Dual-Camera Active Acquisition for Automated Small-Object Dataset Construction

Jackie Wang[*], JC Vaught[†], Douglas Cahl[‡], Yi Wang[§]
Department of Mechanical Engineering, University of South Carolina, Columbia, SC, 29201

Deep learning performance on small objects is frequently bottlenecked by the quality and quantity of training data rather than model architecture. To address this, we propose an active dual-camera system for automated small-object dataset generation, specifically designed to overcome the resolution limits of static wide-angle surveillance. The system leverages a fixed wide-angle camera for target discovery and a PTZ unit for detailed interrogation. The control framework transitions from open-loop predictive slewing to closed-loop visual tracking, compensating for mechanical latencies and slow zoom mechanics. Implemented on an NVIDIA Jetson Orin, the system runs concurrent detector instances, one on the GPU and one on the Deep Learning Accelerator (DLA) to achieve 30 fps throughput. High-resolution object verifications are projected back into the wide frame using a static homography-based calibration, creating high-confidence labels for targets that appear as only a few pixels in the wide view. We validate the system design against a test case of distant aircraft in daylight, analyzing the trade-offs between slew speed (120°/s) and zoom settling time. Preliminary analysis suggests this active acquisition paradigm can improve label precision significantly and reduce the human effort required for small-object dataset curation.

I. Introduction

Detecting small objects in wide-area imagery remains difficult because small targets contain limited discriminative detail, are sensitive to blur and compression, and are easily confused with background clutter or visually similar distractors. Surveys of small-object detection emphasize persistent challenges from low resolution, occlusion, background interference, and class imbalance, while also documenting that improvements frequently depend on better data and scale-aware acquisition rather than architecture changes alone.

This paper addresses the data bottleneck directly. Instead of treating the camera as a passive sensor, we treat sensing as an active process in which a fixed wide-angle camera performs continuous search and a controllable PTZ camera performs confirmation and detail capture on demand. Active PTZ systems have long been studied for tracking and dynamic imaging, and modern work continues to integrate deep detectors with PTZ control for small aerial targets.

The key idea is to exploit complementary camera roles. The wide-angle stream provides persistent coverage and candidate proposals, while the PTZ camera provides a temporary, high-resolution view that increases classification confidence and stabilizes localization. Those high-quality PTZ detections are then projected back to wide-angle frames to produce labels that are spatially consistent with the deployment camera and valuable for retraining the wide-angle detector.

The airplane-in-the-sky test case is representative of many small-object regimes. Targets are frequently fewer than 20–40 pixels in height at wide focal lengths; motion is largely angular with occasional accelerations; backgrounds vary with clouds, haze, sun glare, and camera sensor artifacts. These conditions align with published analyses of small/tiny object detection challenges and with prior work on detection of flying objects using YOLO-family models.

[*] Graduate Student, Department of Mechanical Engineering, Member AIAA.
[†] Graduate Student, Department of Mechanical Engineering, Member AIAA.
[‡] Professor, Department of Mechanical Engineering, Member AIAA.
[§] Professor, Department of Mechanical Engineering, Senior Member AIAA.

II. Related Work

A. Small Object Detection

Recent surveys synthesize methods for small object detection, highlighting multi-scale feature processing, super-resolution or enhancement, context modeling, and data-centric strategies such as targeted acquisition and hard-negative mining.

For airborne targets, YOLO-family detectors are commonly chosen for their speed–accuracy trade-off, and recent work explicitly evaluates YOLOv8-based pipelines for flying object detection.

B. PTZ Tracking and Active Vision

PTZ tracking differs from fixed-camera tracking because the camera motion changes the image formation process and induces highly dynamic backgrounds. Standardized evaluation of PTZ trackers and benchmark methodologies have been studied, emphasizing control latency, re-centering accuracy, and zoom stability.

Modern active vision approaches incorporate deep detection modules into PTZ decision-making for improved observation quality, including PTZ-assisted perception pipelines and PTZ-guided small-object detection strategies.

For small aerial targets, recent systems explicitly coordinate wide sensing with PTZ imaging and tracking to keep fast, erratically moving targets in view.

C. Pseudo-Labeling, Self-Training, and Active Learning

Pseudo-labeling and self-training reduce annotation costs by selecting high-confidence predictions as training targets, but they require careful thresholding and quality control to avoid drift. Adaptive thresholding strategies for pseudo-label selection have been proposed to reduce manual tuning, and recent object detection self-training methods emphasize robust selection mechanisms.

Active learning for object detection studies how uncertainty and diversity can guide sample selection for labeling, including plug-and-play strategies that combine uncertainty- and diversity-based phases.

This paper differs in that the "labeling oracle" is not a human annotator but a sensor action: the PTZ zoom. The acquisition action produces a higher-fidelity view that improves label quality and reduces ambiguity without manual annotation.

III. System Overview

A. Hardware and Software Roles

The system uses two cameras mounted with a fixed relative pose (Fig. 1):

(1) a fixed wide-angle camera providing continuous coverage and running a real-time YOLO-family detector;

(2) a PTZ camera whose pan, tilt, and zoom can be commanded via a control interface and whose telemetry is logged with timestamps.

The wide camera produces candidate detections and tracklets. A scheduling policy selects targets and commands the PTZ to re-center and zoom. The PTZ stream is analyzed to confirm class and refine bounding boxes. High-confidence PTZ detections are transferred back to the wide-angle frame to create dataset labels. A curation module filters low-quality frames and controls redundancy.

B. End-to-End Pipeline Diagram

IV. Problem Formulation

Let I_t^w be the wide-angle frame at time t, and let I_t^p be the PTZ frame captured at time t with telemetry $\tau_t = (\theta_t, \phi_t, z_t)$ representing pan, tilt, and zoom. The wide detector produces candidate detections $D_t = \{(b_{t,i}^w, s_{t,i}^w, c_{t,i})\}_i$ with bounding boxes $b_{t,i}^w$, confidence scores $s_{t,i}^w$, and class labels $c_{t,i}$.

The system chooses actions a_t that either keep scanning (no PTZ) or command the PTZ to acquire a zoomed observation centered on a selected candidate. The goal is to maximize the expected dataset value subject to control and

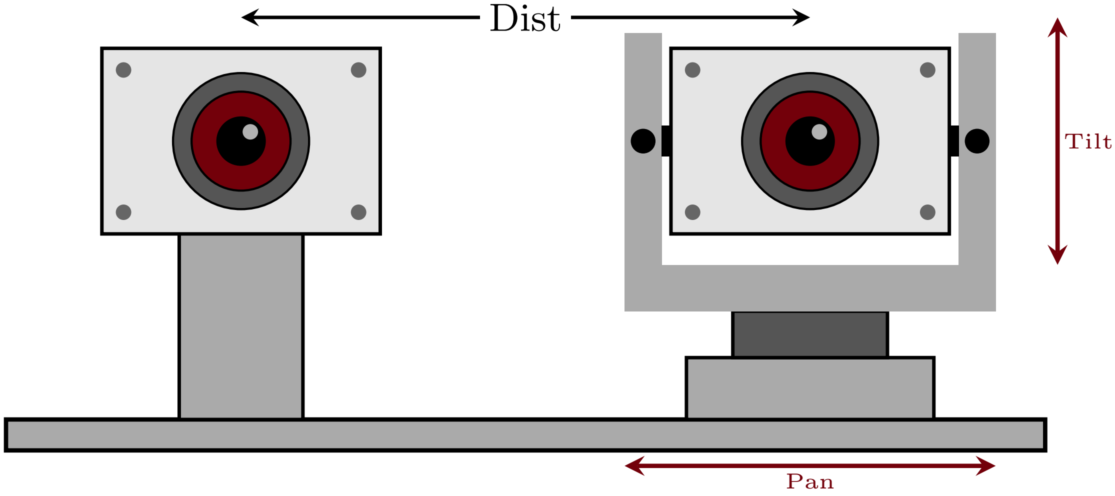


Fig. 1 Physical hardware configuration. A fixed wide-angle camera provides persistent coverage while a 2-axis PTZ camera can be commanded to lock onto and zoom in on candidate targets.

compute constraints:

$$\max_{\pi} \mathbb{E}\left[\sum_{t} U\left(I_t^w, I_t^p, \tau_t\right)\right] \tag{1}$$

subject to PTZ dynamics, latency, bandwidth, and storage budgets. Here $U(\cdot)$ is an acquisition utility that increases with (i) label correctness probability, (ii) localization quality, (iii) novelty/diversity, and (iv) usefulness for training (hard examples and hard negatives).

V. Calibration and Cross-View Geometry

A. Camera Model

Both cameras are modeled with intrinsics (K_w, K_p) and distortion parameters. Let $\Pi(\cdot)$ be perspective projection with distortion, and let R_{pw}, t_{pw} map 3D points from wide-camera coordinates to PTZ-camera coordinates.

For sky targets at long range, parallax between cameras is often negligible if the baseline is small relative to range. In that regime, direction-only transfer is effective: rays from each camera correspond to the same world direction, and cross-view mapping can be performed using ray directions on the unit sphere rather than estimated depth.

B. Mapping Wide Detections to PTZ Pan/Tilt Commands

Let (u, v) be the center of a wide detection box b^w in pixel coordinates. After undistortion, the bearing direction in the wide camera frame is

$$\hat{d}_w = \frac{K_w^{-1}[u\ v\ 1]^\top}{\|K_w^{-1}[u\ v\ 1]^\top\|}. \tag{2}$$

Transforming to the PTZ base frame gives $\hat{d}_p = R_{pw}\hat{d}_w$. Using a conventional yaw–pitch parameterization, the commanded pan (yaw) and tilt (pitch) are

$$\theta = \operatorname{atan2}(\hat{d}_{p,y}, \hat{d}_{p,x}), \quad \phi = \operatorname{atan2}(\hat{d}_{p,z}, \sqrt{\hat{d}_{p,x}^2 + \hat{d}_{p,y}^2}). \tag{3}$$

Because PTZ motion and video pipelines have latency, $\hat{d}_p$ is preferably computed from a short-horizon prediction of target motion rather than the instantaneous detection center.

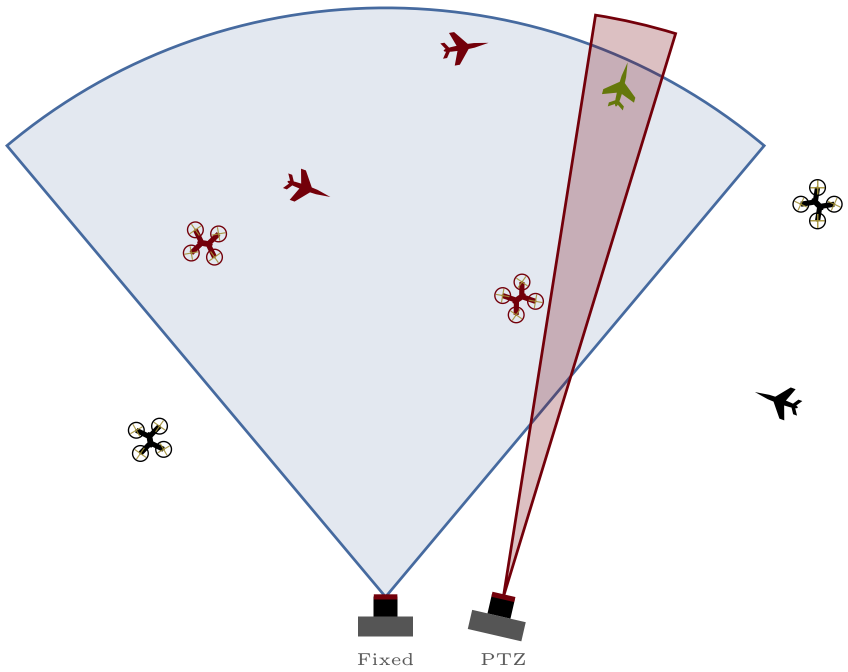


Fig. 2 Field of view comparison. The wide camera (blue) covers approximately $90°$ for persistent surveillance, while the PTZ camera (red) provides a narrow, steerable, high-resolution view that can be directed to any point within the wide FOV.

C. Zoom Selection by Target Angular Size

Let h^w be the detected box height in wide pixels. Under small-angle approximation, the apparent angular height is approximately $\alpha \approx h^w / f_w$, where f_w is the wide focal length in pixels. To obtain a desired PTZ pixel height h^p_{des}, select a PTZ focal length $f_p(z)$ such that

$$h^p_{\text{des}} \approx \alpha f_p(z) \approx \frac{h^w}{f_w} f_p(z), \tag{4}$$

then choose the zoom z whose calibrated $f_p(z)$ best matches the required value, clipped to PTZ limits. This approach avoids explicit range estimation and is well suited for distant aircraft.

VI. PTZ Triggering, Tracking, and Scheduling

PTZ tracking is a coupled perception-and-control problem with dynamic imaging conditions and control delays. Prior PTZ tracking evaluations emphasize that camera motion, latency, and re-centering quality dominate performance differences in practice.

We use wide-camera tracklets to provide temporal coherence and to predict target bearing during PTZ motion. A lightweight predictor (e.g., constant angular velocity with α–β filtering or a Kalman filter) provides a predicted bearing $\hat{d}_w(t + \Delta)$ given command latency Δ, consistent with classical pan/tilt tracking formulations.

A. Utility Function for Triggering and Redundancy Control

The trigger utility balances value and cost. A practical form is a weighted sum of terms:

$$U(o) = \lambda_1 \,\text{Unc} + \lambda_2 \,\text{SizeGain} + \lambda_3 \,\text{Novelty} - \lambda_4 \,\text{Cost}, \tag{5}$$

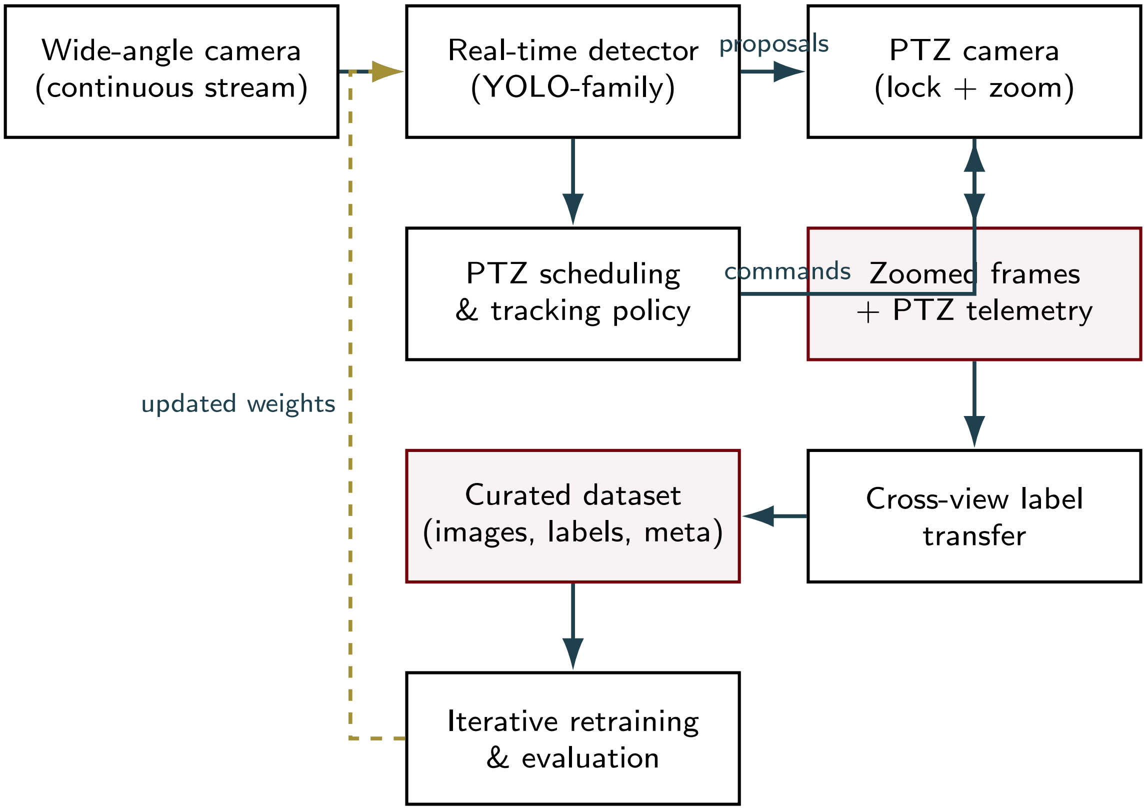


Fig. 3 Dual-camera active acquisition loop. The PTZ camera is triggered by wide-angle detections to improve evidence quality; high-confidence PTZ detections are transferred back to wide-angle frames.

where Unc increases when the wide detector is unsure (so PTZ confirmation is valuable), SizeGain estimates how much the PTZ will increase target pixels, Novelty reduces redundancy by down-weighting near-duplicates, and Cost captures PTZ time-on-target, slew distance, and opportunity cost (only one PTZ target at a time).

This framing is consistent with object-detection active learning literature that combines uncertainty and diversity, although here the action is a sensor query rather than a human label request.

VII. Automated Dataset Construction

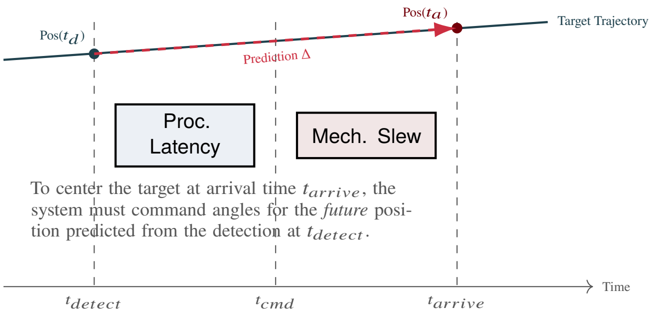
A. Label Sources and Cross-View Transfer

The PTZ view is treated as a higher-fidelity label source when it produces a confirmed detection for the target class. The label transfer module maps the PTZ detection back into the wide image coordinate system.

For distant targets, direction-only transfer proceeds by converting the PTZ bounding box corners into bearing rays, rotating those rays into the wide camera frame, and projecting them into the wide image plane. Let (u_k^p, v_k^p) be PTZ box corners; compute PTZ rays $\hat{d}_k^p$ from K_p^{-1}, rotate to wide frame $\hat{d}_k^w = R_{wp}\hat{d}_k^p$, then project:

$$[u_k^w, v_k^w, 1]^\top \propto K_w \hat{d}_k^w, \quad k \in \{1, 2, 3, 4\}. \tag{6}$$

The transferred wide box is the tight axis-aligned rectangle enclosing the projected corners. This procedure uses only calibrated intrinsics and the relative rotation, plus the PTZ telemetry that defines the effective PTZ optical axis at capture time.



To center the target at arrival time t_{arrive}, the system must command angles for the *future* position predicted from the detection at t_{detect}.

Fig. 4 Latency Compensation. The PTZ command must account for processing delays and mechanical slew time. The target position is extrapolated Δ seconds into the future to ensure the PTZ arrives at the correct bearing.

Algorithm 1 Real-time wide-to-PTZ acquisition

```
Init $f_\theta$ (wide), $\mathcal{T}$ (tracker), $g$ (PTZ), $\mathcal{D}$ (data)
for each wide frame $I_t^w$ do
    $D_t \leftarrow f_\theta(I_t^w)$;    $\mathcal{T} \leftarrow \text{UpdateTracks}(\mathcal{T}, D_t)$
    Compute $U(o)$ for tracks (conf, size, novelty)
    Select $o^\star \leftarrow \arg\max U(o)$ s.t. availability
    if $U(o^\star) > \tau_{\text{trigger}}$ then
        Predict $\hat{d}_w(t + \Delta)$; map to $(\theta, \phi, z)$
        Command PTZ: $g(\theta, \phi, z)$; get $I_{t'}^p$, $\tau_{t'}$
        Run PTZ detector: $D_{t'}^p \leftarrow f_{\theta_p}(I_{t'}^p)$
        if Quality OK AND max $s^p > \tau_{\text{confirm}}$ then
            $b^{w \leftarrow p} \leftarrow \text{Project}(b^p, \tau_{t'})$
            Commit $(I_t^w, b^{w \leftarrow p}, c)$ to $\mathcal{D}$
        end if
    end if
end for
Periodically retrain $f_\theta$ on $\mathcal{D}$
```

B. Quality Gates and Drift Prevention

Self-training and pseudo-labeling can suffer from confirmation bias if low-quality pseudo-labels are admitted. This is widely recognized in semi-supervised detection; adaptive thresholding and robust pseudo-label selection reduce manual tuning and improve stability.

In this system, the primary drift control mechanism is the physical zoom verification: many ambiguous wide detections become unambiguous at higher resolution. Remaining failure modes are handled by quality gates (Fig. 5) that reject samples when motion blur is excessive, exposure is saturated (e.g., sun glare), the target is at the frame boundary, or the PTZ detector disagrees with the wide detector class in a way indicative of confusion (e.g., airplane vs bird). Each gate is implemented as a deterministic predicate so that dataset inclusion is reproducible and auditable.

C. Dataset Schema

Each committed example stores the wide frame (or short clip), the transferred label, and acquisition metadata. Table 1 defines a minimal schema sufficient for training and analysis.

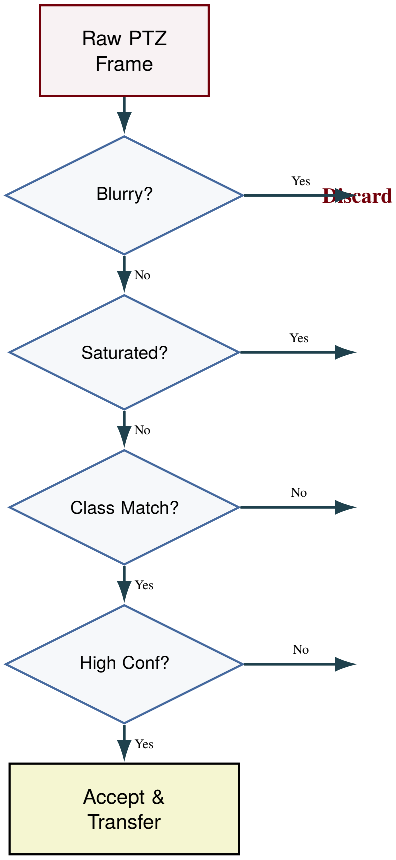


Fig. 5 Quality Gate Logic. To prevent dataset contamination (drift), PTZ frames must pass a series of deterministic checks for image quality (blur, saturation) and content verification before being used to generate labels.

VIII. Airplanes-in-the-Sky Test Case

A. Operating Conditions and Target Characteristics

Airplanes introduce systematic small-object issues. In a wide-angle view, aircraft can be extremely small with strong scale variation as they traverse the sky and change apparent altitude and distance (Fig. 6). Visual appearance changes with viewing angle, contrails, lighting, haze, and compression artifacts, matching known difficulty factors for small objects (low detail, interference, background ambiguity).

Flying-object detection has been studied with YOLOv8 as a practical real-time architecture choice, and YOLO-family surveys emphasize why these models are frequently deployed in resource-constrained real-time settings.

B. Class Taxonomy and Negatives

For this test case, the primary class is *airplane*. Hard negatives arise from birds, insects near the lens, distant drones, clouds with sharp edges, sensor noise, and specular highlights (Fig. 7). The PTZ confirmation step naturally collects informative negatives: wide proposals that are rejected by PTZ as non-airplane can be stored as hard negatives with contextual metadata.

IX. Evaluation Protocol

This paper describes a system design and an evaluation plan intended to be executed on a real deployment. The core evaluation objective is to measure whether PTZ-assisted acquisition yields higher-quality labels and better downstream small-object performance than passive wide-only collection.

Table 1 Minimal dataset record schema for each accepted acquisition.

Field	Description
timestamp_w	Wide-frame timestamp (monotonic)
image_w	Wide image (or video clip key)
bbox_w	Label box in wide coordinates
class	Target class (airplane for the test case)
score_w	Wide detector confidence at time t
timestamp_p	PTZ-frame timestamp used for confirmation
telemetry_p	PTZ pan/tilt/zoom, focus, exposure
bbox_p, score_p	PTZ detector box and confidence
quality_metrics	Blur, saturation, occlusion flags
site_meta	Location, camera orientation, weather

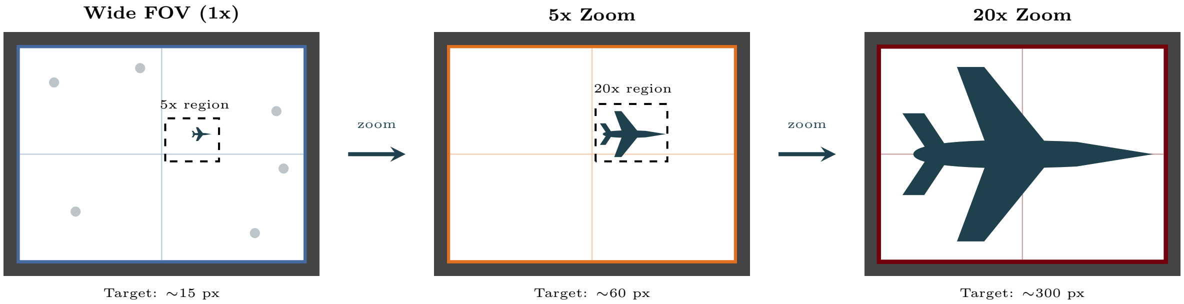


Fig. 6 Zoom level comparison. At wide-angle (left), a distant airplane may occupy only 15 pixels—insufficient for reliable classification. Progressive zoom increases pixel count, enabling confident detection at 20x zoom (right).

A. Baselines

The following baselines support attribution of improvements to PTZ zoom verification rather than to data volume alone:

A wide-only baseline that logs candidate detections from the wide camera without PTZ confirmation;

a PTZ patrol baseline that performs a scripted scan independent of wide detections;

a manual-zoom oracle baseline on a small audited subset, used only to estimate upper bounds and error modes.

B. Metrics

Label quality is assessed on an audited subset using human review or higher-resolution reference footage. Primary metrics include (i) label precision and recall for accepted samples, (ii) bounding-box IoU between transferred labels and audited labels, (iii) confidence uplift $\Delta s = s^p - s^w$ for confirmed samples (Fig. 8), and (iv) downstream detector performance (e.g., small-object mAP on wide-angle imagery) after training on the constructed dataset.

Because PTZ tracking introduces dynamics and latency, system metrics include slew-to-lock time, dwell time per target, fraction of triggers that successfully reacquire the target, and PTZ availability contention (fraction of time PTZ is busy). PTZ tracking literature suggests these measures are essential to understanding real-world performance beyond static detection accuracy.

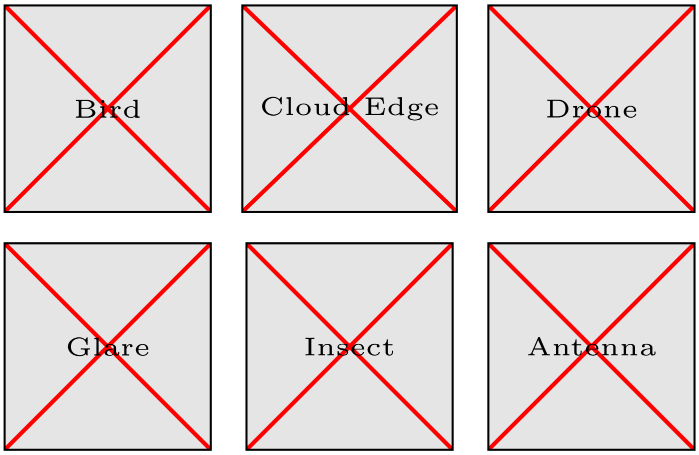


Fig. 7 Hard negative examples. Common false positives in the wide camera include birds, cloud edges, drones, sun glare, insects near the lens, and distant antennas. PTZ verification rejects these as non-airplane.

C. Ablation Factors

Ablations should isolate contributions from (i) trigger thresholds and hysteresis, (ii) motion prediction vs reactive centering, (iii) zoom scheduling, (iv) label transfer method (direction-only vs depth-aware if depth is available), and (v) pseudo-label admission thresholds. Adaptive pseudo-label thresholding methods provide a useful reference point for designing these ablations.

X. Implementation Considerations

A. Real-Time Constraints

The wide-angle detector must run at frame rate to maintain coverage. YOLO-family models are commonly selected for this regime; Ultralytics documentation notes YOLOv8 release and positioning as a speed–accuracy option for detection tasks, and survey work summarizes the evolution of YOLO variants and real-time deployment considerations.

The PTZ control loop must tolerate latency in (i) wide detection, (ii) command transmission, (iii) mechanical motion, and (iv) PTZ video encoding/decoding. A practical design uses a bounded queue for PTZ commands, drops stale commands, and prioritizes keeping the target near the PTZ image center over maximizing instantaneous zoom.

B. Telemetry Synchronization

Accurate label transfer requires time alignment between wide frames, PTZ frames, and PTZ telemetry. The system should log monotonic timestamps at capture and at inference, and it should record the PTZ telemetry state at the exact time a PTZ frame is captured (or as close as the interface permits). If telemetry is sampled asynchronously, interpolation to the frame time reduces geometric error.

C. Safety and Privacy

The airplane test case naturally focuses on sky regions; nonetheless, deployment should constrain PTZ tilt limits and define privacy-preserving regions to avoid capturing ground-level imagery. These constraints can be implemented at the control layer by rejecting commands that enter prohibited zones.

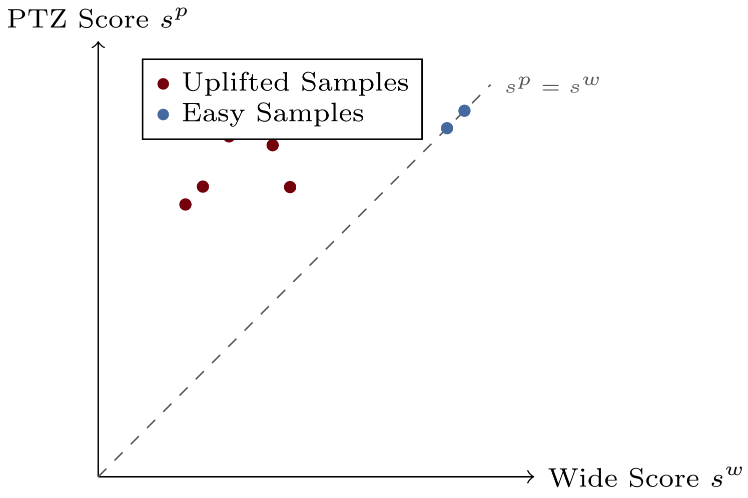


Fig. 8 Confidence uplift visualization. Points above the diagonal $s^p = s^w$ indicate samples where PTZ verification increased detection confidence. Low-confidence wide detections (left region) can achieve high PTZ confidence, validating the zoom-based verification approach.

XI. Discussion

A. When PTZ Verification Helps Most

PTZ verification is most valuable when the wide detector frequently encounters ambiguous, small candidates whose pixel support is insufficient for confident classification. In such regimes, zoom provides additional evidence without changing the wide camera that will ultimately run the detector. This can be especially beneficial when the deployment environment differs from training data and when pseudo-labeling would otherwise be unreliable.

The approach is aligned with modern PTZ-assisted perception systems that explicitly integrate deep detection with PTZ imaging to improve small-target observability.

B. Limitations

The system is constrained by single-PTZ availability and cannot zoom multiple targets simultaneously. Fast-moving targets may leave the PTZ field of view during slew, particularly when latency is high or the wide tracker is unstable. Atmospheric turbulence and haze can reduce the effective benefit of zoom. Direction-only label transfer assumes small parallax; large baselines or nearer targets require depth-aware mapping.

C. Extensions

Multi-PTZ configurations can reduce contention and increase throughput. Joint scheduling across targets can be formulated as a knapsack-like selection over predicted utilities. More advanced multi-view techniques may reduce reliance on explicit calibration in some settings, although the sky-target scenario is favorable for calibration-based geometry due to strong distance scale separation.

XII. Conclusion

This paper presented a dual-camera active acquisition method for fully automated small-object dataset construction using a fixed wide-angle camera with real-time detection and a PTZ camera for zoom-based verification. For airplanes in the sky, the PTZ view provides higher-resolution evidence that can be transferred back to wide-angle frames to produce higher-quality labels, hard negatives, and metadata. We detailed calibration and control formulations, label transfer procedures, and drift-control gates grounded in pseudo-labeling and PTZ tracking insights from the literature. The proposed evaluation protocol measures label quality, confidence uplift, and downstream small-object performance, enabling rigorous assessment of whether sensor-driven "auto-labeling" can replace or substantially reduce manual annotation in small-object regimes.

References

[1] "Small Object Detection: A Comprehensive Survey on Challenges, Techniques and Real-World Applications," *Intelligent Systems with Applications*, 2025. doi: 10.1016/j.iswa.2025.200561

[2] "Small object detection in diverse application landscapes," *Multimedia Tools and Applications*, vol. 83, pp. 40321–40359, 2024. doi: 10.1007/s11042-024-18866-w

[3] "Deep learning-based detection from the perspective of small or tiny objects: A comparative review," *Image and Vision Computing*, vol. 117, p. 104813, 2022. doi: 10.1016/j.imavis.2022.104439

[4] "Real-Time Flying Object Detection with YOLOv8," *arXiv preprint arXiv:2305.09972*, 2023. doi: 10.48550/arXiv.2305.09972

[5] "A Comprehensive Review of YOLO: From YOLOv1 to YOLOv8 and Beyond," *arXiv preprint arXiv:2304.00501*, 2023. doi: 10.48550/arXiv.2304.00501

[6] "Evaluation of trackers for Pan-Tilt-Zoom Scenarios," *arXiv preprint arXiv:1711.04260*, 2017. doi: 10.48550/arXiv.1711.04260

[7] "Reproducible Evaluation of Pan-Tilt-Zoom Tracking," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2015. doi: 10.1109/ICIP.2015.7351162

[8] "Active Visual Perception Enhancement Method Based on Deep Reinforcement Learning," *Electronics*, vol. 13, no. 9, p. 1654, 2024. doi: 10.3390/electronics13091654

[9] "Anomalous object detection by active search with PTZ cameras," *Expert Systems with Applications*, vol. 184, p. 115150, 2021. doi: 10.1016/j.eswa.2021.115150

[10] "VIGIA-E: Density-Aware Patch Selection for Efficient Video Surveillance with PTZ Cameras," in *Proceedings of the 25th International Conference on Computer Analysis of Images and Patterns (CAIP)*, 2025. doi: 10.1007/978-3-032-04968-1_23

[11] "Adaptive Self-Training for Object Detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2023. doi: 10.1109/ICCVW60793.2023.00102

[12] "Improving Object Detection Accuracy with Self-Training Based on Bi-Directional Pseudo Label Recovery," *Electronics*, vol. 13, no. 12, p. 2230, 2024. doi: 10.3390/electronics13122230

[13] "Ten Years of Active Learning Techniques and Object Detection: A Comprehensive Survey," *Applied Sciences*, vol. 13, no. 10, p. 6181, 2023. doi: 10.3390/app13169110

[14] "Plug and Play Active Learning for Object Detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. doi: 10.1109/CVPR52733.2024.01684