Temporal-Semantic Context Fusion for Robust Weakly Supervised Video Anomaly Detection

Yuan Zeng, Yuanyuan Wu✉, Jing Liang, and Wu Zeng

College of Computer Science and Cyber Security,Chengdu University of Technology,Sichuan 610059, Chengdu, China
{2021020855,2021020874,zengwu}@stu.cdut.edu.cn, wuyuanyuan@cdut.edu.cn

1 Supplementary Experiments

To assess the influence of batch size and sampling size on the final experimental results, a series of relevant experiments were conducted, using different batch sizes and sampling rates. The experiment results are shown in the table below.

Our empirical observations indicate that, under the same sampling ratio, an enhancement in the batch size corresponds to a more effective anomaly detection. However, due to the constraints of experimental conditions, our maximum sampling is 60. When maintaining a consistent batch size, optimal results are achieved through balanced sampling, where the ratio of positive to negative samples is 1. In contrast, a sample ratio of 0.5 tilts the balance towards an abundance of negative samples, rendering the model prone to overfitting and leading to poor performance. Conversely, when the positive-negative sample ratio reaches 1.5, a scarcity of negative samples emerges, resulting in a decline in the anomaly detection performance.

Based on the comprehensive experiment results, we found that the best effect is achieved when the batch size is 60 and balanced sampling is used.

Table 1. AUC(%) performance comparison of batch size and sampling ration.

Batch size \ Sample ratio	0.5	1	1.5
60	96.19	96.41	95.96
32	95.77	96.12	95.40
16	95.24	95.46	95.20
8	94.85	95.35	94.96