ALFred: An Active Learning Framework for Real-world Semi-supervised Anomaly Detection with Adaptive Thresholds

Shanle Yao
syao@charlotte.edu
Ghazal Alinezhad Noghre
galinezh@charlotte.edu
Armin Danesh Pazho
adaneshp@charlotte.edu
Hamed Tabkhi
htabkhiv@charlotte.edu

Department of Electrical and Computer Engineering
University of North Carolina at Charlotte
Charlotte, USA

Abstract

Video Anomaly Detection (VAD) can play a key role in spotting unusual activities in video footage. VAD is difficult to use in real-world settings due to the dynamic nature of human actions, environmental variations, and domain shifts. Traditional evaluation metrics often prove inadequate for such scenarios, as they rely on static assumptions and fall short of identifying a threshold that distinguishes normal from anomalous behavior in dynamic settings. To address this, we introduce an active learning framework tailored for VAD, designed for adapting to the ever-changing real-world conditions. Our approach leverages active learning to continuously select the most informative data points for labeling, thereby enhancing model adaptability. A critical innovation is the incorporation of a human-in-the-loop mechanism, which enables the identification of actual normal and anomalous instances from pseudo-labeling results generated by AI. This collected data allows the framework to define an adaptive threshold tailored to different environments, ensuring that the system remains effective as the definition of 'normal' shifts across various settings. Implemented within a lab-based framework that simulates real-world conditions, our approach allows rigorous testing and refinement of VAD algorithms with a new metric. Experimental results show that our method achieves an EBI (Error Balance Index) of 68.91 for Q3 in real-world simulated scenarios, demonstrating its practical effectiveness and significantly enhancing the applicability of VAD in dynamic environments.

1 Introduction

Video Anomaly Detection (VAD) with semi-supervised learning is a crucial domain of computer vision focused on identifying unusual behavior in video streams, with applications ranging from surveillance to healthcare [3]. While traditional pixel-based methods

[12, 22, 45] often raise privacy and bias concerns, especially for marginalized groups, pose-based approaches offer a privacy-preserving and more equitable alternative by relying on skeletal motion data. However, deploying these models outside controlled environments remains a major challenge, requiring innovations to ensure real-world robustness.

The real world demands that VAD systems contend with a host of practical difficulties. Anomaly definition is inherently relative and context-dependent. For instance, running in a library would be flagged as unusual, whereas jogging in a park is perfectly normal—necessitating systems to adapt their understanding of "normal" and "anomalous" across diverse environments. Real-time detection imposes tight constraints on inference time, requiring models to process streaming video data quickly for timely alerts, yet the computational demands of algorithms often outstrip available resources. Defining clear anomaly boundaries further complicates matters, as many VAD algorithms produce anomaly scores that are not inherently percentages [18, 26, 30], complicating the establishment of clear decision boundaries. Moreover, standard metrics like AUC-ROC and AUC-PR rely on labeled test data, which is often unavailable in dynamic, real-world settings. Static, offline-trained models struggle to generalize, leading to frequent false positives.

To confront these challenges head-on, we introduce ALFred (Active Learning Framework for Real-world video anomaly detection Deployment), a novel approach tailored for pose-based VAD. Unlike conventional methods that rely on static training or passive online updates, ALFred leverages active learning to selectively query the most informative data points for labeling, optimizing efficiency and precision. This is paired with a human-in-the-loop mechanism, where human feedback refines the system's understanding of context-specific anomalies, supported by AI-generated pseudo-labels. To enhance decision-making, ALFred incorporates a dynamic threshold adaptation strategy, adjusting detection boundaries based on historical labeled context and feedback — critical for handling non-percentage outputs and ensuring responsiveness in dynamic settings.

To rigorously evaluate ALFred's reliability in real-world scenarios, we introduce the Error Balance Index (EBI), a novel metric that quantifies the trade-off between False Positive Rate (FPR) and False Negative Rate (FNR). Unlike traditional evaluation metrics that often overlook the delicate balance between misclassification types, EBI provides a more robust measure of a system's practical viability in dynamic environments. We assessed ALFred in a lab-based simulation designed to emulate real-world conditions, including continual data and domain shifts. Building on established pose-based models suitable for online learning, such as STG-NF [18], TSGAD [30], and SPARTA [29]. ALFred achieved an EBI of 68.91 for Q3, outperforming other methodologies. This success highlights its ability to deliver fast, privacy-preserving, and adaptable anomaly detection. Beyond performance, ALFred integrates contextual adaptability into real-world evaluation, paving the way for more reliable, deployment-ready VAD frameworks. Future improvements such as optimizing inference speed and expanding dataset diversity promise to further enhance its practical utility.

In summary, the contributions of this paper are: (1) Introducing ALFred, an active learning framework, applicable to existing models, for real-world VAD with adaptive thresholds; (2) Bringing human-in-the-loop functionality to introduce reliable training and validation sets for active learning; (3) Conducting extensive experiments that showcase the benefits of ALFred in adapting to unseen data.

2 Related Works

Historically, traditional VAD methods relied on handcrafted features [8, 9, 23, 36, 42]. These techniques were effective in controlled settings but struggled with generalization to real-world applications due to their reliance on predefined feature sets. The advent of deep learning revolutionized VAD, enabling the development of two primary approaches: pixel-based and pose-based. Pixel-based techniques [1, 5, 6, 14, 15, 16, 20, 24, 37, 43] analyze raw video data to detect irregularities, but their dependence on visual content raises privacy concerns and makes them susceptible to to environmental variations such as lighting changes and occlusions. In contrast, pose-based approaches extract skeletal movement data, offering a privacy-preserving alternative that is robust to background noise.

Semi-supervised learning strategies for pose-based VAD typically train models to internalize normal behavior by solving self-supervised tasks. These include reconstructing the current frame [4, 7, 17, 21, 26, 41] or predicting past or future motion sequences [19, 35, 44]. Several studies [27, 28] have adopted multi-branch frameworks that integrate both objectives, demonstrating anomaly detection performance.

Among publicly available pose-based models, GEPC, STG-NF, SPARTA and TSGAD have emerged as prominent baselines. GEPC [26] employs a graph-based clustering approach where pose sequences are mapped into a latent space, with anomalies detected based on deviations from the learned distribution. STG-NF [18] uses normalizing flows to transform input sequences into a standard distribution, flagging deviations as anomalies. TSGAD [30] extends pose-based detection by incorporating trajectory analysis through a graph variational autoencoder, where discrepancies between predicted and actual trajectories contribute to the anomaly score. SPARTA [29, 32] leverages the power of transformers for both reconstruction and prediction in order to teach the model to learn normal data.

Despite the promise of these models, a major limitation is their reliance on static training paradigms. While some integrate online processing, they lack mechanisms for continuous adaptation to evolving behaviors in dynamic environments. Additionally, these models generally assume fixed thresholds for anomaly detection, which may not generalize well across different contexts. Addressing these issues requires incorporating active learning and human feedback to refine anomaly classification dynamically.

The concept of 'Online Anomaly Detection' varies across studies [11, 12, 13]. Many pixel-based approaches [22, 38, 45] define it as the ability to make real-time decisions without retraining, whereas our work focuses on continuous online learning, where models incrementally update their parameters using streaming data. Recent work [40] underscores the challenges of deploying VAD models in dynamic environments, highlighting the need for adaptive methods that can handle evolving behaviors and environmental shifts. This distinction is crucial, as real-world anomaly detection systems must learn and adapt continuously, rather than relying on fixed decision boundaries set during initial training. Several studies [10, 13, 39] have explored cross-domain evaluation, where models are tested on unseen datasets without prior exposure. However, these works do not propose explicit domain adaptation strategies, particularly in streaming settings where data distribution shifts over time. Effective deployment of VAD systems demands real-time domain adaptation mechanisms that can recalibrate models in response to new environmental conditions.

Overall, existing methods lack adaptive selection of informative samples and efficient human-in-the-loop refinement. Our proposed ALFred framework bridges this gap by integrating active learning with AI-generated pseudo-labels and dynamic thresholding, enabling continuous adaptation and reducing reliance on extensive manual annotations.

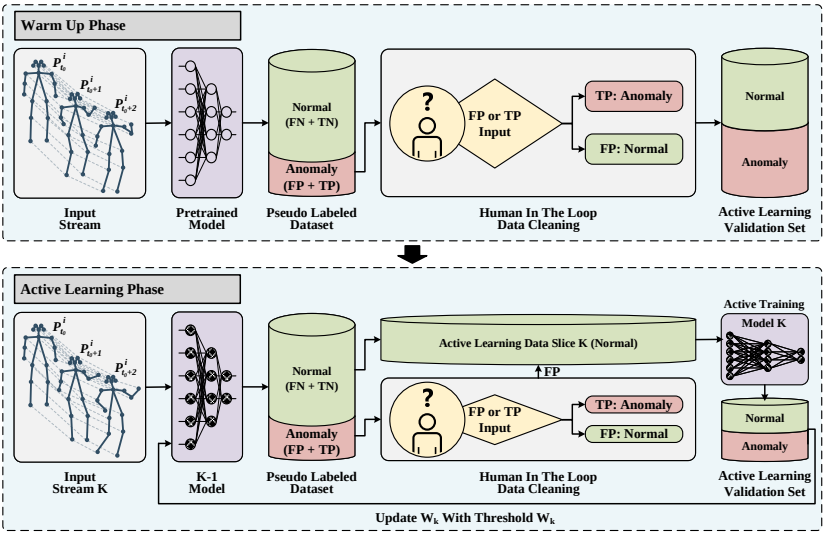

Figure 1: ALFred is a two-phase active learning framework for anomaly detection. The Warm-Up Phase utilizes a pretrained model to generate pseudo-labeled data, which is refined through human-in-the-loop validation before forming the active learning validation set. The Active Learning Phase iteratively improves the model by incorporating human feedback on false positives and true positives, updating the model threshold, and performing active training to enhance anomaly detection accuracy.

3 Methodology

3.1 ALFred

Deploying VAD models in the real world requires more than high performance on standard benchmarks, it demands adaptability to unseen domains and precise threshold calibration, including frame preprocessing. ALFred addresses this challenge through a two-phase strategy: (1) a warm-up phase that collects a representative validation set from the target domain, and (2) an active learning phase that fine-tunes the model and threshold iteratively. The details are provided in Figure 1.

In pose-based VAD, advanced algorithms such as STG-NF [18], GEPC [26], and TS-GAD [30] are frequently evaluated using metrics like AUC-ROC and AUC-PR. While these metrics offer a robust evaluation of model performance across all possible thresholds, they do not directly provide a specific threshold suitable for real-world deployment. In practical settings, selecting an operational threshold is critical, as it governs the trade-off between false positive rate (FPR) and false negative rate (FNR), which varies depending on application-specific requirements. Additionally, the anomaly scores produced by VAD models are often not calibrated probabilities, complicating their direct interpretation as confidence measures.

To address threshold selection, equal error rate (EER) is defined as the point where FPR equals FNR, with more details in Section 3.3. The EER is both model-specific and dependent on the training data. In a new environment, domain shifts can alter the underlying distribution of normal and anomalous events, rendering the original EER threshold ineffective. Thus, adapting the model and its threshold to the target domain is important in real-world scenarios.

To establish an EER threshold specific to the target domain, labeled data from that domain is essential. This labeled data can serve as a validation set for the online-trained model, enabling the computation of an EER threshold tailored to the model's specific weights after training. To address this need efficiently, we introduce ALFred, an active learning framework designed to enhance the deployment of VAD models in real-world applications, consisting of two phases, including a warm-up phase that collects the validation set from the target domain and an active learning phase that deploys a training strategy for the algorithm.

Warm Up Phase: As illustrated in Figure 1, the warm-up phase of ALFred leverages a pre-trained model, initially calibrated with an EER threshold derived from the source domain, and deploys it in the target domain to assign pseudo-labels to the input skeleton data. These pseudo-labels classify segments of the data as normal or anomalous according to the threshold of the source domain. To address potential domain shift, human annotators review the pseudo-labeled anomalous data, correcting misclassifications and ensuring the labels accurately reflect the characteristics of the target domain. This human-in-the-loop approach results in a well-labeled active learning validation set, consisting of a balanced 50:50 ratio of normal to anomalous samples. This validation set is then utilized in the subsequent active learning phase to facilitate the model's adaptation to the new environment.

Active Learning Phase: In the active learning phase, the framework adapts the model to the target domain using the validation set curated during the warm-up phase. The pseudo-labeled normal data from the incoming stream serves as the training set K, leveraging the real-world prevalence of normal instances over anomalies. Although some data points may be misclassified as FNs, they primarily act as noise, which can help the model learn more robust representations of normal behavior. Pseudo-labeled anomalies, on the other hand, are reviewed by human annotators, whose role is critical in identifying TPs and FPs. This human-in-the-loop process ensures high-quality feedback—especially for FPs, which are often challenging for automated systems to detect due to their ambiguous characteristics. These confirmed FPs are then added back to the training set K with the pseudo-labeled normal data, thereby improving the model's understanding of normality and reducing future misclassifications.

It is important to note that the accuracy of pseudo-labels during this phase varies depending on the characteristics and performance of each model, and is tightly linked to the adaptive EER threshold established after each training iteration. This threshold, derived from the active learning validation set, is recalibrated iteratively to reflect the evolving decision boundary of the model, which is shaped by both the updated weights and the data characteristics in each slice K. As a result, both the model and its EER threshold continuously adapt over time, facilitating stable and progressive learning, as illustrated in Figure 1.

3.2 Base Algorithms

As discussed in Section 2, pixel-based algorithms raise significant privacy concerns and face deployment challenges in real-world testbeds. To mitigate these issues and improve robustness, our study adopts a pose-based methodology using three recent state-of-the-art models with available code: STG-NF [18], TSGAD [30], and SPARTA [29]. TSGAD, originally a two-stream model, combines a Graph-Attentive Variational Autoencoder (GA-VAE) for pose reconstruction with trajectory prediction. For fair benchmarking, we use only the pose branch, training it via ELBO maximization to build a latent distribution of normal skeletal poses and detect anomalies based on distributional deviations. The graph-based design improves spatial relationship modeling between keypoints, enhancing anomaly localization in

joint dynamics.

In contrast, STG-NF uses normalizing flows to map skeletal motion directly to a Gaussian distribution, providing exact likelihood scores for anomaly detection without extra heuristics. Its small size (1K parameters) makes it ideal for real-time and resource-limited environments. For consistency in evaluation, we reverse the original class labeling, treating anomalies as positive.

SPARTA [29, 32], a transformer-based model, introduces spatio-temporal and relative pose tokenization with a shared encoder and dual decoders. The current decoder reconstructs present tokens, while the future decoder predicts upcoming ones. For diversity in model paradigms, we use only the prediction branch, enabling a comprehensive evaluation across three major self-supervised anomaly detection frameworks: reconstruction-based, probability-based, and prediction-based.

3.3 Dataset & Metrics

For our framework evaluation, we utilize the HuVAD dataset [31, 33], which contains over five million annotated frames collected from seven CCTV cameras across five days in a real-world community space. The dataset provides continuous recordings with human pose sequences and tracking IDs, supporting accurate cross-camera motion tracking for systematic evaluation of pose-based anomaly detection. It features a wide range of anomalous actions, including throwing, raising hands, lying down, falling, punching, kicking, pushing, and strangling, allowing for a comprehensive performance assessment in complex conditions. In particular, HuVAD includes a context-specific camera (CSC) that captures environments where typically anomalous actions, such as simulated fights, are considered normal due to security personnel training. This unique aspect enables evaluation of model adaptability and domain-specific learning, particularly useful for active learning-based anomaly detection frameworks.

Evaluation of anomaly detection models typically relies on metrics such as **AUC-ROC**, **AUC-PR**, and **EER**, each offering different advantages. AUC-ROC and AUC-PR evaluate performance across thresholds, with AUC-PR being particularly robust to class imbalance—critical in anomaly detection contexts. However, both metrics lack a concrete decision threshold, limiting their operational applicability. In contrast, **EER** (Equal Error Rate) provides a meaningful threshold where false positive and false negative rates are equal, making it valuable for deployment scenarios. To address limitations in existing metrics, we propose the **Error Balance Index (EBI)**, a novel metric that emphasizes the balance between FPR and FNR. While traditional metrics like Balanced Error Rate (BER) use the arithmetic mean of FPR and FNR, they inadequately capture the dominance of one error type over the other. In real-world settings, maintaining this balance is vital, as both false alarms and missed detections carry significant consequences. Drawing inspiration from the F1-score [34], EBI is defined as the harmonic mean of (1-FPR) and (1-FNR): $EBI = (2 \times (1 - FPR) \times (1 - FNR))/((1 - FPR) + (1 - FNR))$, offering increased sensitivity to error imbalances and prioritizing misclassifications over correct predictions.

4 Experiments and Evaluation

As detailed in our dataset description, the HuVAD dataset is specifically engineered to facilitate continual anomaly detection training. For each camera, the dataset is partitioned into

two distinct subsets: a continual anomaly training set and a continual anomaly test set. The training set is further divided into nine slices to emulate a continual training environment. Analysis of pose counts for each continual set indicates that the anomaly rate in the training set remains below one percent per camera, whereas the test set exhibits an anomaly rate of approximately fifty percent, as detailed in Table 1. To assess the performance of our proposed method, ALFred, we employ the continual test set from HuVAD as a simulated validation set. This setup mirrors data collected during an initial warm-up phase, reflecting real-world operational conditions. The nine slices of the continual training set are subsequently utilized during the active learning phase to refine ALFred's capabilities.

	Active Learning Train Set			Active Learning Test Set		
Cam ID	Normal	Anomaly	Anomaly Rate (%)	Normal	Anomaly	Anomaly Rate (%)
0	430,659	4,216	0.97	26,052	26,093	49.96
1	722,160	5,014	0.69	28,523	28,597	49.94
2	661,280	6,813	1.02	25,402	25,449	49.95
3	1,608,498	8,042	0.50	15,786	15,818	49.95
4	415,838	1,530	0.37	37,208	37,274	49.96
5	639,233	4,278	0.66	28,268	28,353	49.92
CSC	730,386	3,108	0.42	28,301	28,343	49.96

Table 1: Total pose count and ratio for active learning from HuVAD[31]

For initializing the weights of the source domain, referred to as K-1, we used the ShanghaiTech dataset [25], a widely recognized benchmark in the VAD domain. These K-1 weights are first validated in the simulated validation set, namely, the HuVAD continual test set, to establish an EER threshold, customized to the target domain. This threshold is determined with human-in-the-loop guidance, based on the trade-off between FNR and FPR. Following this validation, the K-1 weights are used to assign pseudo-labels to incoming target domain data. The accuracy of these pseudo-labels is inherently dependent on the characteristics of each model and the defined EER threshold, and can vary during training. The final pseudo-label accuracy is evaluated against ground-truth labels from each data slice, using the overall FNR and FPR as reference metrics.

Continual Learning [40]: No filtering is applied, and the training set includes all actual positives and negatives from the incoming data.

Pseudo-Continual Learning: Only data pseudo-labeled as normal by the K-1 weights is incorporated for training.

Active Learning: Human annotators identify FPs among the pseudolabeled anomalies - data incorrectly classified as anomalous but normal - and these FPs are added to the pseudolabeled normal data for training.

AL Light: A variant of active learning that minimizes annotator workload by sending only labeled positives with scores between the EER threshold and the rolling 50th percentile of recent max anomaly scores to human-in-the-loop review.

When discussing human-in-the-loop learning, our evaluation goes a step further by analyzing how to reduce annotator workload in an active learning scenario. Figure. 2 illustrates the cumulative score distribution for positive labels during the active learning phase C0, using TSGAD [30] as an example. The blue line represents the 50th percentile between the threshold and the maximum anomaly score at each step. Based on this analysis, limiting annotation requests to scores between the threshold and the midpoint can reduce the labeling workload by approximately 34.22%. Due to the incremental nature of data feeding during training, we adopt a rolling 50th percentile approach, which consistently reduces the annotation load by at least 24.35% at each step.

Model	Training Methodology	AUC-ROC	AUC-PR
		CSC	
STG-NF	Offline Training	43.71	48.29
	Continual Learning [40]	43.51	49.25
	Pseudo-Continual Learning	**44.51**	**49.51**
	Active Learning	43.61	49.21
	AL Light	43.81	49.10
TSGAD	Offline Training	44.31	54.23
	Continual Learning [40]	47.36	55.69
	Pseudo-Continual Learning	47.36	55.52
	Active Learning	**49.48**	**57.95**
	AL Light	48.13	57.25
SPARTA	Offline Training	45.24	50.36
	Continual Learning [40]	48.14	52.61
	Pseudo-Continual Learning	48.11	52.60
	Active Learning	**48.48**	**52.86**
	AL Light	48.14	52.44

Table 2: Comparison AUC-ROC and AUC-PR of CSC for models in different training methods.

We initially assessed the performance of our models using the widely adopted metrics of AUC-ROC and AUC-PR. These metrics were applied to evaluate three distinct training data scenarios, with a baseline model established by training offline on the entire active learning training set without any data slicing. As shown in Table 2, using CSC as an example, all model weights were tested on the active learning validation set. Given the multi-step nature of the continual training process, we recorded the best results from each scenario to provide a comprehensive analysis. More details can be found in the supplementary materials for replication and further investigation.

While the findings align with recent work [40], confirming the effectiveness of our approach in achieving robust detection performance across diverse training conditions and highlighting its potential for practical VAD applications, the trends observed in Table 2 do not necessarily translate to real-world scenarios. This discrepancy, primarily discussed in Section 3.1, arises from the fact that these metrics are computed over varying thresholds and, by their nature, do not provide direct real-world insights.

Since AUC-ROC and AUC-PR do not directly translate to operational performance in real-world scenarios, we adopted a different evaluation strategy for our continual training setup. Specifically, for each incoming data slice, we used the current model's EER threshold to assign pseudo-labels and computed TP, FP, TN, and FN by comparing these pseudo-labels to the ground truth. This method ensures a fair comparison across all scenarios by applying consistent pseudo-labeling criteria and is particularly appropriate for continual training, where offline training results cannot serve as a valid baseline due to the dynamic nature of the data stream. Such a comparison would be unfair given the static nature of offline training datasets. The accumulated results from each step for all cameras were used to calculate the overall FNR and FPR, as presented in Table 2. Detailed results for different cameras can be found in the supplementary materials for replication and further investigation.

Notably, while trends in AUC-ROC and AUC-PR on the validation set, do not always directly reflect model performance in real-world scenarios, these metrics remain essential within the ALFred framework for assessing model performance and determining the EER threshold. This threshold balances the trade-off between false positives and false negatives,

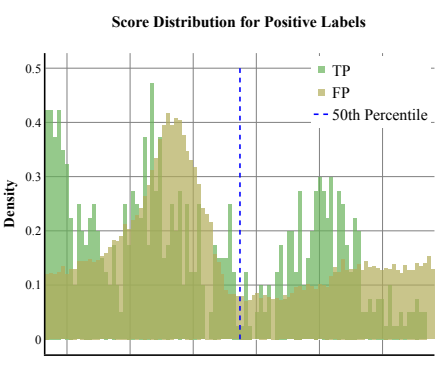

Figure 2: Score distribution for positive labels from C0 during Active Learning phase.

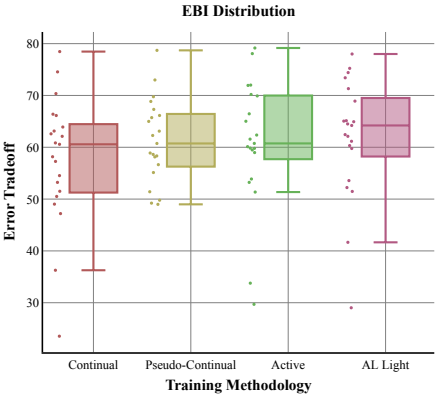

Figure 3: Error tradeoff distribution across different training methodologies.

Training Methodology	FNR %	FPR %	EBI %	FNR %	FPR %	EBI %	FNR %	FPR %	EBI %
	STG-NF [13]			TSGAD [30]			SPARTA [25]		
Continual Learning [10]	32.77	45.92	59.94	38.68	38.20	61.56	**24.25**	41.55	65.98
Pseudo-Continual	37.74	**28.01**	**66.77**	**29.84**	47.82	59.85	25.67	40.61	66.02
Active Learning	**35.24**	39.85	62.36	41.29	**31.29**	63.31	24.96	**39.46**	**67.01**
AL Light	37.67	37.50	62.56	39.33	31.73	**64.24**	24.86	39.78	66.85

Table 3: Comparison FNR and FPR of cumulative data for models in different training

which is critical for practical deployment. According to the study [2], FPR is particularly vital in real-world applications as it directly affects user trust and system reliability.

Based on Table 3, the two selected models exhibit distinct performance patterns in real-world scenarios. The lighter model, STG-NF, demonstrates a lower FPR when trained using the pseudo-continual learning approach, which involves training on a smaller amount of data. However, as we transition to a larger and more complex dataset acquired through active learning, STG-NF fails to maintain this advantage, whereas TSGAD and SPARTA benefit from active learning, achieving improved FPR.

While the tables provide numerical insights, they do not explicitly illustrate the trade-off between different methodologies in terms of balancing FPR and FNR. To address this, we conducted an EBI analysis, introduced in Section 3.3, which quantifies this balance and offers a clearer comparative assessment. As shown in Figure 3, the EBI analysis reveals that Active Learning Light outperforms other methods, achieving the Q3 (25th percentile) at 68.91 and the highest median at 64.21. This indicates that, on average, Active Learning Light provides the most effective trade-off between reducing FPR and FNR, leading to more efficient anomaly classification. Compared to Continual Learning and Pseudo-Continual Learning, the higher Q3 and median suggest that Active Learning consistently achieves better overall reliability in most scenarios. The AL-Light method achieves comparable results to Active Learning while additionally reducing the human-in-the-loop workload by at least 24.35%. However, its wider range of values implies that performance variability is higher in extreme cases, indicating potential fluctuations under certain conditions. This trade-off underscores that while AL Light is the most effective approach for minimizing errors, careful parameter tuning and ongoing monitoring are necessary to manage its occasional instability.

5 Research Questions and Future Directions

ALFred represents a first step toward adaptability in pose-based anomaly detection, but there remains room for improvement. One of the most critical areas for advancement is enhancing the accuracy of the framework, which can be achieved through multiple approaches. These include refining the model architecture, incorporating more diverse data, and leveraging advanced optimization techniques. One particularly promising method is reinforcement learning (RL), which enables active exploration of uncertain samples, allowing the model to prioritize ambiguous or low-likelihood instances for more effective learning. Additionally, adaptive anomaly score calibration through RL facilitates dynamic threshold optimization, improving robustness across varying conditions, while meta-learning and reward-based adaptation enhance generalization to unseen environments.

Beyond accuracy, another direction is evaluating the practical feasibility of the framework by assessing timing constraints, computational efficiency, and hardware requirements. Finally, achieving real-world applicability requires long-term deployment in operational settings, where the system can be continuously monitored to assess robustness, adaptability, and failure points. A structured evaluation and refinement cycle would provide critical insights into the system's strengths and limitations, bridging the gap between research and real-world implementation.

6 Conclusion

This paper addresses the challenges of deploying VAD models in real-world settings where domain shifts degrade performance. Conventional metrics like AUC-ROC and AUC-PR fail to provide actionable thresholds due to uncalibrated anomaly scores. To address this, we introduce ALFred, an active learning framework that efficiently adapts VAD models with minimal labeled data. ALFred employs a two-phase strategy: a warm-up phase using pseudo-labels refined by human annotators, and an active learning phase that fine-tunes the model with human feedback to establish an EER threshold. Experiments on the HuVAD dataset show that ALFred reduces false positives and enhances adaptability, demonstrating the potential of active learning for real-world VAD deployment. This work lays the foundation for further research into domain adaptation and human-in-the-loop refinement of anomaly detection systems.

Acknowledgment

This research is supported by the National Science Foundation (NSF) under Award No. 1831795.

References

[1] Ghazal Alinezhad Noghre, Armin Danesh Pazho, Vinit Katariya, and Hamed Tabkhi. Understanding the challenges and opportunities of pose-based anomaly detection. In *Proceedings of the 8th International Workshop on Sensor-Based Activity Recognition and Artificial Intelligence*, iWOAR '23, New York, NY, USA, 2023. Association for

Computing Machinery. ISBN 9798400708169. doi: 10.1145/3615834.3615844. URL https://doi.org/10.1145/3615834.3615844.

[2] Babak Rahimi Ardabili, Armin Danesh Pazho, Ghazal Alinezhad Noghre, Vinit Katariya, Gordon Hull, Shannon Reid, and Hamed Tabkhi. Exploring public's perception of safety and video surveillance technology: A survey approach. *Technology in Society*, 78:102641, 2024.

[3] Jinan Bao, Hanshi Sun, Hanqiu Deng, Yinsheng He, Zhaoxiang Zhang, and Xingyu Li. Bmad: Benchmarks for medical anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 4042–4053, June 2024.

[4] Mohammad Baradaran and Robert Bergevin. Multi-task learning based video anomaly detection with attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2886–2896, June 2023.

[5] Antonio Barbalau, Radu Tudor Ionescu, Mariana-Iuliana Georgescu, Jacob Dueholm, Bharathkumar Ramachandra, Kamal Nasrollahi, Fahad Shahbaz Khan, Thomas B Moeslund, and Mubarak Shah. Ssmtl++: Revisiting self-supervised multi-task learning for video anomaly detection. *Computer Vision and Image Understanding*, 229:103656, 2023.

[6] Dongyue Chen, Lingyi Yue, Xingya Chang, Ming Xu, and Tong Jia. Nm-gan: Noise-modulated generative adversarial network for video anomaly detection. *Pattern Recognition*, 116:107969, 2021.

[7] Xiaoyu Chen, Shichao Kan, Fanghui Zhang, Yigang Cen, Linna Zhang, and Damin Zhang. Multiscale spatial temporal attention graph convolution network for skeleton-based anomaly behavior detection. *Journal of Visual Communication and Image Representation*, 90:103707, 2023.

[8] Kai-Wen Cheng, Yie-Tarng Chen, and Wen-Hsien Fang. Gaussian process regression-based video anomaly detection and localization with hierarchical feature representation. *IEEE Transactions on Image Processing*, 24(12):5288–5301, 2015.

[9] Serhan Coşar, Giuseppe Donatiello, Vania Bogorny, Carolina Garate, Luis Otavio Alvares, and François Brémond. Toward abnormal trajectory and event detection in video surveillance. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(3): 683–695, 2016.

[10] Armin Danesh Pazho, Ghazal Alinezhad Noghre, Babak Rahimi Ardabili, Christopher Neff, and Hamed Tabkhi. Chad: Charlotte anomaly dataset. In *Scandinavian Conference on Image Analysis*, pages 50–66. Springer, 2023.

[11] Keval Doshi and Yasin Yilmaz. Continual learning for anomaly detection in surveillance videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.

[12] Keval Doshi and Yasin Yilmaz. Online anomaly detection in surveillance videos with asymptotic bound on false alarm rate. *Pattern Recognition*, 114:107865, 2021.

[13] Keval Doshi and Yasin Yilmaz. Towards interpretable video anomaly detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2655–2664, 2023.

[14] Yona Falinie A. Gaus, Neelanjan Bhowmik, Brian K. S. Isaac-Medina, Hubert P. H. Shum, Amir Atapour-Abarghouei, and Toby P. Breckon. Region-based appearance and flow characteristics for anomaly detection in infrared surveillance imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2995–3005, June 2023.

[15] Mariana-Iuliana Georgescu, Antonio Barbalau, Radu Tudor Ionescu, Fahad Shahbaz Khan, Marius Popescu, and Mubarak Shah. Anomaly detection in video via self-supervised and multi-task learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12742–12752, 2021.

[16] Mariana Iuliana Georgescu, Radu Tudor Ionescu, Fahad Shahbaz Khan, Marius Popescu, and Mubarak Shah. A background-agnostic framework with adversarial training for abnormal event detection in video. *IEEE transactions on pattern analysis and machine intelligence*, 44(9):4505–4523, 2021.

[17] Lars Heckler, Rebecca König, and Paul Bergmann. Exploring the importance of pre-trained feature extractors for unsupervised anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2917–2926, June 2023.

[18] Or Hirschorn and Shai Avidan. Normalizing flows for human pose anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13545–13554, 2023.

[19] Chao Huang, Yabo Liu, Zheng Zhang, Chengliang Liu, Jie Wen, Yong Xu, and Yaowei Wang. Hierarchical graph embedded pose regularity learning via spatio-temporal transformer for abnormal behavior detection. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 307–315, 2022.

[20] Xiangyu Huang, Caidan Zhao, Jinhui Yu, Chenxing Gao, and Zhiqiang Wu. Multi-level memory-augmented appearance-motion correspondence framework for video anomaly detection. In *2023 IEEE International Conference on Multimedia and Expo (ICME)*, pages 2717–2722. IEEE, 2023.

[21] Yashswi Jain, Ashvini Kumar Sharma, Rajbabu Velmurugan, and Biplab Banerjee. Posecvae: Anomalous human activity detection. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 2927–2934. IEEE, 2021.

[22] Hamza Karim, Keval Doshi, and Yasin Yilmaz. Real-time weakly supervised video anomaly detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 6848–6856, 2024.

[23] Mustaqeem Khan, Jamil Ahmad, Abdulmotaleb El Saddik, Wail Gueaieb, Giulia De Masi, and Fakhri Karray. Drone-hat: Hybrid attention transformer for complex action recognition in drone surveillance videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 4713–4722, June 2024.

[24] Demetris Lappas, Vasileios Argyriou, and Dimitrios Makris. Dynamic distinction learning: Adaptive pseudo anomalies for video anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 3961–3970, June 2024.

[25] Wen Liu, Weixin Luo, Dongze Lian, and Shenghua Gao. Future frame prediction for anomaly detection–a new baseline. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6536–6545, 2018.

[26] Amir Markovitz, Gilad Sharir, Itamar Friedman, Lihi Zelnik-Manor, and Shai Avidan. Graph embedded pose clustering for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10539–10547, 2020.

[27] Romero Morais, Vuong Le, Truyen Tran, Budhaditya Saha, Moussa Mansour, and Svetha Venkatesh. Learning regularity in skeleton trajectories for anomaly detection in videos. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11996–12004, 2019.

[28] Ghazal Alinezhad Noghre, Armin Danesh Pazho, and Hamed Tabkhi. An exploratory study on human-centric video anomaly detection through variational autoencoders and trajectory prediction. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 995–1004, 2024.

[29] Ghazal Alinezhad Noghre, Armin Danesh Pazho, and Hamed Tabkhi. Posewatch: A transformer-based architecture for human-centric video anomaly detection using spatio-temporal pose tokenization. *arXiv preprint arXiv:2408.15185*, 2024.

[30] Ghazal Alinezhad Noghre, Armin Danesh Pazho, and Hamed Tabkhi. An exploratory study on human-centric video anomaly detection through variational autoencoders and trajectory prediction. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 995–1004, 2024.

[31] Ghazal Alinezhad Noghre, Shanle Yao, Armin Danesh Pazho, Babak Rahimi Ardabili, Vinit Katariya, and Hamed Tabkhi. Pheva: A privacy-preserving human-centric video anomaly detection dataset. *arXiv preprint arXiv:2408.14329*, 2024.

[32] Ghazal Alinezhad Noghre, Armin Danesh Pazho, and Hamed Tabkhi. Human-centric video anomaly detection through spatio-temporal pose tokenization and transformer, 2025. URL https://arxiv.org/abs/2408.15185.

[33] Armin Danesh Pazho, Shanle Yao, Ghazal Alinezhad Noghre, Babak Rahimi Ardabili, Vinit Katariya, and Hamed Tabkhi. Towards adaptive human-centric video anomaly detection: A comprehensive framework and a new benchmark, 2025. URL https://arxiv.org/abs/2408.14329.

[34] Van Rijsbergen. Information retrieval; ; butterworth, 1978. *J. librariansh.*, 11:237, 1979.

[35] Royston Rodrigues, Neha Bhargava, Rajbabu Velmurugan, and Subhasis Chaudhuri. Multi-timescale trajectory prediction for abnormal human activity detection. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 2626–2634, 2020.

[36] Damith Chamalke Senadeera, Xiaoyun Yang, Dimitrios Kollias, and Gregory Slabaugh. Cue-net: Violence detection video analytics with spatial cropping enhanced uniformerv2 and modified efficient additive attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 4888–4897, June 2024.

[37] Guodong Wang, Yunhong Wang, Jie Qin, Dongming Zhang, Xiuguo Bao, and Di Huang. Video anomaly detection by solving decoupled spatio-temporal jigsaw puzzles. In *European Conference on Computer Vision*, pages 494–511. Springer, 2022.

[38] Zhengye Yang and Richard J. Radke. Context-aware video anomaly detection in long-term datasets. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 4002–4011, June 2024.

[39] Ziyi Yang, Iman Soltani, and Eric Darve. Anomaly detection with domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2958–2967, June 2023.

[40] Shanle Yao, Ghazal Alinezhad Noghre, Armin Danesh Pazho, and Hamed Tabkhi. Evaluating the effectiveness of video anomaly detection in the wild: Online learning and inference for real-world deployment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 4832–4841, June 2024.

[41] Shoubin Yu, Zhongyin Zhao, Haoshu Fang, Andong Deng, Haisheng Su, Dongliang Wang, Weihao Gan, Cewu Lu, and Wei Wu. Regularity learning via explicit distribution modeling for skeletal video anomaly detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.

[42] Yuan Yuan, Jianwu Fang, and Qi Wang. Online anomaly detection in crowd scenes via structure analysis. *IEEE transactions on cybernetics*, 45(3):548–561, 2014.

[43] M Zaigham Zaheer, Arif Mahmood, M Haris Khan, Mattia Segu, Fisher Yu, and Seung-Ik Lee. Generative cooperative learning for unsupervised video anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14744–14754, 2022.

[44] Xianlin Zeng, Yalong Jiang, Wenrui Ding, Hongguang Li, Yafeng Hao, and Zifeng Qiu. A hierarchical spatio-temporal graph convolutional neural network for anomaly detection in videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(1):200–212, 2021.

[45] Yuxing Zhang, Jinchen Song, Yuehan Jiang, and Hongjun Li. Online video anomaly detection. *Sensors*, 23(17):7442, 2023.