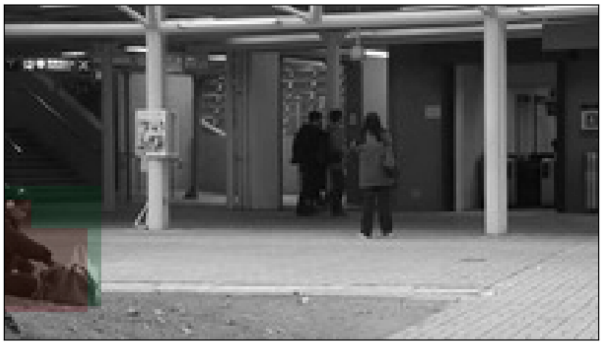

Figure 21. True positive in CUHK Avenue - person interacting with a bag on the grass.

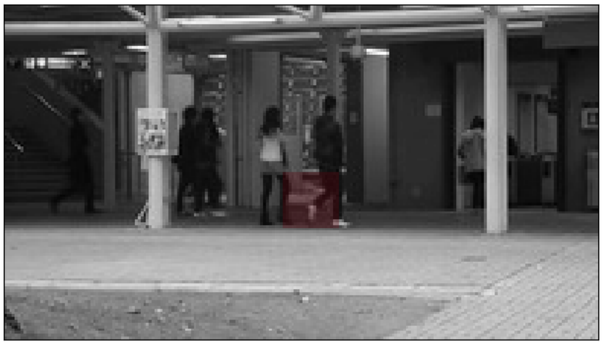

Figure 22. False positive in CUHK Avenue - seemingly random.

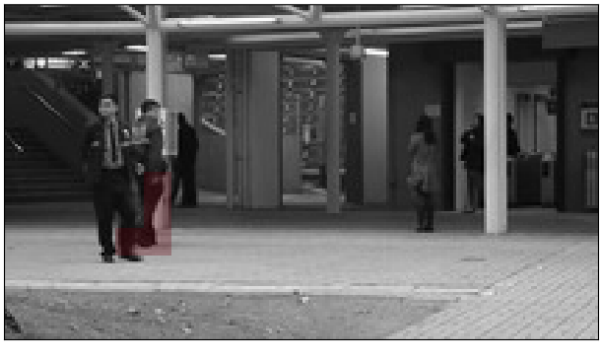

Figure 23. False positive in CUHK Avenue - unusual movement in this region of the camera frame.

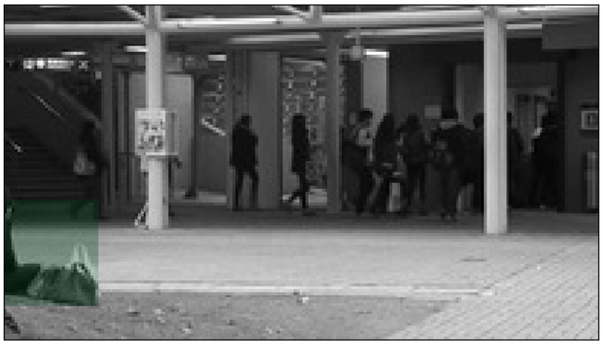

Figure 24. False negative in CUHK Avenue - still, unattended bag.

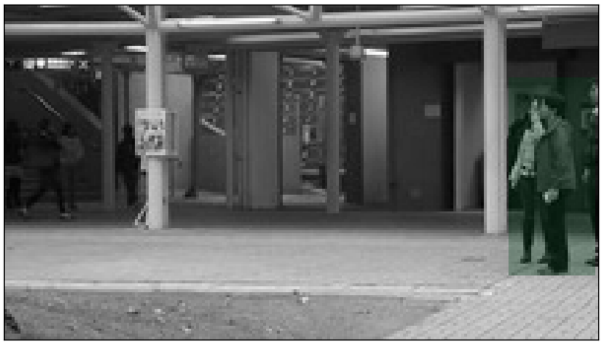

Figure 25. False negative in CUHK Avenue - start of an anomalous event that is seemingly normal.

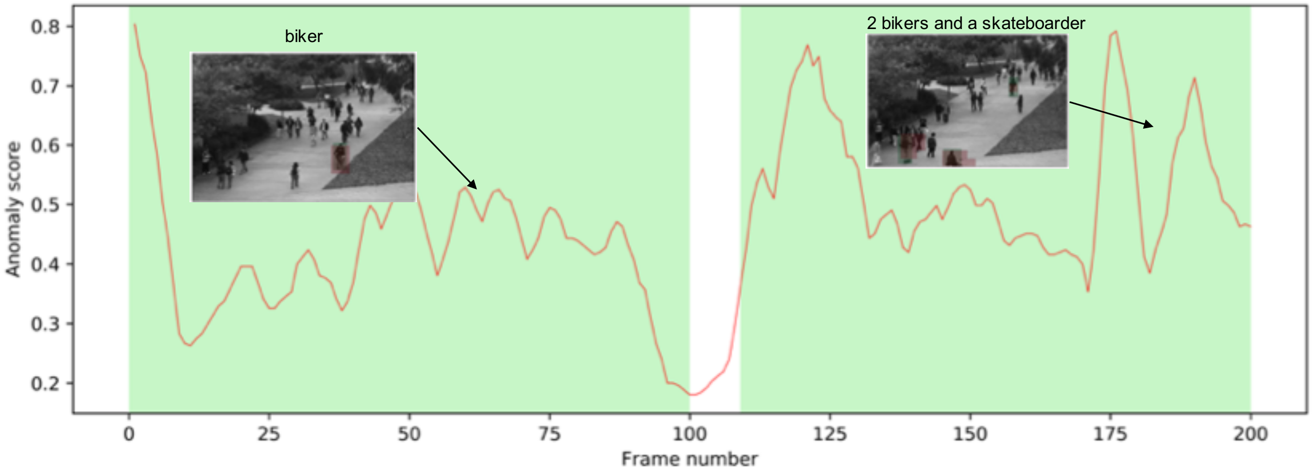

Figure 26. Per-frame anomaly score visualization of UCSD Ped1 Test sequence 006.

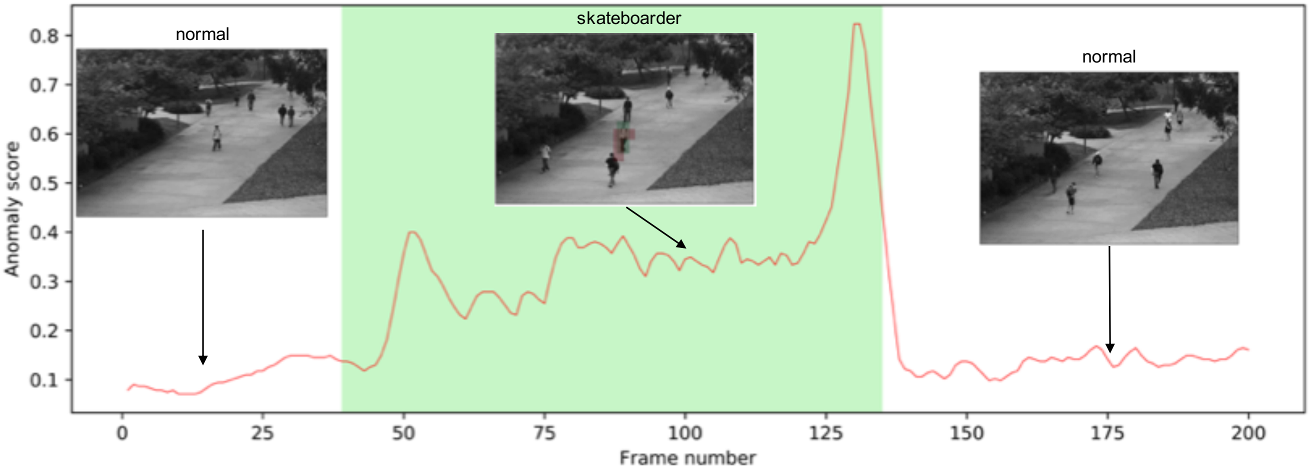

Figure 27. Per-frame anomaly score visualization of UCSD Ped1 Test sequence 025.

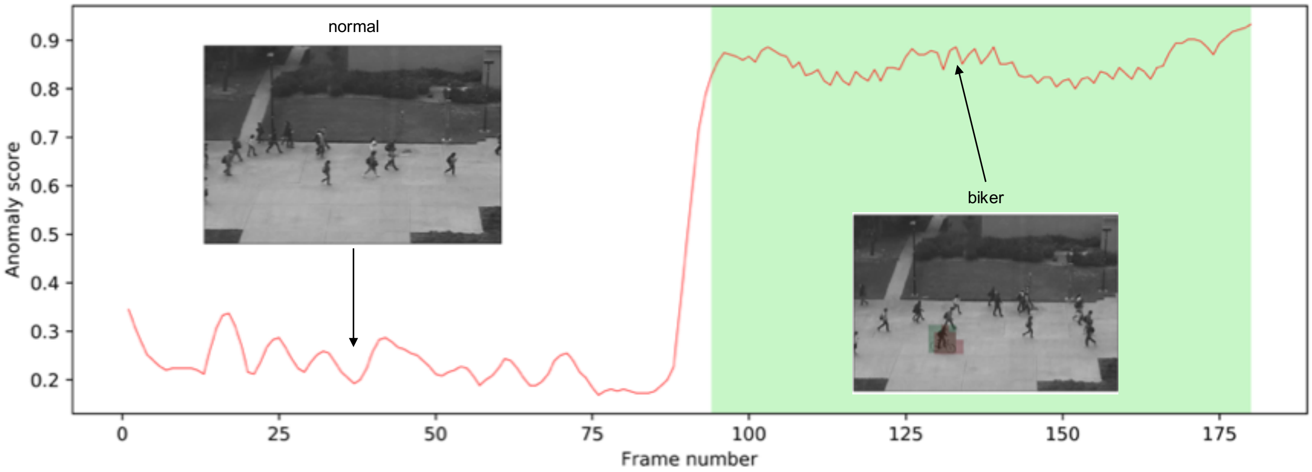

Figure 28. Per-frame anomaly score visualization of UCSD Ped2 Test sequence 002.

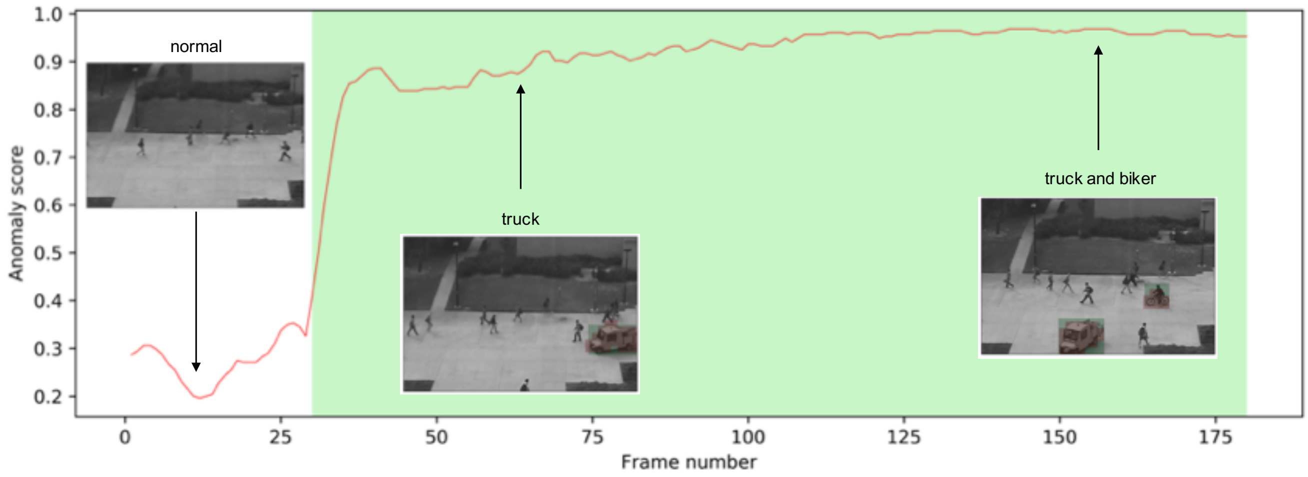

Figure 29. Per-frame anomaly score visualization of UCSD Ped2 Test sequence 004.

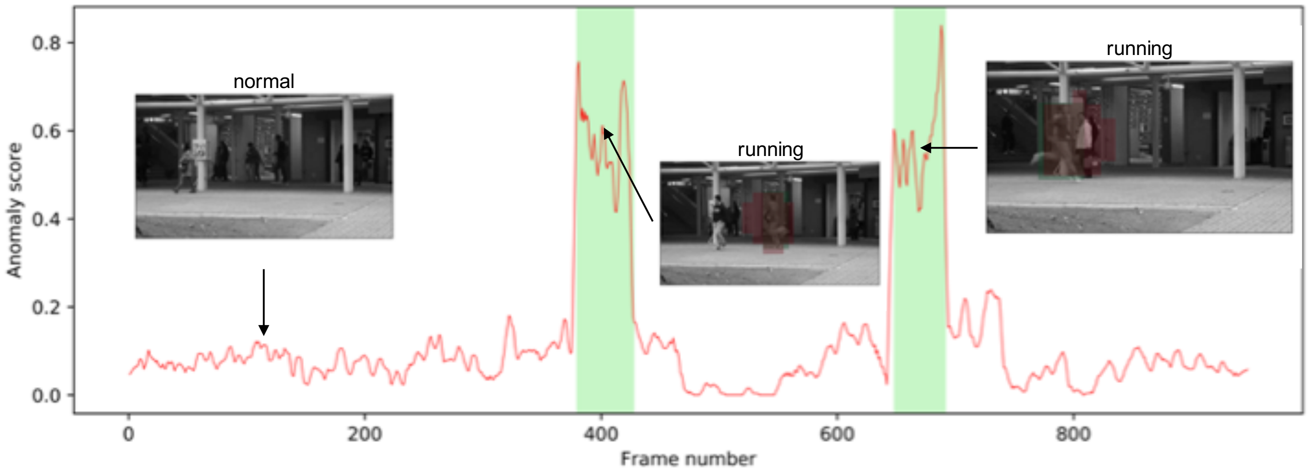

Figure 30. Per-frame anomaly score visualization of CUHK Avenue Test sequence 004.

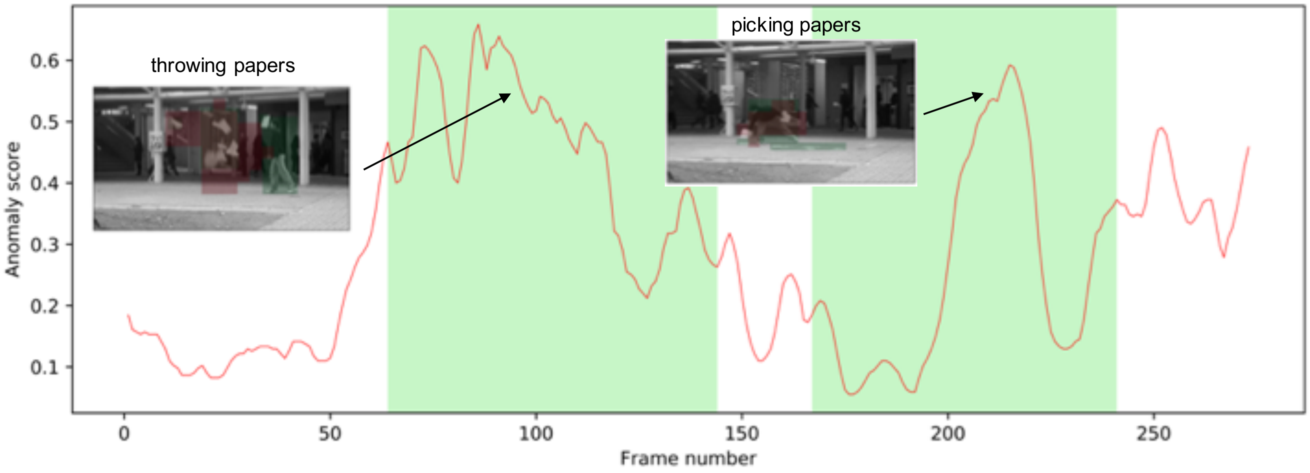

Figure 31. Per-frame anomaly score visualization of CUHK Avenue Test sequence 020.

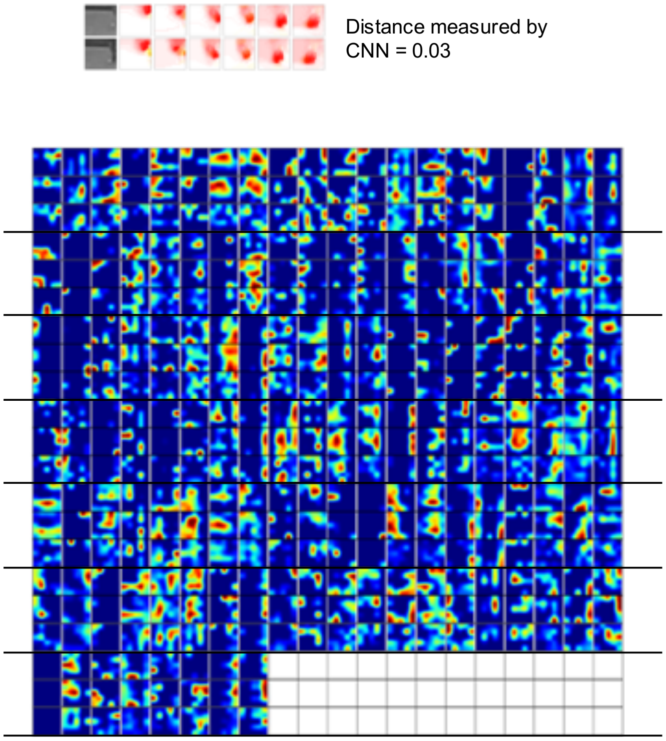

Figure 32. Learned representations and their element-wise difference between 2 video patches in UCSD Ped2, visualized.

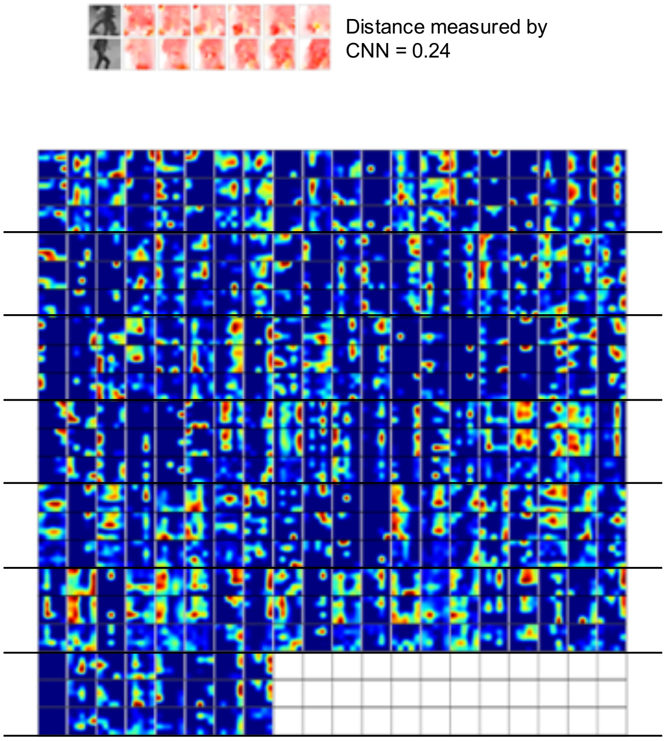

Figure 33. Learned representations and their element-wise difference between 2 video patches in UCSD Ped2, visualized.

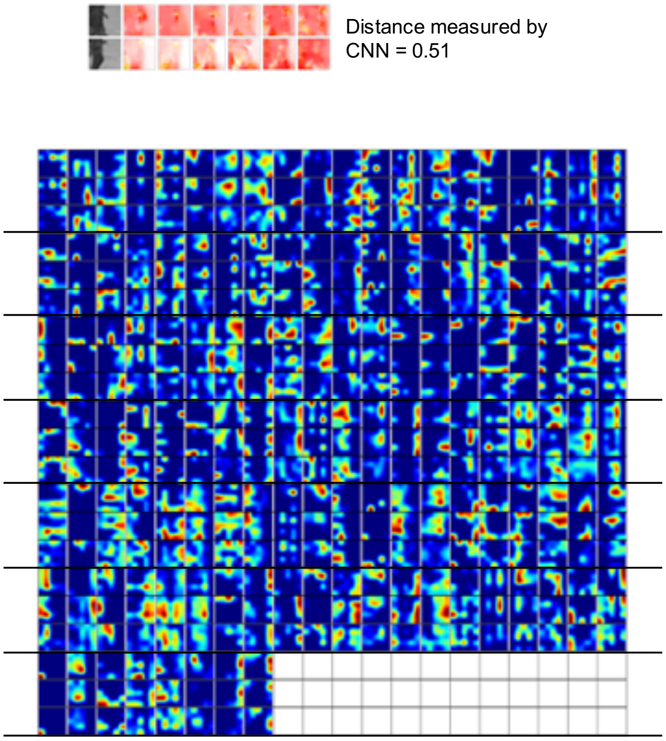

Figure 34. Learned representations and their element-wise difference between 2 video patches in UCSD Ped2, visualized.

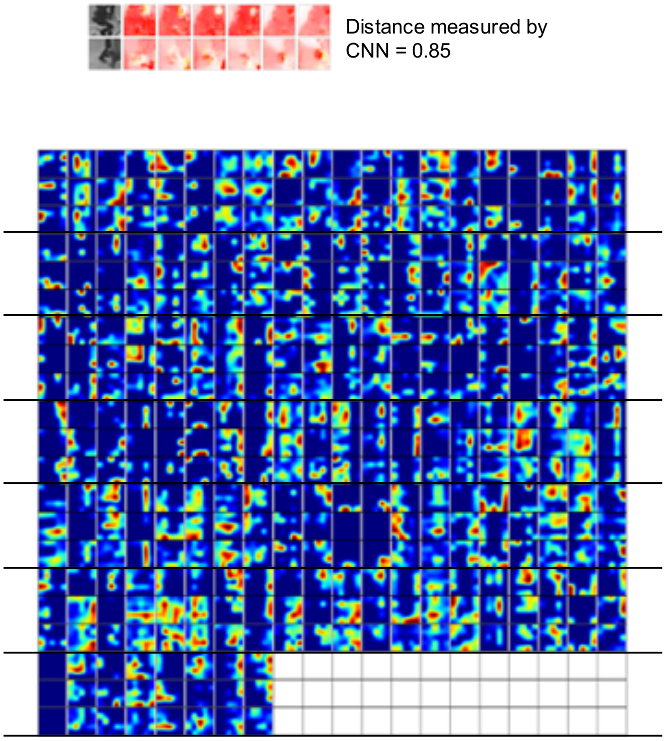

Figure 35. Learned representations and their element-wise difference between 2 video patches in UCSD Ped2, visualized.

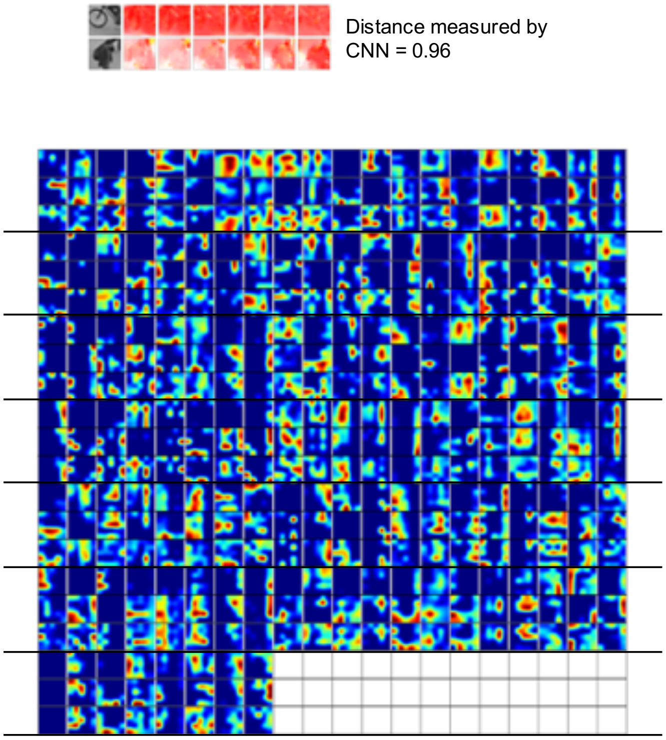

Figure 36. Learned representations and their element-wise difference between 2 video patches in UCSD Ped2, visualized.