

## I. Related Work

### A. Small Object Detection and Resolution Limits

The fundamental bottleneck in small object detection is the scarcity of distinguishing features. When a target occupies fewer than  $15 \times 15$  pixels, class-defining details are often lost entirely [1]. Classical solutions involving super-resolution (SR) attempt to reconstruct this lost detail [2].

However, reliance on SR for scientific data collection is flawed for two reasons: correctness and latency. First, SR is fundamentally generative; it estimates high-frequency details based on learned priors, creating a risk of hallucination where the model reinforces its own biases [5]. Second, the latency advantage of SR is negligible for high-quality restoration. While lightweight models run in  $< 30\text{ms}$  on edge accelerators (e.g., Jetson Orin), high-fidelity generative models required for scientific validity often require  $> 300\text{ms}$  per frame [18]. This exceeds the mechanical slew-and-settle time of our system ( $\approx 150\text{ms}$ ), which is competitive with high-end commercial PTZ units (typically 60–200 ms command latency [16]). Our system therefore chooses the mechanical penalty to obtain optical ground truth rather than the computational penalty for potentially hallucinatory estimation.

### B. Active Acquisition vs. Continuous Tracking

Most PTZ tracking literature focuses on the control problem of keeping a target centered in the frame [7, 8]. This requires mitigating total system latency (video encoding + network + mechanical response), which for IP-based systems frequently ranges from 200–500 ms [17]. Our work addresses a distinct problem: *active acquisition*, or "slew-to-classification." Unlike continuous tracking, where the objective is persistence, our objective is information gain via discrete spot-checks.

Existing active perception systems like VIGIA-E [11] typically optimize for broad area coverage or anomaly detection. In contrast, our system functions as a sparse query mechanism. It identifies specific low-confidence candidates in the wide field and commits the PTZ resource to verifying them individually. This shifts the challenge from long-term stabilization to rapid, precise separate-and-verify maneuvers.

### C. Sensor-Driven Labeling

Reducing manual annotation is a central goal of both semi-supervised learning and active learning. Pseudo-labeling methods such as ASTOD [12] attempt to retrain models using high-confidence predictions, but this approach often fails in the small-object regime where the detector is consistently uncertain [13]. Similarly, active learning strategies like PPAL [15] identify informative samples but still require a human loop [14].

Our proposed "Active Acquisition" creates a fully automated hybrid. We use the selection logic of active learning (targeting less confident samples) but satisfy the label query using the PTZ sensor instead of a human. The success of this automated verification relies on the domain shift provided by optical zoom: while the target is ambiguous at  $15 \times 15$  pixels, the zoomed view restores it to a regime (e.g.,  $> 100 \times 100$  pixels) where off-the-shelf detectors already achieve near-perfect accuracy [6]. By physically bridging the gap between the surveillance view and the high-resolution training distribution of standard models, we convert a difficult "small object" inference problem into a trivial classification task, enabling the generation of verified ground-truth labels at scale.

## References

- [1] M. Nikouei et al., "Small Object Detection: A Comprehensive Survey on Challenges, Techniques, and Real-World Applications," *Intelligent Systems with Applications*, vol. 25, 2025.
- [2] B. Mahaur, N. Singh, and K. K. Mishra, "Road object detection: a comparative study of deep learning-based algorithms," *Multimedia Tools and Applications*, vol. 81, no. 10, pp. 14247–14282, 2022.
- [3] A. Rozantsev, V. Lepetit and P. Fua, "Detecting Flying Objects Using a Single Moving Camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 5, pp. 879–892, 2017.
- [4] G. Chen, H. Pu, W. Luo, and L. Zhang, "A Survey of the Four Pillars for Small Object Detection: Multiscale Representation, Contextual Information, Super-Resolution, and Region Proposal," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 2, pp. 936–953, 2022.
- [5] X. Zhang, Q. Chen, R. Ng, and V. Koltun, "Zoom to Learn, Learn to Zoom," *CVPR*, 2019.

- [6] “Real-Time Flying Object Detection with YOLOv8,” *arXiv preprint arXiv:2305.09972*, 2023.
- [7] “Evaluation of trackers for Pan-Tilt-Zoom Scenarios,” *arXiv preprint arXiv:1711.04260*, 2017.
- [8] “Reproducible Evaluation of Pan-Tilt-Zoom Tracking,” *ICIP*, 2015.
- [9] “Active Visual Perception Enhancement Method Based on Deep Reinforcement Learning,” *Electronics*, vol. 13, no. 9, 2024.
- [10] “Anomalous object detection by active search with PTZ cameras,” *Expert Systems with Applications*, vol. 184, 2021.
- [11] “VIGIA-E: Density-Aware Patch Selection for Efficient Video Surveillance with PTZ Cameras,” *CAIP*, 2025.
- [12] “Adaptive Self-Training for Object Detection,” *ICCVW*, 2023.
- [13] “Improving Object Detection Accuracy with Self-Training Based on Bi-Directional Pseudo Label Recovery,” *Electronics*, vol. 13, no. 12, 2024.
- [14] “Ten Years of Active Learning Techniques and Object Detection: A Comprehensive Survey,” *Applied Sciences*, vol. 13, no. 10, 2023.
- [15] “Plug and Play Active Learning for Object Detection,” *CVPR*, 2024.
- [16] “Pelco Esprit Compact PTZ Technical Specifications,” Pelco by Motorola Solutions, 2024. [Online].
- [17] “Understanding Latency in IP Video Systems,” PTZOptics White Paper, 2023. [Online].
- [18] “NVIDIA Jetson Orin Nano AI Performance Benchmarks,” NVIDIA Developer Blog, 2024. [Online].