VUEBLOX: A Semi-Supervised Hybrid Deep Learning Framework for Occlusion-Aware and Interpretable Crowd Anomaly Detection

Vaichaly Varatharajan
Department of Computing
Informatics Institute of Technology
Colombo, Sri Lanka
vaichaly.20210459@iit.ac.lk

Saadh Jawwadh
Department of Computing
Informatics Institute of Technology
Colombo, Sri Lanka
saadh.j@iit.ac.lk

***Keywords*—Crowd anomaly detection, Explainable AI, Occlusion handling, Semi-supervised learning, Temporal-spatial attention**

I. Introduction

Crowd tracking using computer vision technologies enhances public safety, but detecting crowd anomalies remains challenging due to issues like occlusion and interpretability models' low accuracy [1]. Also, those are irregularly happening constitute a small percentage, so the datasets have lots of missing labeled outliers.This study address these gaps by proposing a semi-supervised hybrid architecture with dedicated modules for occlusion and interpretability.

II. Literature Review

Over time, research on crowd anomaly detection has changed dramatically. As deep learning has advanced, methods that use variational autoencoders [2] and convolutional neural networks have shown better results. Though [3] emphasizes the significance of part-based models and attention processes for managing visual obstructions in pedestrian detection.Semi-supervised learning approaches have emerged to address data scarcity challenges, with methods like AnoRand demonstrating how random labeling strategies can improve anomaly detection performance.

A. Taxonomy of crowd anomaly detection techniques

As shown in Figure 1 the author has presented two primary methods for detecting anomalies in crowds: frame extraction-based anomaly detection and crowd density-driven anomaly identification.

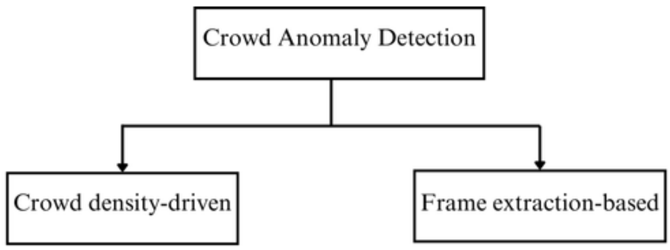

Fig. 1. Taxonomy of crowd anomaly detection techniques

III. Methodology

System architecture has a pipeline structure and consists six components, which are inter-linked with each other.

A. Input Video Acquisition

The process of the pipeline is initiated by the acquisition of video data as the only input source for the system.

B. Preprocessing

Frame extraction, normalization, noise reduction, and background subtraction are the primary preprocessing operation [4]. As a result, the video is transformed from raw data to standardized frames.

C. Occlusion Handling

This module implements an architecture based on a transformer, processing of patches, as well as the application of a GAN inpainting system, to recover the areas which are occluded in crowded scenes. Spatial relationships are modeled by Graph Neural Networks [5].

D. Feature Extraction

The characteristics concerning the video content consists of four techniques applied. Specifically, SimCLR employs self-supervised contrastive learning [6]. The sparse tracking techniques are selected based on the movement patterns. Graph Neural Networks depicts the interactions, and Long-Term Memory networks handle the sequence of frame features to unveil unusual patterns.

https://doi.org/10.31705/ADScAI.2025.32

E. Anomaly Detection

It uses hybrid techniques, which are ensemble voting and threshold-based scoring strategies to hunt down the abnormal occurrences with high precision. This method utilizes multiple detection models, weighted voting, and dynamic confidence score adjustment to ensure that the detection is robust to environment.

F. Interpretability

Interpritation includes Grad-CAM attention maps, graph-based explanations, and prototype-based anomaly association. Those technologies give graphic and text-based demonstrations, outlining localities influencing anomaly result [7].

IV. Results

Author applied the model to evaluate precision, F1-score, AUC, and occlusion robustness.Occlusion-handling module trained on the MOT dataset , and it lasted 7 hours due to limited computational resources.The full end-to-end system needed 24 hours on a mid-range GPU.Without interpretability modules 20 FPS can go high.

TABLE I
SOTA approaches comparison in terms of computational time on UCSD Ped2 dataset

Approaches	**Computational time**
150 FPS [8]	27FPS
SF [9]	25FPS
AMDN [10]	25FPS
sRNN-AE [11]	25FPS
Our method	**20FPS**

TABLE II
Performance Comparison of Crowd Anomaly Detection Methods

Method	**Precision**	**F1-score**	**AUC**	**Occlusion Robust**
Transfer Learning [12]	0.731	0.671	0.798	Low
Semantic Segmentation [13]	0.924	0.238	0.884	Medium
Graph-based Detection [14]	0.941	0.256	0.902	Medium
Hybrid Pixel-Frame [15]	0.957	0.243	0.894	Low
Ours	**0.990**	**0.269**	**0.929**	**High**

V. Conclusion

VUEBLOX serves as a semi-supervised framework to detect crowd anomalies through explainability and occlusion. While the system producing promising outcomes the framework still poses significant challenges because of limited computational resources and shows dataset biases. Rich explanations and enhanced complex occlusion management result in a reduction of FPS. System will focus on merging various behavioral signals while enhancing edge-based systems and improving performance across data domains.

References

[1] Y. Zhang, H. Chen, W. Bao, Z. Lai, Z. Zhang, and D. Yuan, "Handling Heavy Occlusion in Dense Crowd Tracking by Focusing on the Heads," Oct. 2023, arXiv:2304.07705 [cs]. [Online]. Available: http://arxiv.org/abs/2304.07705

[2] Y. Liu, D. Yang, Y. Wang, J. Liu, J. Liu, A. Boukerche, P. Sun, and L. Song, "Generalized Video Anomaly Event Detection: Systematic Taxonomy and Comparison of Deep Models," *ACM Computing Surveys*, vol. 56, no. 7, pp. 1–38, Jul. 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3645101

[3] F. Li, X. Li, Q. Liu, and Z. Li, "Occlusion Handling and Multi-Scale Pedestrian Detection Based on Deep Learning: A Review," *IEEE Access*, vol. 10, pp. 19 937–19 957, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9718221/

[4] M. Zhang, J. Wang, Q. Qi, Z. Zhuang, H. Sun, and J. Liao, "Cognition Guided Video Anomaly Detection Framework for Surveillance Services," *IEEE Transactions on Services Computing*, vol. 17, no. 5, pp. 2109–2123, Sep. 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10542401/

[5] G. Almahadin, M. Subburaj, M. Hiari, S. Sathasivam Singaram, B. P. Kolla, P. Dadheech, A. D. Vibhute, and S. Sengan, "Enhancing Video Anomaly Detection Using Spatio-Temporal Autoencoders and Convolutional LSTM Networks," *SN Computer Science*, vol. 5, no. 1, p. 190, Jan. 2024. [Online]. Available: https://doi.org/10.1007/s42979-023-02542-1

[6] M. Abdalla, S. Javed, M. A. Radi, A. Ulhaq, and N. Werghi, "Video Anomaly Detection in 10 Years: A Survey and Outlook," Jul. 2024, arXiv:2405.19387. [Online]. Available: http://arxiv.org/abs/2405.19387

[7] R. Srivastava and G. Chhabra, "Anomaly Detection Approach for Human Detection in Crowd Based Locations," in *2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET)*, Sep. 2023, pp. 150–157. [Online]. Available: https://ieeexplore.ieee.org/document/10303413

[8] C. Lu, J. Shi, and J. Jia, "Abnormal Event Detection at 150 FPS in MATLAB."

[9] V. Reddy, C. Sanderson, and B. C. Lovell, "Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture," in *CVPR 2011 WORKSHOPS*. Colorado Springs, CO, USA: IEEE, Jun. 2011, pp. 55–61. [Online]. Available: http://ieeexplore.ieee.org/document/5981799/

[10] D. Xu, E. Ricci, Y. Yan, J. Song, and N. Sebe, "Learning Deep Representations of Appearance and Motion for Anomalous Event Detection," Oct. 2015, arXiv:1510.01553 [cs]. [Online]. Available: http://arxiv.org/abs/1510.01553

[11] W. Luo, W. Liu, D. Lian, J. Tang, L. Duan, X. Peng, and S. Gao, "Video Anomaly Detection with Sparse Coding Inspired Deep Neural Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 3, pp. 1070–1084, Mar. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/8851288/

[12] P. Liyanage and P. Fernando, "Suspicious Human Crowd Behaviour Detection - A Transfer Learning Approach," in *2021 21st International Conference on Advances in ICT for Emerging Regions (ICter)*. Colombo, Sri Lanka: IEEE, Dec. 2021, pp. 63–68. [Online]. Available: https://ieeexplore.ieee.org/document/9774784/

[13] F. Abdullah and A. Jalal, "Semantic Segmentation Based Crowd Tracking and Anomaly Detection via Neuro-fuzzy Classifier in Smart Surveillance System," *Arabian Journal for Science and Engineering*, vol. 48, no. 2, pp. 2173–2190, Feb. 2023. [Online]. Available: https://link.springer.com/10.1007/s13369-022-07092-x

[14] M. Yang, Y. Feng, A. S. Rao, S. Rajasegarar, S. Tian, and Z. Zhou, "Evolving graph-based video crowd anomaly detection," *The Visual Computer*, vol. 40, no. 1, pp. 303–318, Jan. 2024. [Online]. Available: https://link.springer.com/10.1007/s00371-023-02783-4

[15] H. Tutar, A. Güneş, M. Zontul, and Z. Aslan, "A Hybrid Approach to Improve the Video Anomaly Detection Performance of Pixel- and Frame-Based Techniques Using Machine Learning Algorithms," *Computation*, vol. 12, no. 2, p. 19, Jan. 2024. [Online]. Available: https://www.mdpi.com/2079-3197/12/2/19