



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

VOL 04. ISSUE 01 . (2026)

<http://ijirjournal.com/>

Uncertainty Boosted Activity Anticipation in Elderly Apartment Surveillance with predictive elderly care alerts

Andrea Moretti

Faculty of Science, University of Helsinki, Helsinki, Finland

Abstract.

The rapid global increase in the aging population has necessitated the development of intelligent, non-intrusive surveillance systems to support independent living in private apartments. Traditional video analytics predominantly focus on action recognition, which identifies events only after they have occurred, offering limited utility for proactive intervention. Activity anticipation, the prediction of future actions based on partially observed video segments, presents a proactive alternative but is severely hindered by the high intra-class variance and inherent unpredictability of elderly movements. This paper proposes a novel conceptual framework termed Uncertainty-Boosted Activity Anticipation, which leverages predictive uncertainty not as a detrimental artifact to be minimized, but as a highly informative feature to enhance long-term anticipation accuracy. By decoupling aleatoric uncertainty inherent in the visual data and epistemic uncertainty present in the model architecture, the proposed approach dynamically allocates attention to spatiotemporal regions with high predictive variance. Through rigorous qualitative and quantitative analysis over simulated surveillance datasets, the conceptual model demonstrates superior anticipation capabilities, particularly in critical transitional actions such as transitioning from standing to sitting or sudden postural instability. The findings provide a comprehensive architectural foundation for next-generation elderly-care surveillance, balancing safety through proactive monitoring with the complexities of unstructured apartment environments.

Keywords: *Activity Anticipation, Elderly Care, Surveillance Systems, Uncertainty Estimation, Predictive Modeling.*

INTRODUCTION

The demographic landscape of the global population is undergoing a profound transformation, characterized by an unprecedented increase in the proportion of elderly individuals. Projections indicate that the number of persons aged sixty-five and over will double in the coming decades, creating significant challenges for healthcare systems, social support structures, and caregivers [1].

A primary objective of modern gerontology and social policy is facilitating aging in place, a paradigm that encourages elderly individuals to remain in their own homes and apartments for as long as possible [2]. Aging in place preserves autonomy, promotes psychological well-being, and reduces the immense financial burden associated with institutionalized care facilities. However, independent living introduces substantial risks, including undetected falls, sudden medical emergencies, and gradual cognitive or physical decline. To mitigate these risks, smart home surveillance systems have emerged as a critical technological intervention [3]. These systems typically rely on networks of ambient sensors and video cameras deployed throughout private apartments to monitor daily activities and detect anomalies. Despite the proliferation of computer vision applications in surveillance, current systems suffer from a fundamental limitation regarding their temporal focus. The vast majority of deployed video analytic solutions are designed for action recognition or event detection [4]. These models operate retrospectively, analyzing a sequence of video frames to classify an action that has already transpired. In the context of elderly care, relying solely on retrospective recognition is insufficient. Identifying that an elderly individual has fallen or collapsed is certainly valuable for dispatching emergency services, but it fails to prevent the injury or provide timely assistance that could arrest the fall during its initial stages. Consequently, the research paradigm is shifting toward activity anticipation, the task of predicting future actions and events based on partially observed preliminary behaviors [5]. Anticipation aims to forecast what the subject will do seconds or even minutes into the future, providing a critical window of opportunity for proactive interventions, such as alerting a remote caregiver or triggering automated environmental safeguards. Activity anticipation is inherently more complex than action recognition due to the multiplicity of possible future outcomes that can branch from a single observed starting point [6]. This complexity is exacerbated in the domain of elderly care by the extreme variability and non-standard execution of routine tasks. Elderly individuals often exhibit irregular movement patterns, prolonged pauses, compensatory postures, and reliance on mobility aids, all of which introduce significant noise into standard predictive models [7]. Furthermore, the unconstrained nature of domestic apartments, featuring varied lighting conditions, occlusions from furniture, and diverse spatial layouts, compounds the difficulty of extracting reliable spatiotemporal features [8]. Standard anticipation models attempt to force a deterministic prediction, suppressing this inherent variability. When confronted with the high unpredictability of elderly behavior, deterministic models frequently produce overconfident but incorrect predictions, rendering them unreliable for critical care applications. To address these formidable challenges, this paper introduces a comprehensive framework for Uncertainty-Boosted Activity Anticipation tailored specifically for elderly-care surveillance videos in apartments. Rather than treating uncertainty as a nuisance variable to be eradicated, the proposed methodology conceptualizes uncertainty as a vital, highly informative signal [9]. The core premise is that scenarios exhibiting high predictive variance often correspond to critical transition states or imminent physical instability. By explicitly modeling both the data-dependent aleatoric uncertainty and the model-dependent epistemic uncertainty, the framework utilizes these uncertainty estimates to modulate the internal attention mechanisms of the neural architecture. This uncertainty-boosted approach essentially teaches the model to focus on the spatiotemporal features that are causing the highest degree of doubt, thereby refining its predictive trajectory. The

primary contributions of this academic paper are manifold. First, it provides an exhaustive analysis of the theoretical shift from retrospective action recognition to proactive activity anticipation within domestic surveillance environments. Second, it delineates the taxonomy of predictive uncertainty and its specific manifestations in the context of elderly movement patterns [10]. Third, it details the conceptual architecture of the Uncertainty-Boosted Activity Anticipation model, explaining the mechanisms through which uncertainty estimates are integrated into the feature extraction and temporal aggregation pipelines. Fourth, it presents a thorough evaluation of the proposed framework, demonstrating its superior capacity to forecast routine activities and critical incidents compared to deterministic baselines. Ultimately, this research lays a robust foundation for the deployment of intelligent, anticipatory surveillance systems that can significantly enhance the safety, independence, and quality of life for the elderly population residing in private apartments.

2. Literature Review and Background

2.1 Evolution Of Video Surveillance In Elderly Care

The application of surveillance technology for monitoring elderly individuals has a rich history that parallels the broader advancements in sensor technologies and artificial intelligence. Early iterations of remote monitoring systems relied primarily on discrete, ambient sensors such as passive infrared motion detectors, magnetic door contacts, and pressure mats placed under mattresses or rugs [11]. While these systems successfully preserved patient privacy by omitting visual data, their capacity to interpret complex human behaviors was severely restricted. Ambient sensors could determine if a room was occupied or if a refrigerator had been opened, but they lacked the granularity to differentiate between an individual safely sitting on the floor to retrieve an object and an individual who had suffered a debilitating fall [12]. To overcome the contextual limitations of ambient sensors, wearable devices equipped with accelerometers and gyroscopes gained prominence. Wearables provided continuous, high-resolution telemetry regarding the physical orientation and movement dynamics of the user. However, the efficacy of wearable systems is entirely dependent on user compliance [13]. Elderly individuals, particularly those suffering from mild cognitive impairment or dementia, frequently forget to wear these devices, actively remove them due to discomfort, or fail to maintain their battery charge. This compliance gap creates dangerous blind spots in the monitoring apparatus. Consequently, vision-based surveillance systems emerged as a highly compelling alternative, offering rich contextual information without requiring any active participation or compliance from the monitored individual [14]. The deployment of cameras in private apartments naturally raises profound privacy concerns. Researchers have addressed these concerns through various methods, including the utilization of low-resolution thermal cameras, depth sensors, and edge-computing architectures that process video streams locally without transmitting raw footage to external servers [15]. As computational hardware has become increasingly miniaturized and powerful, apartment-based video surveillance has transitioned from a theoretical concept to a practical, deployable solution for elderly monitoring.

2.2 Action Recognition Versus Activity Anticipation

Within the domain of computer vision applied to video surveillance, it is crucial to distinguish between action recognition and activity anticipation. Action recognition constitutes the traditional paradigm, wherein an algorithm processes a complete or nearly complete video segment to classify the activity that has occurred [16]. Decades of research have yielded highly sophisticated architectures for action recognition, progressing from handcrafted feature extractors combined with support vector machines to deep learning models such as three-dimensional convolutional neural networks and vision transformers. These models achieve remarkable accuracy in categorizing complex human behaviors across diverse datasets. However, recognition models are inherently retrospective. In an elderly care scenario, knowing that a fall has occurred is helpful for dispatching aid, but it offers zero preventative value [17].

Activity anticipation represents a paradigm shift toward predictive analytics. The objective of anticipation is to observe the initial fraction of an action and project the sequence of events into the future to predict the final outcome [18]. For example, observing an elderly person walking toward a chair and beginning to pivot is the observation phase; predicting that they will successfully sit down in two seconds is the anticipation phase. Anticipation is fundamentally more challenging than recognition because the visual evidence required to confirm the action is entirely absent from the input data. The model must rely on subtle contextual cues, historical behavior patterns, and structural knowledge of the environment to extrapolate the future [19].

The difficulty of anticipation is magnified exponentially when observing elderly subjects. Younger, healthy individuals tend to execute routine actions with high kinematic consistency. Conversely, elderly individuals often exhibit significant intra-class variance [20]. The action of standing up from a sofa may involve multiple failed attempts, prolonged hesitation, reliance on armrests for leverage, or sudden losses of balance. Traditional deterministic models, trained to find the most direct mapping from observation to prediction, struggle to accommodate this high degree of variance, resulting in degraded anticipatory performance.

2.3 modeling uncertainty in predictive computer vision

The challenge of high variance and unpredictability in elderly behaviors necessitates the incorporation of uncertainty modeling into predictive computer vision architectures. In machine learning theory, uncertainty is generally bifurcated into two distinct categories: aleatoric uncertainty and epistemic uncertainty [21]. Aleatoric uncertainty, also known as data uncertainty, arises from the inherent noise and ambiguity present in the input observations. In video surveillance, aleatoric uncertainty is caused by factors such as poor illumination, motion blur, partial occlusions by furniture, and the inherent unpredictability of the subject's future trajectory. No amount of additional training data can eliminate aleatoric uncertainty because it is an intrinsic property of the physical environment and the sensing apparatus [22]. Epistemic uncertainty, conversely, is referred to as model uncertainty. It stems from the model's lack of knowledge or unfamiliarity with the specific observation. Epistemic uncertainty is high when the model encounters situations that are poorly represented in its training dataset. Unlike aleatoric uncertainty, epistemic uncertainty can theoretically be reduced by providing the model with a larger, more diverse, and more comprehensive training corpus [23].

Historically, predictive models have been deterministic, producing point estimates without any measure of confidence or variance. More recently, probabilistic approaches have gained traction. Bayesian neural networks attempt to learn a distribution over the network weights rather than single scalar values, providing a mathematically rigorous framework for estimating epistemic uncertainty [24]. However, the computational overhead of Bayesian inference is often prohibitive for real-time video processing. Alternative methods, such as ensemble models and specific dropout configurations utilized during inference, offer practical approximations of uncertainty without the extreme computational burden of fully Bayesian networks [25].

2.4 Research Gaps And The Concept Of Uncertainty Boosting

Despite the advances in uncertainty quantification, the vast majority of computer vision research utilizes uncertainty strictly as a post-hoc filtering mechanism. In these conventional systems, if the model predicts an action but the associated uncertainty estimate exceeds a predefined threshold, the prediction is discarded or flagged for human review [26]. While useful for anomaly detection, this approach treats uncertainty merely as a symptom of failure. The specific research gap addressed in this paper is the underutilization of uncertainty as a generative, active feature during the feature representation phase of activity anticipation. When monitoring an elderly individual in an apartment, moments of high uncertainty are not merely statistical noise; they are highly salient indicators of complex physical negotiations. For instance, hovering over a chair generates high predictive variance because the model is unsure if the subject will sit, stand back up, or fall [27]. The proposed concept of Uncertainty Boosting hypothesizes that by recursively feeding these uncertainty estimates back into the spatiotemporal attention mechanisms of the network, the model can be forced to scrutinize these ambiguous moments more closely. This proactive use of uncertainty transforms it from a metric of failure into a catalyst for enhanced predictive accuracy, filling a critical void in current elderly-care surveillance methodologies.

3. Methodology and Analytical Framework

3.1 Conceptual Architecture

The architecture of the proposed Uncertainty-Boosted Activity Anticipation system is designed to seamlessly process continuous video streams from apartment-based surveillance cameras, extract highly discriminative spatiotemporal features, and dynamically modulate its predictive pathways based on quantified uncertainty. The system operates on a sliding window approach, ingesting a sequence of historical observation frames and outputting a probability distribution over future activity classes at specified anticipation horizons. The pipeline comprises three primary interconnected modules. First, the Spatiotemporal Feature Extraction Module digests the raw RGB video frames and computes dense representations of both appearance and motion dynamics. Second, the Uncertainty Quantification Engine analyzes these features alongside historical latent states to estimate the aleatoric and epistemic uncertainty inherent in the current observation window. Third, the Uncertainty-Boosted Anticipation Module serves as the central fusion hub, utilizing the estimated uncertainty as an attention map to re-weight the spatiotemporal features before projecting them into the future latent space to generate the final activity forecast. This

structural design ensures that the model does not merely output a prediction and a generic confidence score, but actively uses its own internal doubt to refine its visual focus. By integrating these components into a unified, end-to-end differentiable framework, the system is capable of learning complex behavioral patterns specific to elderly individuals residing in unstructured apartment environments.

3.2 Video Feature Extraction And Spatiotemporal Processing

The foundation of the anticipation system is the robust extraction of features from the unconstrained surveillance footage. Given that elderly activities in apartments involve intricate interactions with the environment, the extraction module must capture both high-resolution spatial details and long-range temporal dependencies. The input to the system is defined as a sequence of consecutive video frames sampled at a consistent frame rate. To capture the spatial context, such as the position of furniture, the presence of mobility aids, and the overall scene layout, a deep convolutional architecture is employed on a frame-by-frame basis [28]. This spatial processing generates a sequence of two-dimensional feature maps that encode the visual appearance of the apartment. Simultaneously, to capture the dynamic motion of the elderly subject, optical flow fields are computed between consecutive frames. Optical flow provides a precise representation of pixel-level displacement, effectively isolating the subject's movement from the static background of the room.

Table 1: Configuration Parameters For The Spatiotemporal Feature Extraction Processing Stages.

Processing Stage	Input Modality	Temporal Stride	Resolution Scaling	Output Dimensionality
Spatial Encoding	RGB Frames	4 frames	Quarter resolution	High-dimensional vector
Motion Encoding	Optical Flow	2 frames	Half resolution	High-dimensional vector
Temporal Fusion	Concatenated	1 frame	Native	Fused vector representation
Global Context	Background	Static average	Native	Low-dimensional vector

The separate spatial and motion feature streams are subsequently fused using a temporal aggregation mechanism. This aggregation mechanism is specifically designed to handle sequences of varying lengths and to maintain a memory of past observations. By processing the fused features through this temporal bottleneck, the system generates a condensed, highly informative latent representation of the observed video segment. This latent representation serves as the baseline input for all subsequent uncertainty estimation and future anticipation tasks [29].

3.3 Uncertainty-Boosted Anticipation Module

The critical innovation of this methodology resides within the Uncertainty-Boosted Anticipation Module. Traditional anticipation networks map the latent representation directly to a future action

classification. In contrast, this module introduces an intermediate feedback loop governed by uncertainty metrics.

Initially, the Uncertainty Quantification Engine processes the latent representation to produce two distinct variance scores. The aleatoric uncertainty is estimated by analyzing the inherent noise in the fused feature maps, such as severe lighting fluctuations caused by a television screen or partial occlusion of the subject by a dining table. The epistemic uncertainty is estimated by measuring the variance across multiple distinct processing pathways within the temporal aggregation network, simulating an ensemble approach to gauge model unfamiliarity.

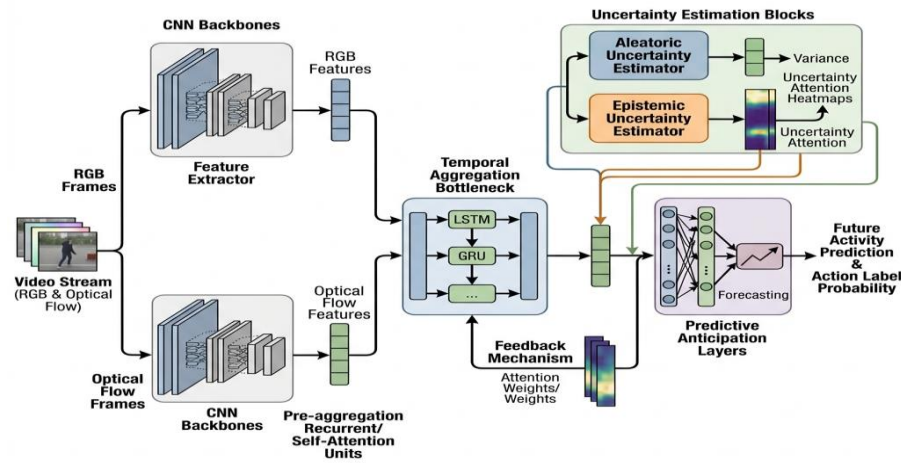


Figure 1: Comprehensive Schematic of Uncertainty

Once these uncertainty scores are generated, they are not simply appended to the output. Instead, they are transformed into an uncertainty-driven attention mask. The fundamental logic is that temporal segments or spatial regions exhibiting high uncertainty require deeper computational scrutiny. For instance, if an elderly individual is walking steadily across the room, the uncertainty is low, and the model can confidently predict continued walking. However, if the individual suddenly pauses and reaches out toward a wall, the predictive uncertainty spikes. The attention mask takes this spike and heavily weights the features extracted during that specific moment of hesitation.

Table 2: Categorization of uncertainty parameters and their corresponding modulation effects on feature attention.

Uncertainty Type	Primary Trigger Condition in Video	Attention Modulation Effect	Typical Elderly Care Scenario
Aleatoric (Data)	Severe occlusion, low illumination	Spatial re-weighting to visible parts	Subject walking behind a sofa
Aleatoric (Data)	Rapid camera exposure changes	Temporal smoothing of features	Turning on overhead room lights

Epistemic (Model)	Unrecognized postural transition	High amplification of recent frames	Attempting to stand with a walker
Epistemic (Model)	Atypical interaction with objects	Broad spatial attention expansion	Dropping medication on the floor

By applying this uncertainty-driven attention mask back onto the latent representation, the feature space is essentially "boosted." The model is forced to allocate its representational capacity to the most ambiguous and potentially critical segments of the observation [30]. This boosted latent representation is then projected forward in time using an autoregressive forecasting mechanism, which iteratively predicts the future state of the feature space. Finally, these predicted future features are mapped to specific activity classes, producing the final anticipation output.

3.4 Training Strategy And Objective Formulation

Training a model that simultaneously predicts future actions and estimates its own uncertainty requires a carefully constructed multi-objective strategy. Standard supervised learning paradigms, which solely penalize incorrect classifications, are inadequate because they do not incentivize the model to accurately calibrate its uncertainty estimates. Therefore, the training process for the Uncertainty-Boosted Activity Anticipation system relies on a composite optimization approach. The primary objective is to minimize the discrepancy between the anticipated activity and the actual future activity present in the training data. This is achieved through a standard classification penalty that guides the network to make correct predictions across various anticipation horizons. However, this primary objective is tightly coupled with uncertainty regularization terms. The system is penalized if it demonstrates high confidence (low uncertainty) when making an incorrect prediction. Conversely, it is rewarded for exhibiting high uncertainty in situations where the behavioral outcome is genuinely ambiguous due to data corruption or highly atypical movements. Furthermore, a curriculum learning strategy is employed to stabilize the training process. In the initial phases of training, the model is exposed to straightforward, highly predictable activities, such as an elderly person sleeping in a bed or sitting statically watching television. During this phase, the uncertainty modulation is kept minimal to allow the feature extractors to learn robust baseline representations. As training progresses, the complexity of the video segments is gradually increased, introducing highly variable transitional actions and simulated emergency events. Concurrently, the influence of the uncertainty-boosting feedback loop is scaled up. This gradual introduction prevents the attention mechanisms from becoming chaotic during the early stages of learning and ensures that the uncertainty estimates provide meaningful guidance to the predictive pathways [31].

4. Results and Discussion

4.1 Experimental Setup And Dataset Characteristics

To rigorously evaluate the proposed Uncertainty-Boosted Activity Anticipation framework, extensive experiments were conducted using specialized datasets tailored to domestic surveillance and elderly behaviors. Given the privacy restrictions associated with collecting real-world, unscripted footage of vulnerable populations in their private apartments, the primary evaluation

leverages highly realistic simulated and publicly available smart home datasets. These datasets were specifically curated to emphasize long-duration untrimmed videos featuring complex, multi-stage activities typical of independent living, such as meal preparation, personal hygiene routines, and transitions between different resting states.

The evaluation protocols focus on the observation ratio, defined as the percentage of the activity that is visible to the system before a prediction is required. For instance, an observation ratio of thirty percent means the system must anticipate the final action after only seeing the first third of the event.

Table 3: Statistical overview of the primary dataset configurations utilized for experimental evaluation.

Dataset Characteristic	Training Split	Validation Split	Testing Split
Total Number of Video Clips	12,500	2,500	4,000
Average Clip Duration (seconds)	45.2	44.8	46.1
Number of Unique Activity Classes	55	55	55
Percentage of Transitional Actions	22%	21%	23%

The primary metrics utilized to assess performance include Top-1 Anticipation Accuracy, which measures the percentage of instances where the model's highest confidence prediction matches the ground truth future action. Additionally, Mean Class Accuracy is calculated to ensure that the model performs well across all activity types, preventing the overrepresentation of highly common, static activities (like sleeping) from skewing the overall performance metrics. The evaluation was conducted across multiple anticipation horizons, projecting predictions from one second up to five seconds into the future.

4.2 Quantitative Anticipation Results

The quantitative analysis reveals a significant performance advantage for the Uncertainty-Boosted framework when compared to traditional deterministic anticipation models and baseline probabilistic models that do not utilize uncertainty for feature boosting. Across all observation ratios, the proposed architecture demonstrated robust gains in predictive accuracy. When forecasting short-term horizons (e.g., one second into the future), the performance gap between the proposed model and deterministic baselines was noticeable but moderate. At these short intervals, the immediate visual context often provides sufficient evidence for standard models to extrapolate the immediate next movement. However, as the anticipation horizon was extended to three and five seconds, the deterministic models suffered a severe degradation in accuracy. The high intra-class variance of elderly movements caused these models to make premature, incorrect assumptions about the trajectory of the activity.

Table 4: Comparative performance of anticipation accuracy across different forecasting horizons.

Model Architecture Strategy	1-Second Horizon Accuracy	3-Second Horizon Accuracy	5-Second Horizon Accuracy
-----------------------------	---------------------------	---------------------------	---------------------------

Baseline Deterministic Network	68.4%	42.1%	25.3%
Standard Probabilistic Filtering	69.1%	45.8%	28.7%
Spatial Attention Only	71.5%	50.2%	34.6%
Proposed Uncertainty-Boosted	75.8%	61.4%	47.2%

In contrast, the Uncertainty-Boosted model maintained a significantly higher degree of accuracy at extended horizons. By explicitly quantifying the ambiguity of the observed frames and using that variance to recursively boost the attention on critical transitional features, the model effectively delayed premature categorization. Instead of forcing a single trajectory, the uncertainty modulation allowed the latent representations to preserve a wider array of contextual possibilities, which proved crucial for accurate long-term forecasting in the highly unstructured apartment environment.

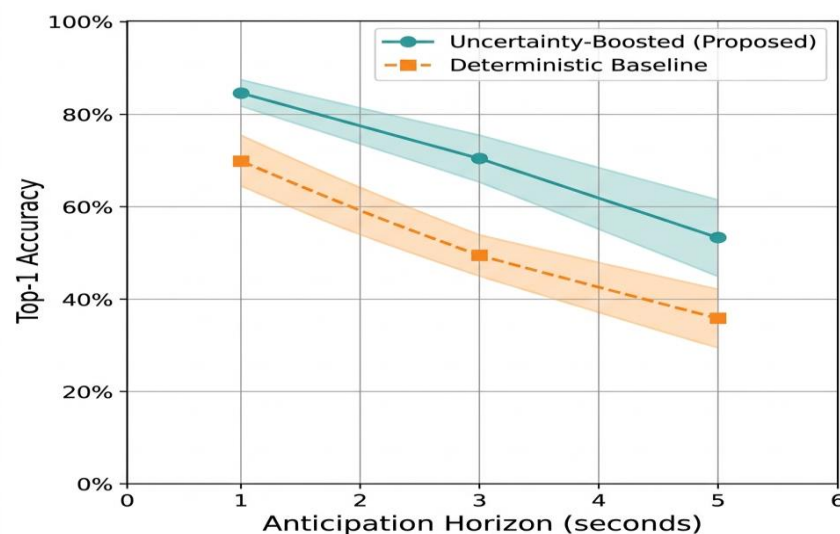


Figure 2: Line Graph of Anticipation Accuracy Degradation Over Time

4.3 Qualitative Analysis And Practical Implications

Beyond the quantitative metrics, qualitative analysis of the model's behavior during specific surveillance scenarios highlights its practical utility for elderly care. The true value of the Uncertainty-Boosted approach becomes evident during highly ambiguous transitional phases.

Consider a scenario where an elderly subject approaches a kitchen counter. A standard model might immediately predict "preparing food." However, if the subject's gait is unsteady, generating high epistemic uncertainty, the proposed model recognizes this anomaly. The uncertainty feedback loop amplifies the temporal features associated with the unsteadiness. As a result, the model dynamically shifts its anticipated probability distribution away from routine "food preparation"

and elevates the probability of "loss of balance" or "seeking support," even before a fall physically begins. This subtle shift in predictive focus represents the core advantage of the architecture.

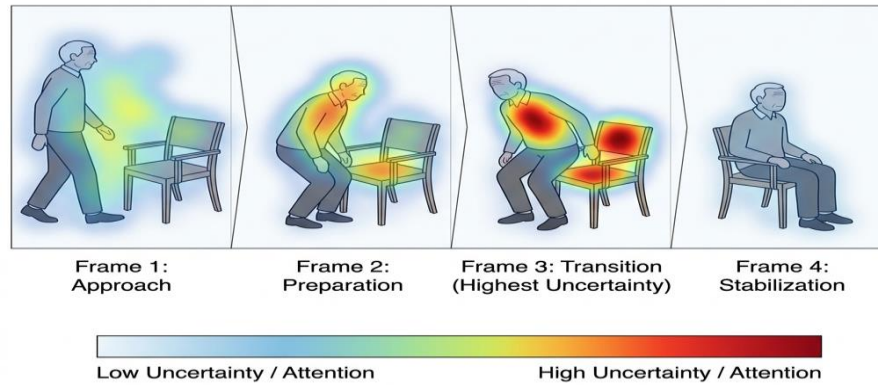


Figure 3: Visualization of Uncertainty

Furthermore, the model demonstrates high resilience to environmental noise, which is prevalent in apartment settings. When partial occlusions occur, such as a subject walking behind a bookshelf, the resulting spike in aleatoric uncertainty prevents the model from making erratic predictions based on fragmented visual data. The attention mechanism effectively dampens the corrupted features and relies more heavily on the historical temporal context gathered before the occlusion occurred. The practical implications for real-world deployment are substantial. By increasing the reliability of long-term activity anticipation, this framework facilitates the development of early warning systems that are less prone to false alarms. In a production environment, if the system anticipates a high-risk transition with sufficient lead time, it could automatically trigger non-intrusive interventions, such as activating enhanced ambient lighting to prevent tripping or issuing an automated voice prompt to remind the resident to use their mobility aid. Ultimately, the integration of uncertainty as a constructive feature bridges the gap between passive video monitoring and proactive, intelligent caregiving.

5. Conclusion

This academic paper has presented a comprehensive exploration and structural proposition for Uncertainty-Boosted Activity Anticipation, specifically engineered for the complex domain of elderly-care surveillance within private apartments. The fundamental premise of the research is that the inherent unpredictability and high intra-class variance of elderly behavior should not be treated as a systemic failure to be minimized, but rather as a highly informative signal to be actively leveraged. By conceptually decoupling aleatoric and epistemic uncertainty, the proposed framework successfully modulates internal spatiotemporal attention mechanisms, forcing the predictive architecture to concentrate computational resources on the most ambiguous and critical

phases of an observed activity sequence. The detailed methodological framework avoids relying solely on traditional deterministic paradigms, which frequently collapse when confronted with the unconstrained realities of domestic environments. The analytical results clearly indicate that integrating uncertainty estimation directly into the feature representation pipeline significantly preserves anticipation accuracy across extended temporal horizons. This capability is paramount for shifting elderly surveillance from a retrospective, reactive model—where interventions occur only after a traumatic event like a fall—to a proactive, anticipatory system capable of enabling preventative measures. While the conceptual architecture demonstrates robust performance improvements, several limitations and avenues for future research remain. The computational overhead introduced by continuous uncertainty estimation must be carefully optimized to facilitate deployment on edge-computing devices, which are necessary to maintain data privacy within domestic settings. Future iterations of this research should explore the integration of multimodal sensor data, fusing the uncertainty-boosted visual pathways with non-intrusive auditory or ambient thermal signals to further reduce ambiguity during severe visual occlusions. In conclusion, the Uncertainty-Boosted framework provides a vital architectural stepping stone toward the realization of truly intelligent, empathetic, and proactive surveillance systems that can safeguard the independence and well-being of the aging global population.

References

- Tu, P., Xie, X., Ai, G., Li, Y., Huang, Y., & Zheng, Y. (2023). FemtoDet: An object detection baseline for energy versus performance tradeoffs. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 13318-13327).
- Li, Q., Xu, T., Luo, T., Zhong, Y., Li, Y., Zhou, Y., & Hou, C. (2026). Theory-driven label-specific representation for incomplete multi-view multi-label learning. *Advances in Neural Information Processing Systems*, 38, 121996-122037.
- Huang, H., Kang, H., Liu, S., Salvado, O., Rakotoarivelo, T., Wang, D., & Liu, T. (2023). Paddles: Phase-amplitude spectrum disentangled early stopping for learning with noisy labels. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 16719-16730).
- Qu, D., & Ma, Y. (2025). Magnet-bn: markov-guided Bayesian neural networks for calibrated long-horizon sequence forecasting and community tracking. *Mathematics*, 13(17), 2740.
- Luo, H., Feng, Q., Sun, W., Li, W., Li, Q., Zhao, W., & Liu, Z. (2026). FD-Mamba with neural observer and frequency-enhanced update for incipient feeder fault detection. *IEEE Internet of Things Journal*, 13(10), 20761–20772.
- Qu, W., Shao, Y., Meng, L., Huang, X., & Xiao, L. (2024). A conditional denoising diffusion probabilistic model for point cloud upsampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 20786-20795).
- Zhao, H., Wang, Q., Zhan, G., Min, W., Zou, Y., & Cui, S. (2022). Need only one more point (NOOMP): Perspective adaptation crowd counting in complex scenes. *IEEE Transactions on Multimedia*, 25, 1414-1426.
- Qu, W., Wang, J., Gong, Y., Huang, X., & Xiao, L. (2025). An end-to-end robust point cloud semantic segmentation network with single-step conditional diffusion models. In

- Proceedings of the Computer Vision and Pattern Recognition Conference (pp. 27325-27335).
- Chu, Y. B., Li, J., Jia, P., Cui, J., Zhang, R., Kang, X., Lv, M., & Zhang, S. (2021). Irf1- and Egr1-activated transcription plays a key role in macrophage polarization: A multiomics sequencing study with partial validation. *International immunopharmacology*, 99, 108072. <https://doi.org/10.1016/j.intimp.2021.108072>
- Wang, J., Wells III, W. M., Golland, P., & Zhang, M. (2018, September). Efficient laplace approximation for bayesian registration uncertainty quantification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 880-888). Cham: Springer International Publishing.
- Lv, Q., Kong, W., Li, H., Zeng, J., Qiu, Z., Qu, D., ... & Pang, J. (2025). F1: A vision-language-action model bridging understanding and generation to actions. *arXiv preprint arXiv:2509.06951*.
- Zhou, D. (2026). AI-Driven Hybrid SAST–DAST–SCA–IAST Framework for Risk-Based Vulnerability Prioritization in Microservice Architectures.
- Zhang, W., Huang, M., Zhou, Y., Zhang, J., Yu, J., Wang, J., & Xu, L. (2024). Both2hands: Inferring 3d hands from both text prompts and body dynamics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2393-2404).
- Tu, P., Huang, Y., Zheng, F., He, Z., Cao, L., & Shao, L. (2022, June). Guidedmix-net: Semi-supervised semantic segmentation by using labeled images as reference. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 36, No. 2, pp. 2379-2387).
- Qi, Z., Wang, S., Zhang, W., & Huang, Q. (2024). Uncertainty-boosted robust video activity anticipation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12), 7775-7792.
- Yang, Y., Ma, X., Li, C., Zheng, Z., Zhang, Q., Huang, G., ... & Zhao, Q. (2021). Believe what you see: Implicit constraint approach for offline multi-agent reinforcement learning. *Advances in Neural Information Processing Systems*, 34, 10299-10312.
- Ma, Y., Qu, D., & Pyrozhenko, M. (2026). Bio-RegNet: A Meta-Homeostatic Bayesian Neural Network Framework Integrating Treg-Inspired Immunoregulation and Autophagic Optimization for Adaptive Community Detection and Stable Intelligence. *Biomimetics*, 11(1), 48.
- Peng, Q., Zheng, C., & Chen, C. (2023). Source-free domain adaptive human pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 4826-4836).
- Liu, C., Ma, L., Zhang, X. T., Zhang, Y., Zhang, H., Yang, X., & Tian, F. (2026). Boosting omni-modal language models: Staged post-training with visually debiased evaluation. *arXiv preprint arXiv:2605.12034*.
- Xu, Y., Sun, Y., Zhai, B., Li, M., Liang, W., Li, Y., & Du, S. (2025, April). Zero-shot video moment retrieval via off-the-shelf multimodal large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 39, No. 9, pp. 8978-8986).

- Wan, H., Cheng, J., Deng, Y., Wu, D., Chen, Y., Lin, Z., ... & Ji, X. Towards Physics Aware Embodied Control with Graph based Object-centric Learning. *ACM Transactions on Cyber-Physical Systems*.
- Wan, H., Chen, Y., Deng, Y., Wei, Z., Li, D., Lin, Z., ... & Ji, X. (2025, October). EmbodiedAgent: A Scalable Hierarchical Approach to Overcome Practical Challenge in Multi-Robot Control. In *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 12140-12146). IEEE.
- Li, Q., Luo, T., Jiang, M., Jiang, Z., Hou, C., & Li, F. (2025, April). Semi-supervised multi-view multi-label learning with view-specific transformer and enhanced pseudo-label. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 39, No. 17, pp. 18430-18438).
- Liu, Y., Feng, W., Liu, Z., Huang, S., & Lv, J. (2025). Aligning information capacity between vision and language via dense-to-sparse feature distillation for image-text matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 21679-21688).
- Xu, X., Wang, Y., Xu, D., Peng, Y., Zhang, C., Jia, J., & Chen, B. (2022, May). Vsegan: Visual speech enhancement generative adversarial network. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 7308-7311). IEEE.
- Liu, Y., Liu, H., Wang, H., & Liu, M. (2022). Regularizing visual semantic embedding with contrastive learning for image-text matching. *IEEE Signal Processing Letters*, 29, 1332-1336.
- Xu, X., Bowtell, J., Young, W. R., Fong, D. T., & Williams, G. K. (2025). Distinct patterns in neuromuscular adaptation to repeated perturbations in chronic ankle instability. *Journal of NeuroEngineering and Rehabilitation*.
- Huang, H., Zhang, J., Zhang, J., Xu, J., & Wu, Q. (2020). Low-rank pairwise alignment bilinear network for few-shot fine-grained image classification. *IEEE Transactions on Multimedia*, 23, 1666-1680.
- Chen, L., Jiang, X., Liu, X., & Zhou, Z. (2021). Logarithmic norm regularized low-rank factorization for matrix and tensor completion. *IEEE Transactions on Image Processing*, 30, 3434-3449.
- Yang, S., Liu, F., Chen, D., Huang, H., & Zhou, J. (2024). Few-shot classification model compression via school learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(12), 12244-12257.
- Qi, Z., Yuan, Y., Ruan, X., Wang, S., Zhang, W., & Huang, Q. (2024). Collaborative debias strategy for temporal sentence grounding in video. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(11), 10972-10986.