저조도 환경 감시 영상에서 시공간 패치 프레임을 이용한 이상행동 분류

Widia A. Samosir[1], 공성곤 [1]
[1] 세종대학교 컴퓨터공학과

23110148@sju.ac.kr, skong@sejong.edu

Spatiotemporal Patched Frames for Human Abnormal Behavior Classification in Low-Light Environment

Widia A. Samosir[1], Seong G. Kong[1]
[1]Dept. of Computer Science and Engineering, Sejong University

ABSTRACT

Surveillance systems play a pivotal role in ensuring the safety and security of various environments, including public spaces, critical infrastructure, and private properties. However, detecting abnormal human behavior in low-light conditions is a critical yet challenging task due to the inherent limitations of visual data acquisition in such scenarios. This paper introduces a spatiotemporal framework designed to address the unique challenges posed by low-light environments, enhancing the accuracy and efficiency of human abnormality detection in surveillance camera systems. We proposed the pre-processing using lightweight exposure correction, patched frames pose estimation, and optical flow to extract the human behavior flow through t-seconds of frames. After that, we train the estimated-action-flow into autoencoder for abnormal behavior classification to get normal loss as metrics decision for normal/abnormal behavior.

1. Introduction

In the realm of video surveillance and security, the accurate classification of human behaviors in challenging environments is of paramount importance. Surveillance systems are tasked with monitoring and safeguarding public spaces, critical infrastructure, and commercial establishments, where low-light conditions often prevail. Recognizing and categorizing abnormal human behaviors under such circumstances is inherently complex due to the reduced visibility and quality of surveillance footage.

This paper introduces a novel approach, the "Spatiotemporal with Patched Frames Framework for Human Abnormal Behavior Classification in Low-Light Environment Surveillance Video". Our research is driven by the imperative need to develop robust methods for enhancing the capabilities of surveillance systems in low-light conditions. Abnormal behavior detection holds significant implications for public safety, crime prevention, and incident response, making it an area of critical concern.

The proposed spatiotemporal framework aims to bridge the gap in the surveillance domain by leveraging cutting-edge techniques in computer vision, deep learning, and spatiotemporal modeling. By integrating these elements, we endeavor to provide a comprehensive solution for the accurate identification and classification of abnormal human behaviors, even in challenging low-light scenarios.

2. Related Works

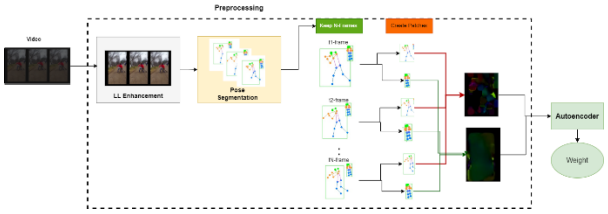
In this paper [1] Yang et. al. proposed unified classification and prediction fusion framework designed to detect various types of abnormal events in surveillance videos. By leveraging the fusion of normality scores between classification streams using pose with ST-GCN and prediction streams using U-Net optical flow, they could achieve best AUC performance compared with stated-of-the-art on UCSD, Ped2 and SHTC datasets.

However, video anomaly detection needs to be versatile in terms of illumination or light environment to able to get the best results for all-time monitoring. Based on that we propose a method for enabling not only the spatial but temporal understanding of human abnormal behavior in terms of sequences of frames. This paper also focuses on increasing the ability of the model to classify the flow of motion in a low-light environment of surveillance video.

3. Proposed Method

3.1. Low-light Estimated Flow Pre-processing

In the realm of computer vision and video analytics, addressing the challenges posed by low-light environments has been a long-standing concern. In this context, one critical aspect is the preprocessing of low-light video data to enhance its quality and facilitate accurate abnormal behavior classification. We are using the Illumination Adaptive Transformer (IAT) [2] that is specifically designed for image enhancement and exposure correction in low-light conditions. By incorporating this transformative technique along with patching method into our preprocessing pipeline in Fig. 1, we aim to significantly improve the quality of low-light video streams, thereby paving the way for more accurate and robust abnormal behavior classification systems.



(Figure 1) Pre-processing and Classification Flow for Our Proposed Method

3.2. Human Abnormal Behavior Classification

We chose autoencoder with skip connection as a classifier for abnormal behavior because this method proved for better performance compared to traditional autoencoder [5][4]. Traditional autoencoders may lose fine spatial details during the encoding and decoding process. Meanwhile, the skip connections in the autoencoder concatenate feature maps from the encoder to the corresponding layers in the decoder, allowing the model to access higher-level features and better preserve spatial information during up-sampling.

Detecting abnormal behavior through the patch cropping to feed to autoencoder also is a novel approach that we aim to explore in our research. By leveraging the concept of patch cropping, we intend to enhance the effectiveness of anomaly detection in various contexts. This technique involves the aggregation or consolidation of localized image or video patches to provide a more comprehensive view of the scene, enabling us to capture both local and global information within an image, which can be particularly useful when dealing with complex scenes or situations where abnormalities may occur at different scales.

In our inference or testing, we need to make quantitative measurements to evaluate our model. We are using PSNR (Peak Signal to Noise Ratio at frame level shown in 1.

$$P_t = \frac{\left[M_{\hat{I}_t}\right]^2}{\frac{1}{R}\left\|\hat{I}_t - I_t\right\|_F^2} \quad (1)$$

We also do optimal threshold experiment using Youden's J Statistic that required TPR (True Positive Rate) and FPR (False Positive Rate) from ROC calculation. The statistic calculated as shown below:

$$J = Sensitivity + Specificity - 1 \quad (2)$$

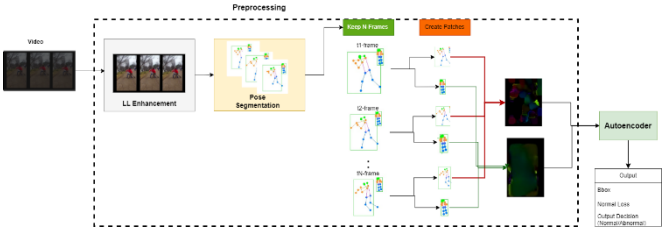
$$Sensitivity = TPR$$
$$Specificity = 1 - FPR \quad (3)$$

$$J = TPR - FPR \quad (4)$$

where J is the optimal threshold that is calculated from the sensitivity and specificity in 4.

3.3. Inference in Low-light Environment Condition

During the inference phase in Fig. 2, we replicate the preprocessing steps employed during training by patching the bounding box. Following this preprocessing, we employ a threshold-based decision-making process to classify instances as either abnormal or normal along with the bounding box of human action.



(Figure 2) Inference process of Our proposed Method

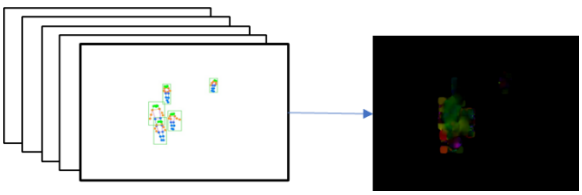
4. Results

4.1. Dataset

In this paper, we are using the Anomaly Detection UCF Dataset [5] as our primary source of video data. This dataset features a diverse range of video clips, encompassing various abnormal behaviors, including instances of abuse, arrests, arson incidents, and assaults. In addition to these anomalous scenarios, the dataset also includes a substantial collection of normal video clips, providing a comprehensive and balanced representation of both typical and atypical activities for our research and analysis. This rich dataset with its diverse content allowed us to conduct a thorough investigation into anomaly detection methods and their effectiveness in distinguishing between normal and abnormal behaviors within video streams.

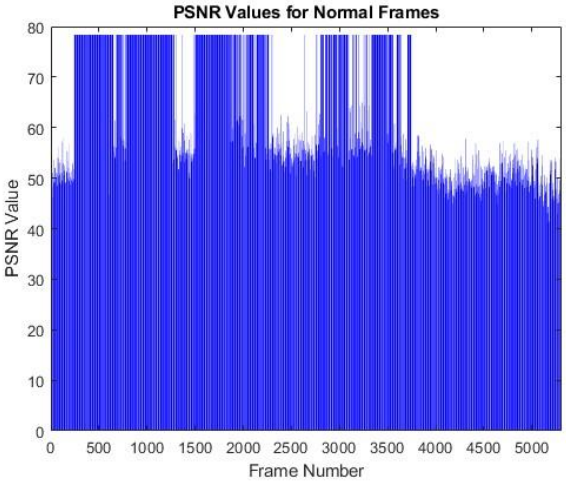
4.2. Preliminary Result

In our methodology, we first perform preprocessing using MMPose to selectively extract the human pose component. Figure 3 illustrates the outcomes of this pose estimation process. Subsequently, we employ Lucas-Kanade optical flow estimation on the temporally extracted human pose data, which allows us to capture the motion flow within the sequential frames, as depicted in Figure 3.

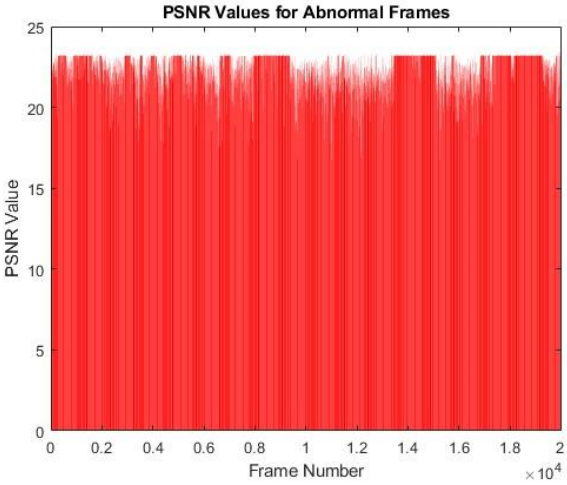


(Figure 3) Extracting pose and motion flow of sequence of frames

To effectively discern anomalies from normal human behavioral patterns, we employ an Autoencoder for training our dataset. Notably, for this preliminary outcome in Fig. 4 and Fig.5, we refrain from implementing the cropping patch component in the input fed into the Autoencoder. This approach allows us to showcase the initial capabilities of our method in distinguishing between normal and abnormal human behavior flows.



(Figure 4) Preliminary Experiment Result for Normal Behavior Frames



(Figure 5) Preliminary Experiment Result for Abnormal Behavior Frames

After we calculate the threshold using the Youden's J statistic, we get the best threshold of PSNR value equal to 24dB. The higher PSNR means the lower anomaly score of frame sequences and the lower PSNR means the higher anomaly score of frame sequences. We get the accuracy around 100% using comparison between total true positives to the length of frame sequences in video level.

5. Conclusions and Future Work

Our ongoing challenge and focus for future research efforts revolve around improving the accuracy of classifying human abnormal behavior using an unsupervised learning model. This challenge is compounded by the need to incorporate multiple pre-processing steps into the pipeline. Furthermore, optimizing real-time processing of classification within video streams poses a particularly complex task, given the requirement to minimize latency. In results we got high accuracy in terms of video-level evaluation. However, we want to precise the accuracy in frame level to get the exact number of anomaly score in real-time video data streams.

In our future work, we intend to address these dual objectives: enhancing the accuracy of classification while concurrently streamlining the real-time processing pipeline for the Spatiotemporal with Patched Frames Framework for Human Abnormal Behavior Classification in Low-Light Environment Surveillance Video.

ACKNOWLEDGEMENT

This work was supported in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant funded by the Korean Government (MSIT) under Grant 2019-0-00231, and in part by the Development of artificial Intelligence-Based Video Security Technology and Systems for Public Infrastructure Safety.

References

[1] Yang, Y., Fu, Z., & Naqvi, S. M. (2023). Abnormal event detection for video surveillance using an enhanced two-stream fusion method. *Neurocomputing*, *553*, 126561.

[2] Cui, Z., Li, K., Gu, L., Su, S., Gao, P., Jiang, Z., ... & Harada, T. (2022, November). You Only Need 90K Parameters to Adapt Light: a Lightweight Transformer for Image Enhancement and Exposure Correction. In *BMVC* (p. 238).

[3] Collin, A. S., & De Vleeschouwer, C. (2021, January). Improved anomaly detection by training an autoencoder with skip connections on images corrupted with stain-shaped noise. In *2020 25th International Conference on Pattern Recognition (ICPR)* (pp. 7915-7922). IEEE.

[4] Yan, H., Liu, Z., Chen, J., Feng, Y., & Wang, J. (2023). Memory-augmented skip-connected autoencoder for unsupervised anomaly detection of rocket engines with multi-source fusion. *ISA transactions*, *133*, 53-65.

[5] Sultani, W., Chen, C., & Shah, M. (2018). Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6479-6488).