Depth-Based Attitude Anomaly Detection for Small UAVs Using a Convolutional Autoencoder for Health Inspection

1st Yoshinobu Tanno
Faculty of Information Design
Tokyo Information Design Professional University
Tokyo, Japan
tanno@tid.ac.jp

2nd Toshiaki Hirata
Faculty of Information Design
Tokyo Information Design Professional University
Tokyo, Japan
hirata@tid.ac.jp

3rd Yasuo Matsuoka
DPMSs (Drone Predictive Maintenance System & Services) LLC
Kawasaki, Japan
y_matsuoka@dpms.co.jp

4th Yutaka Matsuoka
DPMSs LLC
Kawasaki, Japan
syuuzou01@dpms.co.jp

***Abstract*—We propose an anomaly detection framework for small unmanned aerial vehicles (UAVs) aimed at automated preflight inspection. The system uses depth imaging from an external camera, requiring no onboard sensors to capture attitude during hovering flight. A simple three-axis viewpoint correction—X/Y ROI stabilization and Z-axis normalization with a session-wise anchor and a globally fixed depth range—ensures consistent representation across sessions. A convolutional autoencoder (CAE) trained on normal hovering data detects deviations using an anomaly score. Experiments with payloads of 0–39 g demonstrate that the CAE identifies both subtle and heavy load-induced anomalies, including nonmonotonic responses linked to flight-controller compensation. The framework provides a nuanced health assessment beyond a binary classification by treating the anomaly score as a continuous diagnostic score. This approach provides a practical solution for UAV fleet safety and maintenance.**

***Index Terms*—Small UAVs, Convolutional autoencoder, Depth imaging, Attitude monitoring**

I. Introduction

The rapid expansion of unmanned aerial vehicle (UAV) applications across diverse industries, including logistics, precision agriculture, infrastructure inspection, and emergency response, has created an urgent need for automated health-monitoring systems. UAVs are critical components of industrial workflows. Ensuring their operational reliability through predictive maintenance is paramount for both safety and economic efficiency. Traditional inspection methods that rely heavily on manual examinations and scheduled maintenance intervals are becoming increasingly inadequate for managing large UAV fleets operating in complex environments.

Attitude stability during flight is a fundamental indicator of UAV health. Here, attitude refers to roll, pitch, and yaw during hovering flight. Attitude deviations can signal various underlying issues, including mechanical wear, motor imbalances, structural damage, and excessive payload stress. Early detection of such anomalies enables proactive maintenance interventions, potentially preventing catastrophic failures and extending the equipment lifespan. However, developing robust attitude-monitoring systems faces several challenges: the dynamic nature of flight conditions, subtle manifestation of early-stage degradation, and the practical difficulty of obtaining comprehensive labeled failure data for supervised learning approaches.

We address these challenges by proposing an unsupervised anomaly detection framework based on convolutional autoencoders (CAEs) trained exclusively on normal operational data. The proposed approach leverages depth imaging technology to capture rich spatial information about UAV attitude from a fixed front-oblique viewpoint. In particular, the system acquires flight data nonintrusively using an external camera without attaching additional sensors to the UAV, complying with operational restrictions in real-world settings. By learning to reconstruct normal attitude patterns, the CAE produces elevated anomaly scores when anomalous states are encountered, thereby effectively flagging potential issues without requiring explicit failure examples during training.

Depth sensing offers several advantages over traditional RGB imaging, including invariance to lighting conditions, direct measurement of three-dimensional (3D) structures, and reduced sensitivity to surface textures or colors. Furthermore, we can leverage the feature-extraction capabilities of convolutional neural networks originally developed for color image processing by transforming depth maps into pseudo-color representations.

In terms of flight patterns, although UAVs are capable of hovering, rotational, and translational maneuvers, this study specifically focused on hovering. During the hovering experiments, a 3D viewpoint correction was applied to ensure accurate capture of attitude dynamics. These design choices align

the framework with the practical goal of automated preflight inspection, thereby offering a scalable and deployment-ready solution for UAV fleet health management.

Our primary contributions are threefold: (1) we develop a comprehensive depth-based attitude anomaly detection framework for UAV health monitoring that uses only externally mounted depth sensors and requires no onboard sensors; (2) we demonstrate the capability of the system to detect a spectrum of load-induced attitude deviations without requiring labeled anomaly data; and (3) we provide a detailed analysis of the relationship between anomaly score patterns and flight controller compensation mechanisms, offering insights into UAV behavioral dynamics under stress conditions.

II. Related Work

A. Evolution of UAV Attitude Monitoring Technologies

UAV attitude monitoring has advanced from inertial measurement unit (IMU)/global positioning system (GPS)-based stabilization to vision and multisensor methods. Scaramuzza and Fraundorfer [1] provided a seminal tutorial on visual odometry (VO), and Achtelik et al. [2] demonstrated the fusion of IMU and monocular vision for micro aerial vehicle (MAV) navigation in GPS-denied environments. Forster et al. [3] proposed semi-direct visual odometry (SVO) to achieve faster and more accurate pose estimation, although vision-based methods remain vulnerable to blurring, illumination changes, and computational load. Multisensor fusion can enhance robustness but often introduces cost and complexity. These challenges have motivated the use of simple depth sensing combined with unsupervised learning. Unlike many prior UAV attitude monitoring approaches that rely on onboard sensors, our framework explicitly uses an external depth camera and requires no onboard instrumentation, aligning with pre-flight health inspection settings.

B. Deep Learning Approaches to Anomaly Detection

Deep learning (DL) is crucial to anomaly detection, particularly when anomalies are rare. Pang et al. [4] and Chalapathy and Chawla [5] conducted surveys that emphasized unsupervised methods. Autoencoders learn the compact representations of normal data using a reconstruction-based anomaly score. Bergmann et al. [6] introduced a teacher–student framework that improved performance but increased complexity. A single-model approach is sufficient and practical for monitoring the UAV attitude.

C. Depth Sensing in Computer Vision Applications

Depth sensors have proven effective in motion and scene analyses. Hara et al. [7] demonstrated a strong action recognition performance using 3D residual networks. Kwolek and Kepski [8] combined depth maps with accelerometer data to achieve a reliable fall detection. For UAV applications, Oleynikova et al. [9] developed Voxblox, an incremental 3D mapping system for MAVs using onboard depth sensors. Although these studies highlighted the versatility of depth data (including RGB-D modalities), their application to UAV health monitoring remains limited. Compared with RGB-D inputs that increase compute and memory footprint, we adopt a depth-only pseudo-color (three-channel) representation, yielding a lighter and more deployment-friendly pipeline.

D. Industrial Anomaly Detection Systems

However, industrial deployment has encountered some practical challenges. Hirata et al. [10], [11] applied CAE-based anomaly detection to air-conditioning systems and demonstrated the importance of preprocessing, domain-specific model design, and interpretable outputs. Real-world operations also require the handling of sensor drift, seasonal variations, and false alarms. These lessons inform the proposed framework's focus on robustness and usability.

E. Research Gap and Our Contribution

Although UAV monitoring, deep anomaly detection, and depth sensing have advanced, their integration is limited. UAV health monitoring largely relies on telemetry or labeled failure data, whereas vision research focuses on navigation and mapping. We addressed this gap by applying unsupervised DL to depth-based attitude monitoring. The proposed method is lighting-invariant and does not require collecting failure data, and it provides interpretable anomaly score patterns for practical maintenance.

III. System Architecture and Design

A. Overall System Design Philosophy

The core of the proposed approach centers on learning a manifold of normal attitude patterns through unsupervised training using healthy small UAV data. During the operation, deviations from this learned manifold manifested as increased anomaly scores, providing a continuous measure of the anomaly severity. This continuous scoring approach offers several advantages: it supports a continuous, context-aware interpretation of anomaly severity, enables early warning of developing issues, and provides diagnostic value by indicating the magnitude of deviation.

B. Data Acquisition and Preprocessing

As an overall design, we adopted an external, fixed-geometry, depth-only configuration for preflight attitude monitoring that is, in principle, applicable to multiple flight patterns (hovering, rotations about the vertical Z axis, and short-axis-aligned translations along x, y, and z). We focused on hovering to avoid the influence of large translations and rotations and to clearly observe load-induced attitude changes.

Three-axis viewpoint correction. (X/Y) During the initial 10 s calibration (300 frames), we computed the image-plane centroid of the UAV at each frame and obtained the average $(\bar{c}_x, \bar{c}_y)$. We then used a fixed region of interest (ROI) of 250×150 pixels centered at $(\bar{c}_x, \bar{c}_y)$ for all subsequent frames. This fixed-ROI aligns the initial viewpoint by centering the crop at the calibration mean and reduces sensitivity to session-to-session take-off offsets. We do not recenter per frame; therefore, in-session x/y translations remain in the input and

are analyzed by the model. (Z) Depth values are normalized by anchoring at the session-wise minimum valid depth of the UAV body and by applying a globally fixed normalization width computed once from the 0 g sessions (details below), which acts as an out-of-plane (range) correction.

To capture the UAV's hovering attitude, we used an Intel RealSense D455 depth camera installed externally in a fixed front-oblique view (with no onboard sensors) and recorded at 30 Hz with a resolution of 848×480.

The raw depth frames were colorized with fixed mapping for reproducibility, and the resulting three-channel pseudo-color heatmaps were used as the CAE input. The colorizer parameters were fixed across sessions such that the same depth was always mapped to the same RGB triplet, maintaining per-channel intensity distributions comparable and avoiding domain shift.

We deliberately adopted pseudo-colorized depth-only (three-channel) inputs instead of RGB-D to reduce computational and memory costs and to simplify the sensor configuration (see also Sec. II-C).

The depth maps were normalized with a session-specific anchor and a globally fixed width to ensure consistency across payload conditions and sessions. Let $d_{\min}^{(s)}$ be the minimum valid depth of the UAV body during session s. We computed the fixed range width once from all 0 g sessions as

$$w_{\text{fix}} = \left(\max_{s\in\mathcal{S}_{0g}} d_{\min}^{(s)} - \min_{s\in\mathcal{S}_{0g}} d_{\min}^{(s)}\right) + t_{\text{body}},$$

where t_{body} denotes the physical thickness of the UAV. In our setup, this yielded $w_{\text{fix}} \approx 0.298\,\text{m}$. Each frame was then normalized by

$$d' = \text{clip}\left(\frac{d - d_{\min}^{(s)}}{w_{\text{fix}}},\, 0,\, 1\right),$$

thereby maintaining a consistent scale (global width) while adapting to small takeoff offsets (session-wise anchor). This approach maintains color mapping focused on UAV attitude while avoiding background-driven scaling. Invalid (zero) depth pixels were excluded from all calculations.

Data recording began only after the calibration phase determined a fixed ROI.

As shown in Fig. 1, the session-wise anchor plus global-width normalization, fixed-parameter pseudo-colorization, and fixed-ROI stabilization produced consistent three-channel inputs across sessions.

C. Convolutional Autoencoder Architecture

The CAE architecture was carefully designed to balance the representational capacity with the generalization ability. The encoder comprised five convolutional layers with progressively increasing filter counts (64, 64, 128, 128, and 256), each followed by rectified linear unit activation.

The decoder mirrors the encoder structure with transposed convolutions for upsampling, maintaining the architectural symmetry that facilitates gradient flow during training.

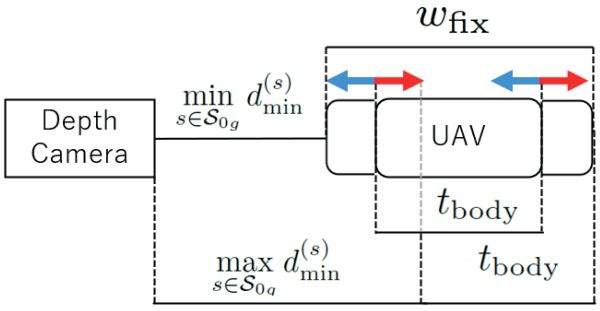


Fig. 1. Preprocessing and three-axis viewpoint-correction pipeline before CAE inference.

The loss function is mean squared error computed pixel-wise between the input and output (reconstruction).

The anomaly score for a given input frame was computed as follows:

$$S = \frac{1}{N}\sum_{i=1}^{N} \|x_i - \hat{x}_i\|^2 \tag{1}$$

where x_i and $\hat{x}_i$ represent the original and reconstructed pixel values, respectively, and N denotes the total number of pixels.

IV. Experimental Methodology

A. Experimental Platform and Setup

For our experiments, we used a DJI Mini 4 Pro, which is a widely available consumer-grade quadrotor representative of the current small UAV platforms. The baseline mass (without additional payload) was 420 g.

Fig. 2 shows the experimental setup. The depth camera mounting configuration placed an Intel RealSense D455 on a rigid tripod in a fixed front-oblique view relative to the hover center. The camera was positioned at a height of approximately 1.2 m above the floor with a horizontal offset of approximately 1.8 m from the hover center, maintaining the UAV within the sensor's optimal measurement range while providing a sufficient field of view to accommodate minor position drift.

B. Payload Configuration and Testing Protocol

Calibrated payloads ranging from 0 to 39 g were applied to simulate various stress conditions that might indicate mechanical issues or operational limits. These weights were measured precisely and attached to the UAV using a consistent mounting protocol that ensured repeatability. The payload range includes light, medium, and heavy loads, spanning from imperceptible perturbation (13 g represents approximately 3% of UAV weight) to near-operational limits (32.5 and 39 g conditions) that approach the maximum additional payload for stable flight.

The payload conditions slightly differed between the two experimental days (hereinafter referred to as the **first test** and **second test**, respectively): 32.5 g was evaluated in the first test as a conservative upper bound, whereas 39 g was added

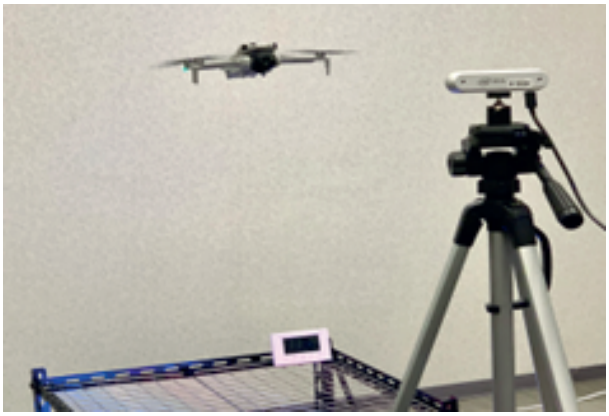


Fig. 2. Photograph of the experimental setup. A DJI Mini 4 Pro and an Intel RealSense D455 depth camera mounted on a tripod in a fixed front-oblique view during hovering.

in the second test after confirming a stable UAV performance in initial trials.

C. Data Collection and Processing

We refer to the two experimental campaigns as the **first test** and the **second test**. The pseudo-color mapping used in this study is shown in Fig. 3; gray/black denotes out-of-range/invalid pixels and warmer colors indicate shorter distances. The CAE was trained exclusively on concatenated 0 g hovering sessions from both campaigns, and payload frames were retained for evaluation.

For each experimental session, the framewise anomaly scores were obtained from the trained CAE model. A 50-frame moving average (MA-50) was applied to the per-frame scores to highlight low-frequency trends and suppress frame-level noise. Thus, the temporal dynamics are represented by the frame index on the x-axis and the MA-50 value on the y-axis.

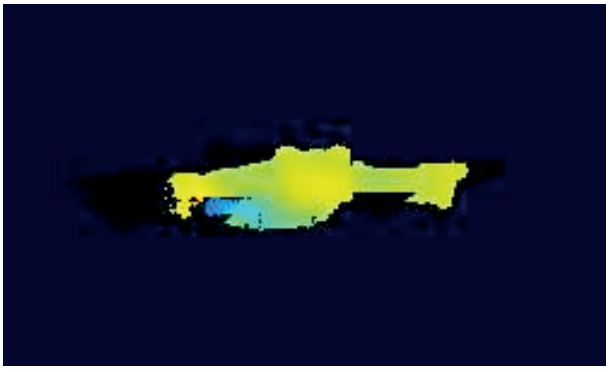


Fig. 3. Example of a depth heat map of a small UAV. Gray regions represent out-of-range values, black indicates missing data, blue tones correspond to larger distances, and red tones correspond to closer regions.

Table I summarizes the two test configurations (first and second tests). Each payload condition was repeated across multiple sessions, and session-wise trajectories were compared to verify reproducibility.

TABLE I
EXPERIMENTAL TEST CONDITIONS

Test	**Payloads (g)**
First test(FT)	0, 13, 26, 32.5
Second test(ST)	0, 13, 26, 39

TABLE II
MA-50 ANOMALY SCORE (MEAN AND STD.) BY SESSION

Condition	Test	Mean	Std.
0 g	First test (FT)	4.99×10^{-5}	8.93×10^{-6}
0 g	Second test (ST)	6.69×10^{-5}	1.07×10^{-5}
13 g	First test (FT)	2.58×10^{-4}	2.84×10^{-4}
13 g	Second test (ST)	1.39×10^{-4}	1.42×10^{-4}
26 g	First test (FT)	1.01×10^{-4}	5.25×10^{-5}
26 g	Second test (ST)	3.79×10^{-4}	5.00×10^{-4}
32.5 g	First test (FT)	7.06×10^{-4}	5.80×10^{-4}
39 g	Second test (ST)	2.45×10^{-2}	9.23×10^{-3}

V. RESULTS AND ANALYSIS

A. Baseline Performance

Under the baseline 0 g condition, the trained CAE model exhibited extremely low anomaly scores, on the order of 10^{-5} across sessions. This established a stable reference level for anomaly detection.

B. Error Dynamics under Payload Conditions

To improve readability, we have separated the time-series plots into three figures. Figure 4a shows the baseline 0 g behavior in the first and second tests (abbreviated as *FT* and *ST*). Figure 4b summarizes the light/medium payloads (13 g, 26 g), where session-dependent transients appear early and intermittently. Figure 4c presents the heavy payloads (32.5 g, 39 g), where the anomaly scores diverged from the baseline; the 39 g condition (ST) exhibited an order-of-magnitude or greater increase.

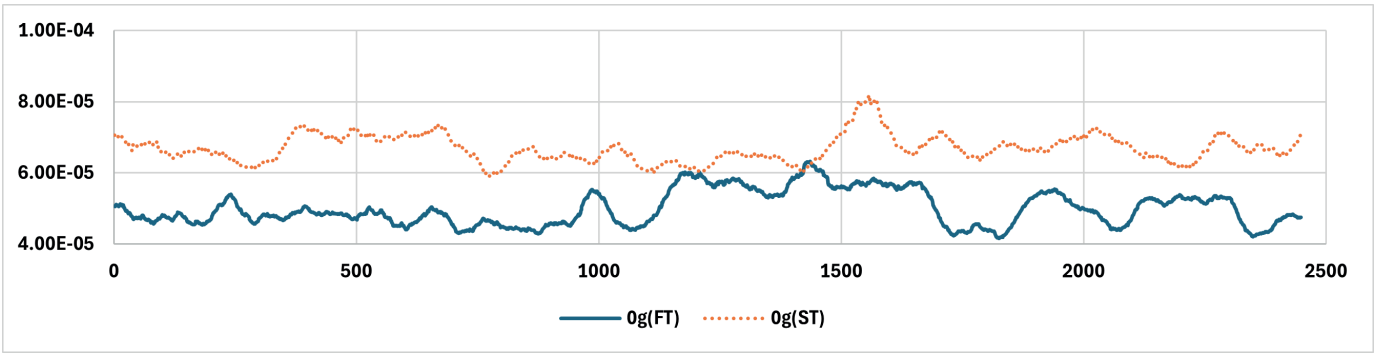
C. Statistical Summary

Table II summarizes the mean and standard deviations of the MA-50 anomaly scores for the representative payload conditions. The baseline remained low and stable, medium payloads exhibited session-dependent variation, and heavy payloads exhibited a clear increase.

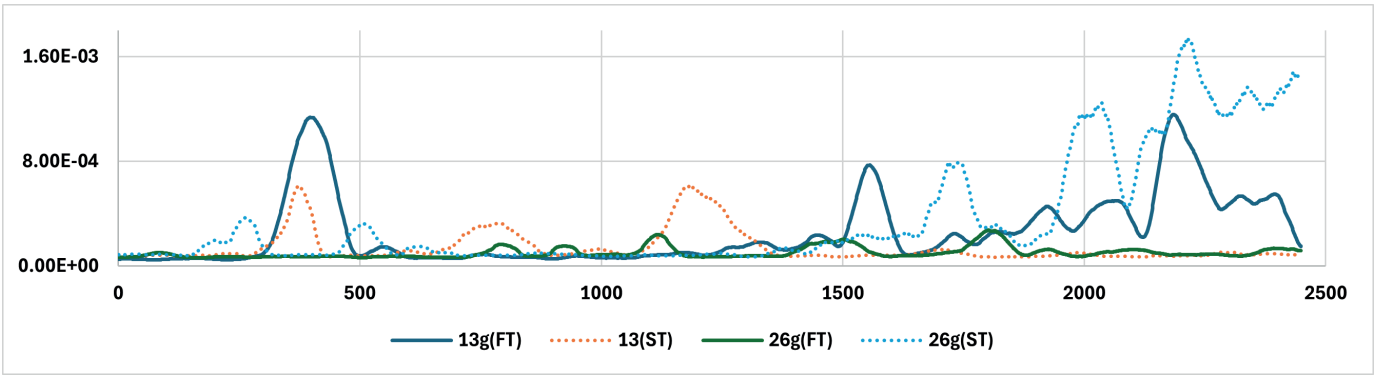
VI. DISCUSSION

A. Quantifying Payload-Induced Attitude Change

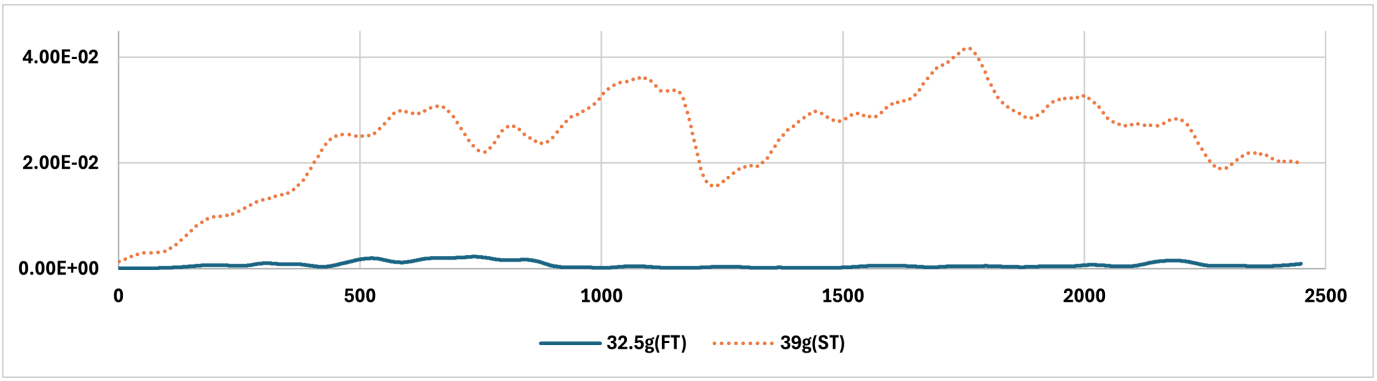
Fig. 4 shows that the application of payloads increases and separates the MA-50 anomaly score from the baseline. The 0 g condition remained low and stable, thus serving as a reliable reference. Light and medium loads exhibited early or delayed transient peaks with some session-to-session variability, whereas heavy loads (32.5/39 g) separated early from the baseline and remained elevated.



(a) Baseline 0 g. FT = solid, ST = dotted.



(b) Light/medium payloads (13 g, 26 g). FT = solid, ST = dotted.



(c) Heavy payloads (32.5 g, 39 g). FT = solid, ST = dotted.

Fig. 4. MA-50 anomaly-score time series (aligned at takeoff). Line style encodes the test campaign: FT = first test (solid), ST = second test (dotted). x-axis: frame index (samples). y-axis: MA-50 anomaly score (unitless).

B. Continuous Health Assessment Relative to a 0 g Baseline

Rather than relying on single peak values, we evaluated attitude changes using a continuous MA-50 time series, which was empirically more robust. The payload conditions tended to exhibit higher scores and earlier separation than 0 g, whereas light/medium loads may show session-dependent differences. For operational use, two thresholds can be derived from the 0 g baseline, $T_1 = \mu_0 + 2\sigma_0$ and $T_2 = \mu_0 + 3\sigma_0$, yielding a three-state policy on the same (MA-50) score: normal for $S < T_1$, attention for $T_1 \leq S < T_2$, and warning for $S \geq T_2$.

C. Interpreting Nonmonotonicity

Two factors plausibly explain the nonmonotonic responses. *Control-side:* near-hover compensation can introduce short-lived overshoot or oscillation. *Experimental-side:* the battery state-of-charge and limits of manual repeatability can shift the timing and amplitude of observed peaks.

D. Limitations and Next Steps

This study used a fixed external depth camera viewpoint and an indoor setting with a single UAV platform. Next, we will extend the evaluation to indoor yaw rotations (about the Z-axis) and short-axis-aligned translations along x, y, and z.

VII. Conclusion

We presented a comprehensive framework for attitude anomaly detection for small UAVs using depth imaging and CAEs. The system successfully identified both subtle and significant attitude deviations without requiring labeled anomaly data, demonstrating the viability of unsupervised learning for practical health-monitoring applications. This depth-based approach provides a robust, lighting-invariant operation suitable for diverse deployment scenarios.

Our experimental results revealed complex relationships between mechanical stress and anomaly manifestations, highlighting the importance of understanding the control system dynamics in the interpretation of detection results. The nonmonotonic response to the loading conditions suggests that simple threshold-based approaches may be insufficient, motivating continued research on context-aware anomaly interpretations.

The modular architecture and unsupervised learning approach of the framework facilitate adaptation to different small UAV platforms and operational requirements. The system supports nuanced decision-making and early warnings of developing issues by providing continuous anomaly scores instead of binary classifications. These characteristics make this approach particularly suitable for integration into predictive maintenance workflows.

Future studies will focus on field validation, specifically for automated preflight inspection, extension to additional UAV platforms, and integration with complementary sensing modalities. The ultimate goal is to develop comprehensive and automated health monitoring systems that ensure the safety and reliability of increasingly autonomous UAV operations across diverse application domains.

References

[1] D. Scaramuzza and F. Fraundorfer, "Visual odometry [tutorial]," *IEEE Robotics & Automation Magazine*, vol. 18, no. 4, pp. 80–92, 2011.

[2] M. Achtelik, M. Achtelik, S. Weiss, and R. Siegwart, "Onboard imu and monocular vision based control for mavs in unknown in- and outdoor environments," in *2011 IEEE International Conference on Robotics and Automation*, 2011, pp. 3056–3063.

[3] C. Forster, M. Pizzoli, and D. Scaramuzza, "Svo: Fast semi-direct monocular visual odometry," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 15–22.

[4] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep learning for anomaly detection: A review," *ACM computing surveys (CSUR)*, vol. 54, no. 2, pp. 1–38, 2021.

[5] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," *arXiv preprint arXiv:1901.03407*, 2019.

[6] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 4183–4192.

[7] K. Hara, H. Kataoka, and Y. Satoh, "Learning spatio-temporal features with 3d residual networks for action recognition," in *Proceedings of the IEEE international conference on computer vision workshops*, 2017, pp. 3154–3160.

[8] B. Kwolek and M. Kepski, "Human fall detection on embedded platform using depth maps and wireless accelerometer," *Computer Methods and Programs in Biomedicine*, vol. 117, no. 3, pp. 489–501, 2014.

[9] H. Oleynikova, Z. Taylor, M. Fehr, R. Siegwart, and J. Nieto, "Voxblox: Incremental 3d euclidean signed distance fields for on-board mav planning," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 1366–1373.

[10] T. Hirata, K. Yoshida, K. Koido, and S. Takahashi, "Anomaly detection in air conditioners using iot technologies," in *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*, 2021, pp. 1553–1559.

[11] T. Hirata, Y. Kuriyama, T. Hinohara, S. Takahashi, K. Yoshida, T. Ogasawara, and M. Go, "Model creation method for anomaly detection in refrigeration and air conditioning systems," in *2024 International Conference on Information Networking (ICOIN)*, 2024, pp. 29–34.