





VideoPatchCore: An Effective Method to Memorize Normality for Video Anomaly Detection (*Supplementary Materials*)

Sunghyun Ahn¹, Youngwan Jo¹, Kijung Lee¹, and Sanghyun Park^{1,*}

Department of Computer Science, Yonsei University, Seoul, Republic of Korea
{skd, jyy1551, rlwj4177, sanghyun}@yonsei.ac.kr

A Qualitative Evaluation

A.1 Anomaly Score Visualization

To intuitively understand the use of multiple memories, we compared PatchCore (PC), which utilizes appearance information, with VideoPatchCore (VPC), which incorporates not only appearance information, but also motion and high-level information, on the SHTech and Corridor datasets.

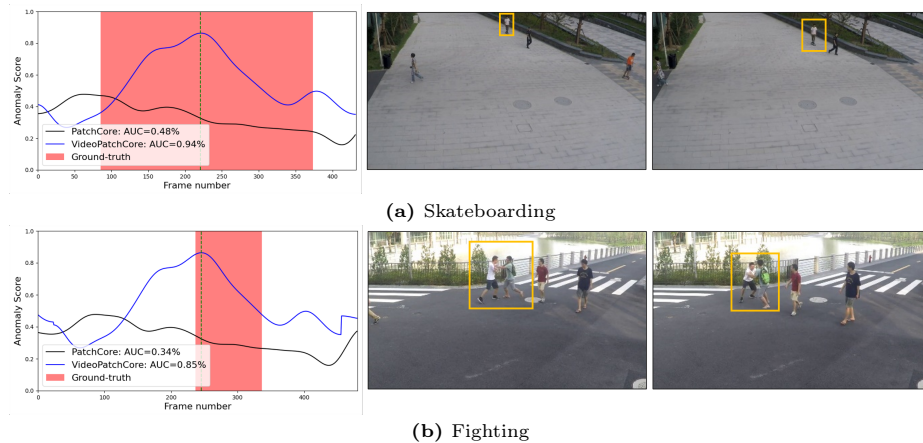


Fig. 1: Comparison of anomaly scores between VPC and PC in the SHTech dataset.

Fig. 1a shows anomaly frames depicting the action of skateboarding, necessitating consideration of motion information. Therefore, temporal memory plays a crucial role in this scenario. Fig. 1b shows anomaly frames depicting the action of fighting, necessitating consideration of interactions between two peoples. Therefore, high-level semantic memory plays a crucial role in this scenario. It

* Corresponding author: Sanghyun Park (sanghyun@yonsei.ac.kr)

is known that VPC, which utilizes temporal and high-level semantic memory, detects anomalies well, whereas PC, which does not utilize them, does not detect anomalies.

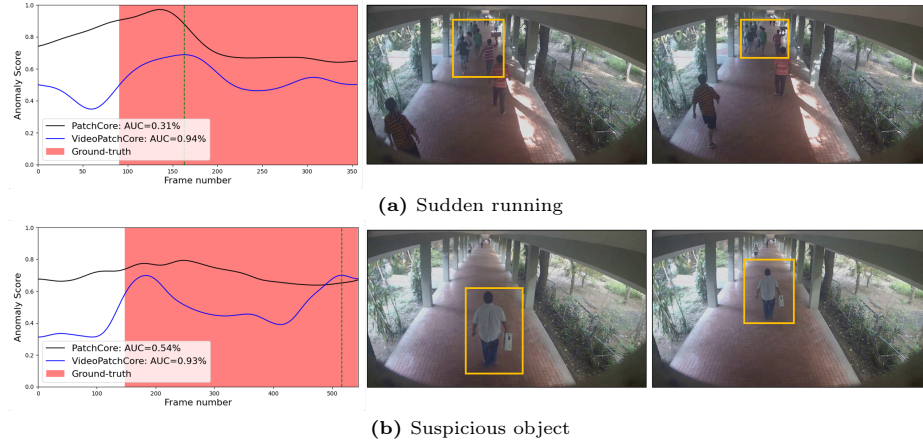


Fig. 2: Comparison of anomaly scores between VPC and PC in the Corridor dataset.

Fig. 2a shows anomaly frames depicting the action of sudden running, necessitating consideration of motion information. Therefore, temporal memory plays a crucial role in this scenario. Fig. 2b shows anomaly frames depicting the action of moving a suspicious object, necessitating consideration of relationship between the person and object. Therefore, high-level semantic memory plays a crucial role in this scenario. VPC, which employs high-level and temporal memory, effectively distinguishes between abnormal and normal frames. In contrast, PC tends to produce false positives.

A.2 Object-wise Anomaly Score Visualization

For a deeper understanding of memory effectiveness, we computed anomaly scores using each memory module for two specific objects shown in Fig. 3. One is the most anomalous object, and the other is the most normal object, as determined by the proposed model. Each object is characterized by the spatial (S) and temporal anomaly scores (T), while frames containing these objects are assigned a high-level anomaly score (H). The experimental results demonstrate precise prediction of anomalous objects within frames by the proposed model, with each memory module effectively fulfilling its role in various scenarios. For instance, in the bicycle scenario involving abnormal appearance, S increases due to spatial memory, while in the running scenario with abnormal behavior, T increases due to temporal memory. Finally, challenging anomalies such as "wrong direction" in the local stream are detected by H increasing in high-level memory. This validates the effectiveness of the three memory modules in VAD.



Fig. 3: Three anomaly scores per object on the Avenue, SHTech, and Corridor datasets. H: high-level anomaly score, S: spatial anomaly score, T: temporal anomaly score.

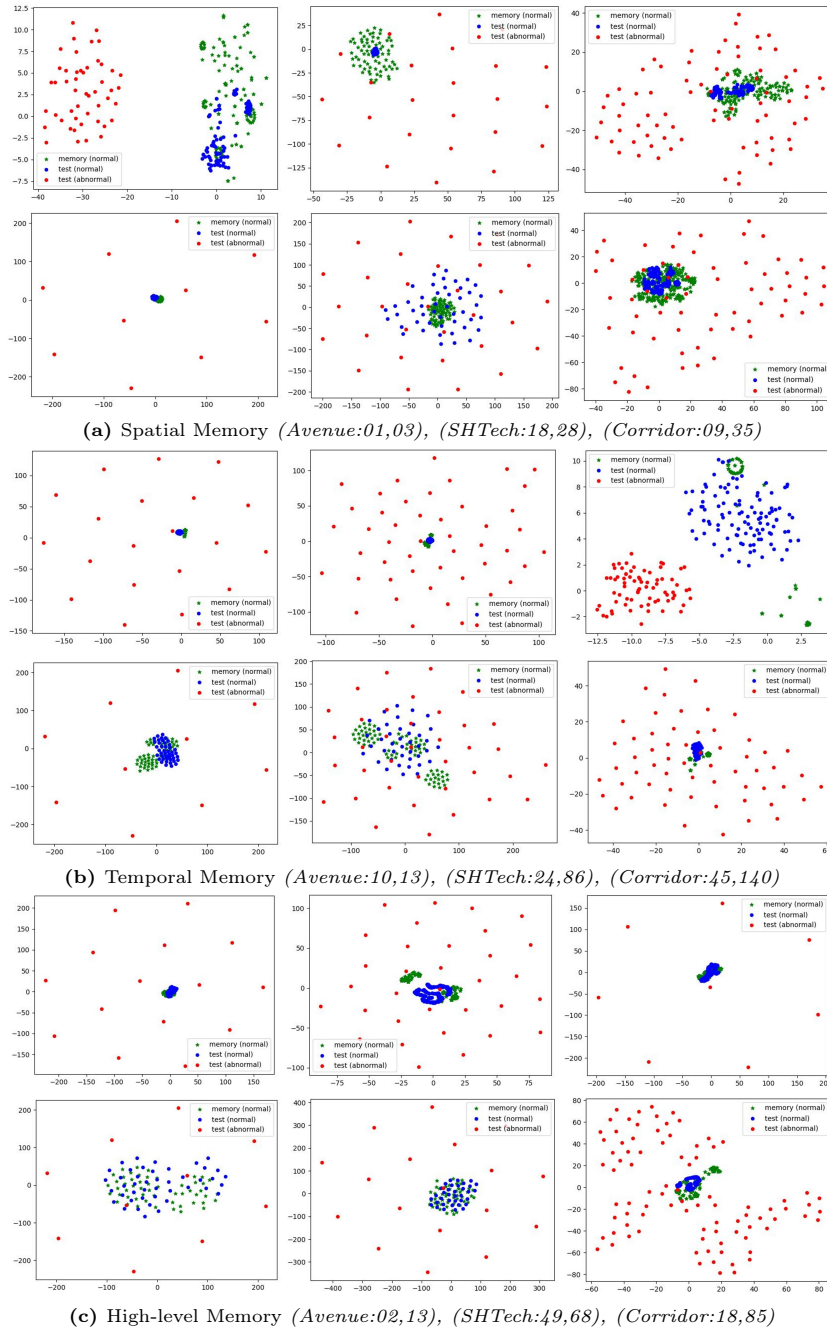


Fig. 4: t-SNE visualization of memory and test patches. The columns, from left to right, correspond to the Avenue, SHTech, and Corridor datasets.

Table 1: Comparison of AUROC scores for all memory banks on the Avenue, SHTech and Corridor datasets. The best results are **red** and the second best results are **blue**.

Avenue	1%	10%	25%	50%	75%	99%
Spatial	0.848	0.831	0.828	0.825	0.828	0.828
Temporal	0.669	0.669	0.669	0.669	0.669	0.669
High-level	0.844	0.845	0.848	0.844	0.844	0.844
Total	0.928	0.918	0.914	0.912	0.912	0.912
SHTech	1%	10%	25%	50%	75%	99%
Spatial	0.748	0.744	0.747	0.747	0.746	0.746
Temporal	0.788	0.788	0.788	0.788	0.788	0.788
High-level	0.671	0.675	0.684	0.673	0.673	0.674
Total	0.846	0.850	0.851	0.851	0.851	0.851
Corridor	1%	10%	25%	50%	75%	99%
Spatial	0.690	0.705	0.705	0.705	0.706	0.705
Temporal	0.735	0.735	0.735	0.735	0.735	0.735
High-level	0.664	0.672	0.673	0.674	0.675	0.660
Total	0.760	0.764	0.763	0.763	0.763	0.764

A.3 Patch Visualization

Fig. 4 depicts the t-SNE plots of normal patches stored in memory (denoted by green 'o'), along with normal (blue 'o') and anomalous patches (red 'o') from the test data. The results show that normal patches are clustered closely together compared to anomalous patches, aligning closely with the distribution of memory patches. In contrast, anomalous patches exhibit a wider dispersion and tend to be farther away from the memory patches. These findings suggest that the proposed memory effectively stores the normalcy of videos, making it suitable for VAD.

B Quantitative Evaluation

B.1 Detailed Analysis of Subsampling Ratio

We conducted a detailed analysis of performance changes based on subsampling ratios in the Avenue, Shanghai, and Corridor datasets as shown in Tab. 1. In the SHTech and Corridor datasets, the performance was better when using more memory, whereas in the Avenue dataset, the performance tended to be higher with less memory.

These results can be explained with two reasons. First, due to the diversity of normal data in the SHTech and Corridor datasets, the performance improves as more normal patches are stored in memory. Second, the Avenue dataset mainly consists of action anomalies, making it challenging to distinguish them from

normal instances without using temporal information. Therefore, from a spatial memory perspective, filtering out normal patches that resemble anomalies enhances the performance. Meanwhile, other memories that utilize temporal information exhibit robust performance regardless of their size. Consequently, using only 1% of memory overall yields the best performance.

However, as evidenced by the experimental results, the performance difference between using 10% and 99% of memory is very small. Therefore, in practical use, sufficiently good performance can be maintained even with memory usage set at 10% or lower.