**2026 NDIA MICHIGAN CHAPTER**
**GROUND VEHICLE SYSTEMS ENGINEERING**
**AND TECHNOLOGY SYMPOSIUM**
**APPLIED ARTIFICIAL INTELLIGENCE (AAI) TECHNICAL SESSION**
**AUGUST 11-13, 2026 - NOVI, MICHIGAN**

# BEYOND DETECTION: ZERO-SHOT UAS PAYLOAD IDENTIFICATION, PASSIVE RANGING, AND POSE ESTIMATION

**Gregory Matousek[1], Nathan Varberg[1], Pete Torrione[1], Namdi Brandon[1], Matt Inkawhich[1], Joe Camilo[1]**

[1]CoVar, LLC, 1765 Greensboro Station Pl, Ste 900, McLean, VA 22102

## ABSTRACT

*Unmanned Aerial Systems (UAS) pose a growing threat on the modern battlefield, demanding rapid detection and characterization capabilities for the warfighter. Existing single-model solutions are inadequate for Counter-UAS (C-UAS), as they struggle across varying ranges and cannot provide detailed contextual information beyond bounding boxes. We present ZEUS (Zero-shot Explainable Universal Segmentation), a multi-model detection and recognition system that integrates several machine learning approaches. ZEUS employs a high-performance UAS detector trained on synthetic, internally collected, and open-source datasets, with real-time capability demonstrated on edge hardware across both electro-optical and infrared modalities. For classification, ZEUS uses a zero-shot approach: detected UAS are segmented and compared against a library of 3D reference models rendered at various poses, enabling identification of new UAS types without retraining. This methodology additionally provides UAS pose and range estimates critical for threat assessment and engagement decisions.*

**Citation:** G. Matousek, N. Varberg, P. Torrione, N. Brandon, M. Inkawhich, J. Camilo "Beyond Detection: Zero-Shot UAS Payload Identification, Passive Ranging, and Pose Estimation," In *Proceedings of the Ground Vehicle Systems Engineering and Technology Symposium* (GVSETS), NDIA, Novi, MI, Aug. 11-13, 2026.

## 1. INTRODUCTION

Unmanned Aerial Systems (UAS) have become a defining feature of modern warfare; there is an urgent need for technologies that enable rapid detection and characterization of these threats. Traditional single-model approaches often fall short in Counter-UAS (C-UAS) scenarios which demand effectiveness across varying ranges

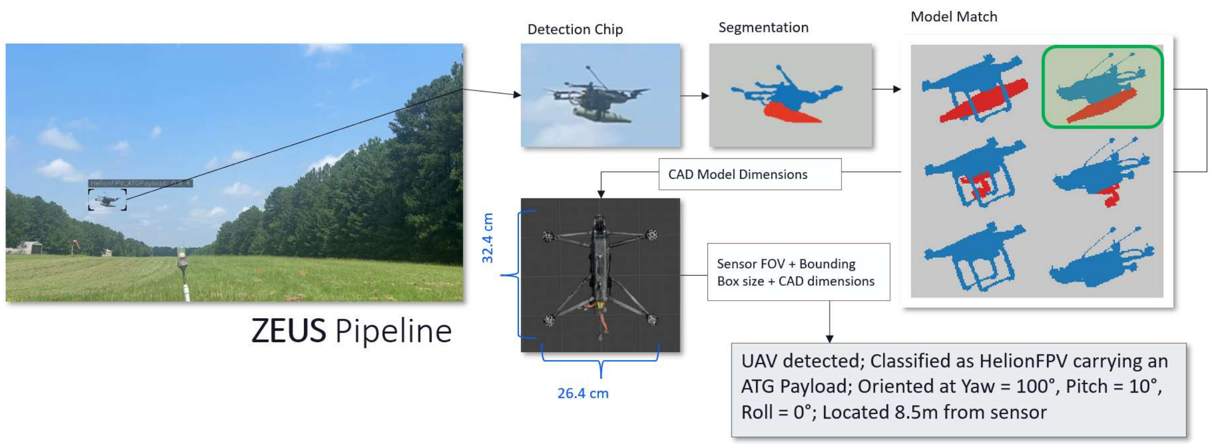


**Figure 1:** Overview of the ZEUS pipeline. First, a detection model predicts bounding boxes from the frame. Each detection is then chipped, and a segmentation model predicts a mask of what pixels correspond to UAS vs. background. We then compare each segmentation to a collection of 2D renders generated from a model library containing 3D CAD models from all UAS of interest. This model match chip classification also provides an explicit pose estimation which we can use to estimate range.

and additional contextual information for enhanced situational awareness and threat assessment.

In this report we present an overview of the ZEUS pipeline (summarized in Figure 1). The first step in the pipeline begins with a highly generalizable object detection module to predict bounding boxes. Details on this module are in Section 2.1. We then chip out each detected region from a given frame and use a lightweight segmentation model to predict which pixels in the chip correspond to the UAS. Details on the segmentation model are in Section 2.2. Finally, in Section 2.3 we detail our model match classification module which predicts class, pose, and range for each chipped detection by comparing its predicted segmentation mask against a library of 2D renders of known UAS CAD models. We recently successfully demonstrated our pipeline in real-time on a Jetson Orin edge device, which is presented in Section 2.4. Section 2.5 outlines our active research into recognizing a UAS' attached payload.

## 2. ZEUS Pipeline

### *2.1. UAS Detection*

UAS detection systems have become a practical necessity on the battlefield as the proliferation of small drones and other airborne threats outpaces what human operators can reasonably track on their own. Ultimately, the usefulness of any C-UAS pipeline is fundamentally limited by what happens at the front end. If a UAS isn't picked up by the detector, nothing downstream can compensate for that miss. Our philosophy is that the detector should be highly generalizable to various operating conditions and drone classes, favoring high recall over optimal precision as additional filtering of false alarm detections can easily be performed downstream in the classification module. Modern battlefields are highly dynamic, with new UAS emerging regularly. Thus, having a detection model that can reliably detect any airborne object will remain effective in the field far longer than a detector trained on a discrete set, one which will require constant re-training.

To achieve these goals, we trained a custom convolutional neural network UAS detector. To ensure the model generalizes well across UAS types and environments, we train it on data drawn from a wide range of sources. For EO detection, the dataset consists of over 500,000 annotated UAS. EO training data comprises open-source drone detection datasets, CoVar internal data collections, and synthetic drone footage. Segmented drone chips were inserted at random locations and scales into frames drawn from open-source datasets of nature trails, parks, and urban environments, producing a diverse set of visually cluttered, challenging training samples. This approach addresses a common gap in open-source data, where UAS data is typically captured against open sky and at relatively close range to the sensor, conditions that generally do not reflect the difficulty of real-world detection scenarios.

To further diversify training, we apply several image-based augmentations such as adjusting brightness, contrast, temperature, blur and compression. In IR we train a similar detector using sources such as internal data from CoVar collections and a significant amount of synthetic data generated in-house at CoVar. We use Falson, an Unreal Engine-based synthetic data simulation software developed by Duality AI [1], which enables us to ingest any arbitrary CAD file, assign its thermal properties, and place it into an expansive realistic world to capture synthetic LWIR imagery from a variety of viewing angles and ranges.

We evaluated our EO detector against an off-the-shell Vision Language Model (VLM) YOLO-World [2]. YOLO-World is an open-vocabulary detection model that leverages text prompts to identify arbitrary object categories in real time, without requiring task-specific training data, making it a strong baseline for comparison against a domain-specialized detector. As a real-time VLM, YOLO-World is also a compelling option for evaluating potential deployment scenarios. In this test, we evaluate two data collections which were held-out during our EO detector training pipeline: a long-range (30m+) data collect taken with a BASLER EO sensor, and a data collection taken from the Warsaw University of Technology (WUT) [3] which contains approximately 54,000 uncorrelated drone imagery taken at a variety of angles and environments. We use probability of detection (PD) vs. false alarm rate (FAR) as a metric for measuring model effectiveness. In both data sets we see significant improvements in our dedicated UAS detector compared to YOLO-World. In operationally relevant FAR regions our UAS detector consistently achieves substantially higher PD compared to YOLO-World. Empirically we see significant improvement in detection performance in situations where a UAS is among cluttered backgrounds (e.g. buildings, tree lines). This is especially important for a warfighter because UAS often operate in visually complex environments, and reliable detection in clutter ensures earlier awareness and more informed decision-making.

### *2.2. Segmentation Methodology*

Once a detection is made, the next phase of the ZEUS pipeline is to chip the detected region and segment the UAS pixels from the background. Like the detector, we want this segmentation model to be as generalizable as possible and not overfit to a particular drone set. While there have been significant advances in foundational off-the-shelf segmentation models like SAM [4], we find that they are inadequate for our task. First, SAM models do not perform well on small blurry objects, especially in the IR domain (see Figure 2 for an example of a SAM 2 [5] drone segmentation in the “easier” RGB domain). Second, SAM models are quite large and computationally expensive, which makes them a suboptimal choice for real-time deployment on compute-restricted edge hardware. Instead, we opt for training our own purpose-built segmentation model.

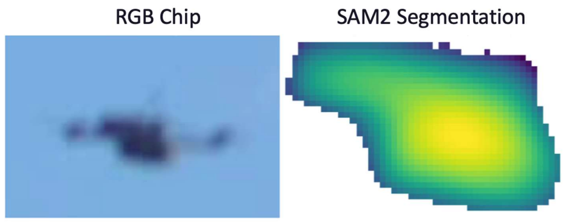


**Figure 2:** Off-the-shelf SAM2 struggles to segment few pixel blurry objects like distant UAS.

Due to the scarcity of labeled IR UAS segmentations, only true synthetic segmentation masks were used during training. In Figure 3, snapshots of segmentation performance during model training are shown. Preliminary results quickly showed synthetic-only trained IR segmentation could generalize well to real IR imagery. A similar trend was not seen in EO. For instance, image compression artifacts that are absent in synthetic data often mislead synthetic-only trained EO segmentation models to hallucinate drone/payload pixels. As a result, a small, yet heavily weighted sample of hand-labeled real EO segmentation masks were folded into the training pipeline, yielding more accurate results downstream.

Current results indicate our segmentation training pipeline produces results that generalize well to un-seen footage and environmental conditions. In Figure 4, sample frames from an internal IR and EO sensor data collect demonstrate our capability to distinguish drone and payload at moderate range. The IR imagery shown originated from a FLIR BOSON camera, which is available commercially. RGB imagery was obtained via a Logitech Webcam. In both cases, at closer range, enough detail is resolved with respect to the payload to identify it (e.g. a Go-Pro camera or an anti-tank grenade).

### *2.3. Model Match, Pose, Ranging*

Once a UAS is detected, chipped, and segmented, ZEUS uses a novel 3D model match algorithm to predict object class, pose, and range in a zero-shot fashion. To

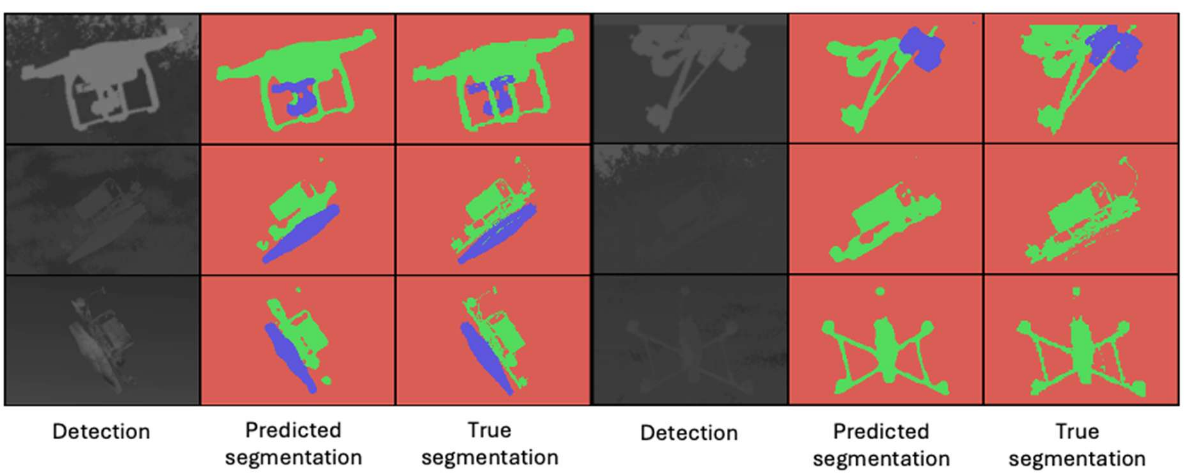


**Figure 3:** Held-out synthetic test examples of drones with payloads. Ground truth bounding boxes are used to give example detections to train the segmentation model. Predicted semantic segmentation masks are compared to true segmentation masks. Our IR segmentation model is trained using exclusively synthetic data.

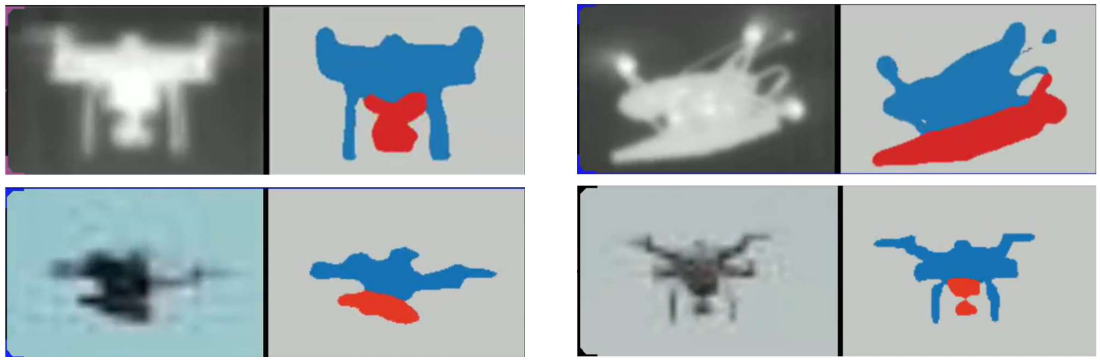

**Figure 4:** Sample chip segmentations in IR (top) and EO (bottom). Left images depict chips detected by our UAS detector – to their right are masks predicted by our trained segmentation model. Gray pixels correspond to predicted background, blue to the drone body, and red to the payload. Our segmenters can distinguish drone body and drone payload, offering more fine-grained classification capabilities downstream.

accomplish this, we maintain a library of CAD models for all UAS of interest.

Prior to deployment, we created a library of 2D renders from an extensive list of viewing angles, storing these masks with their corresponding yaw, pitch, and roll. During deployment, we compute the similarity of the predicted segmentation mask of a detected chip to the library of known possible renders and select the render with the closest match as the predicted class and pose. See Figure 1 for a visualization of this process.

One of the main benefits of the 3D CAD model match classifier is the fact that it is *zero-shot*, meaning that when a novel UAS class appears we do not need any explicit in-domain training data of it. Instead, we simply estimate a CAD model of the UAS and add it to the model library. Because the detection and segmentation models are trained to be generalists, most novel airborne vehicles will be detected and segmented appropriately. The zero-shot nature of ZEUS makes it a robust solution for C-UAS that is more appropriate for modern dynamic environments than traditional models (e.g., YOLO [6]).

A key feature of ZEUS' CAD model-based matching pipeline is its ability to perform passive ranging. Because we know the UAS' physical size (from the CAD model), pose,

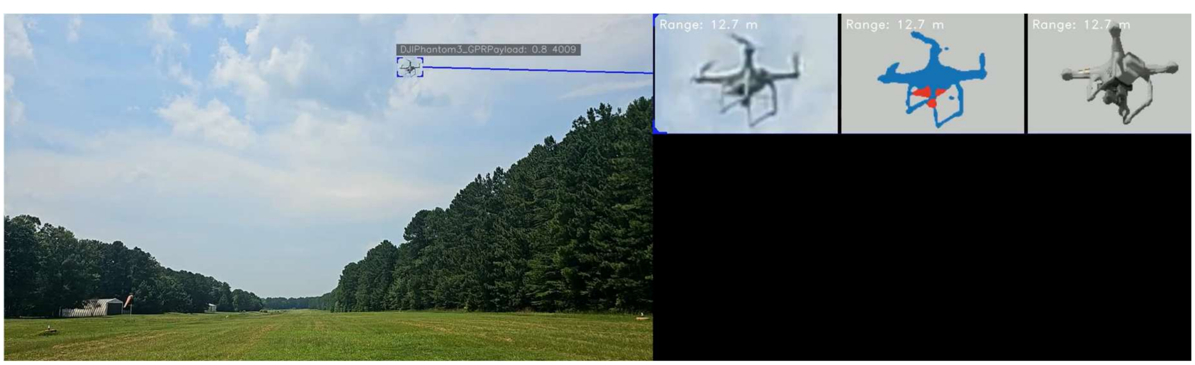


**Figure 5:** ZEUS pipeline evaluated on BASLER EO sensor data at an internal data collect. The left image corresponds to the full detection input frame. The row of chips to the right correspond to outputs of the pipeline. From left-to-right, the UAS image chip is recovered by the detector, receives a drone-payload pixel mask by our segmenter, which is then assigned to a model library match. For visualization purposes, a 3D render of the matched CAD model is shown at its predicted pose. At the top left of each image chip, the result of the passive ranging module is shown in meters.


**Figure 6:** ZEUS pipeline evaluated on FLIR BOSON IR sensor data at an internal demo. In this footage, a DJI Phantom 3 carrying a Go Pro payload is correctly classified by our pipeline.

and pixel extent, as well as the sensor's FOV, we can use simple trigonometry to estimate the UAS' range. Estimating range from raw imagery is a powerful feature of ZEUS, as it provides invaluable metadata without compromising concealment on the battlefield via active laser-emitting rangefinders. Furthermore, this capability comes with negligible storage and computational requirements in addition to the existing pipeline.

Figure 5 provides a detailed example of how the ZEUS pipeline processes EO sensor inputs, showing each stage from detection to model matching. A similar pipeline has been configured for IR input. The full pipeline incorporates several additional modules: object tracking to maintain consistent identities across frames, class smoothing to prevent abrupt classification changes caused by momentarily poor segmentation, and quaternion-based pose smoothing to stabilize orientation estimates over time.

### 2.4. Edge Deployment

In January 2026, the CoVar team performed a live demo of both our IR and RGB zero-shot drone detection pipelines on separate Jetson Orins. For the RGB sector we supplied video frames from a Logitech webcam, and in IR we used a commercially available FLIR BOSON camera. A representative frame from the IR data collect is shown in Figure 6. Throughout the demonstration, real-time ZEUS pipeline visualizations were displayed across dual monitors, one for each modality. Across the full detection, segmentation, model matching, and visualization pipeline, both channels sustained an average of 12 FPS. This demonstration marks a significant milestone: a fully operational, zero-shot drone detection capability running in real time on edge hardware, across two modalities, with no domain-specific training required.

### 2.5. Payload Identification

In contemporary operational settings UAS are frequently equipped with ad-hoc payloads including improvised explosives, ISR pods, and communication relays. Operators need an explainable, adaptable, and sensor efficient description of both the platform and its payload to prioritize engagement and allocate limited defensive resources. In this subsection, we outline our current work into using ZEUS to disambiguate drone from payload using a similar CAD model match approach.

To demonstrate this payload-aware ZEUS capability, we assembled a test set of internally collected EO imagery containing drones equipped with various payloads. Our model library included multiple representative drone bodies and several payload configurations. Matching was performed using the zero-shot segmentation-to-render comparison process described above. Overall performance showed consistent separation of both drone geometry and associated payload type, with remaining ambiguities occurring primarily among visually similar configurations (e.g. ±180 degree pose ambiguities for symmetric drones). From an operational standpoint, the most critical outcome is determining whether a given platform is carrying any payload at all. Under this criterion, the ZEUS pipeline demonstrated reliable discrimination of payload and no-payload cases. Model match results shown in Figures 5 and 6 showcase this capability. Our results indicate that, given sufficient pixel-level detail, ZEUS can sustain tactically relevant payload identification across a variety of ranges, background conditions, and drone types, providing richer situational context for nuanced C-UAS decision making.

## 3. CONCLUSIONS

This work introduces a Zero-Shot Explainable Universal Segmentation-Based (ZEUS) object detector with important C-UAS operational capabilities packed into a real-time, edge deployable pipeline. We developed a state-of-the-art single frame UAS detector robust to a variety of environmental conditions, with training data pooled from multiple open-source datasets, internal CoVar collections, and synthetic datasets. Our UAS detections are passed through a lightweight segmentation model and a 3D model-matching system to provide zero-shot class, pose, and range estimates from EO and IR chips. We also demonstrate effective payload disambiguation, a critical tool for threat assessment. Real-time tests on a Jetson Orin confirmed that the full detection, segmentation, and model-match pipeline can operate reliably at real-time frame rates.

## 4. REFERENCES


[1] A. Shah, "The Digital Twin Encapsulation Standard: An Open Standard Proposal for Simulation-Ready Digital Twins," in *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)*, Orlando, FL, 2023.

[2] T. Cheng, L. Song, Y. Ge, W. Liu, X. Wang and Y. Shan, "YOLO-World: Real-Time Open-Vocabulary Object Detection," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* 2024.

[3] M. Ł. Pawełczyk and M. Wojtyra, "Real World Object Detection Dataset for Quadcopter Unmanned Aerial Vehicle Detection," *IEEE,* vol. 8, pp. 174394-174409, 2020.

[4] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár and R. Girshick, "Segment Anything," *ICCV,* 2023.

[5] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, E. Mintun, J. Pan, K. V. Alwala, N. Carion, C.-Y. Wu, R. Girshick and P. Dollár, "SAM 2: Segment Anything in Images and Videos," *ICLR,* 2025.

[6] Y. Tian, Q. Ye and D. Doermann, "YOLOv12: Attention-Centric Real-Time Object Detectors," *CoRR,* 2025.

## 5. CONTACT INFORMATION

Joe Camilo
1765 Greensboro Station Pl, Ste 900
McLean, VA
joe@covar.com
984-297-0833

## 6. ACKNOWLEDGMENT

This work was funded under an SBIR D2P2 from DEVCOM Army Research Lab. This work made use of the Falcon simulation platform developed by Duality AI to generate high-fidelity synthetic data.