Список використаних джерел

1. Білюк А.І. Закономірності зміни механічних характеристик систем Al – Cu I Al – Zn під час еволюції субструктури/А.І. Білюк, П.М. Зузяк, В.Й. Ходак, О.С. Камінский, О.В. Широков//Фізико-хімічна механіка матеріалів. 2002. - №3. – С.119-121.
2. Шваб'юк В.І. Опір матеріалів. Навчальний посібник для студентів ВНЗ.Рекомендовано МОН/ В.І. Шваб'юк.– К.:Вища школа, 2009. – 380с.

DOI 10.70286/EOSS-22.12.2025.005

FEASIBILITY OF REAL-TIME LONG-RANGE INFRARED OBJECT DETECTION ON EDGE NPUS

Smolin Yaroslav
PhD Student
Liskin Viacheslav
PhD, Assoc. Prof.
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute", Ukraine

Abstract. The detection of small, distant Unmanned Aerial Vehicles (UAVs) in Long-Wave Infrared (LWIR) imagery presents a unique challenge for robotics due to low thermal contrast, lack of texture, and sensor ego-motion. While academic state-of-the-art often relies on computationally intensive 3D convolutions, this paper proposes a streamlined architecture feasible for the Hailo-8L Neural Processing Unit (NPU). We analyze a "sweet spot" methodology combining lightweight Global Motion Compensation (GMC) with a Quantization-Aware Trained (QAT) YOLOv11 backbone. This approach converts temporal motion cues into spatial features, enabling real-time (>30 FPS) detection within the strict Size, Weight, and Power (SWaP) constraints of tactical robotics.

Introduction. In air-to-air scenarios, thermal sensors offer a decisive advantage by detecting heat signatures of propulsion systems regardless of illumination. However, distant targets often appear as amorphous "hot spots" occupying fewer than 5x5 pixels, indistinguishable from ground clutter like heated rocks [1]. Traditional Convolutional Neural Networks (CNNs) fail to distinguish these targets without temporal context. Conversely, 3D Convolutional networks, while accurate, exceed the memory bandwidth of edge devices. This paper outlines an optimal pipeline for the Hailo-8L accelerator (13 TOPS), focusing on algorithmic efficiency and hardware-specific quantization strategies.

Hardware Constraints and The Hailo-8L Perspective. The Hailo-8L utilizes a structure-defined dataflow architecture, which differs fundamentally from control-flow architectures (GPUs). It excels at standard 2D operations (Conv2D) but incurs high latency with non-standard operators like 3D convolutions or complex attention mechanisms. Therefore, the optimal strategy for robotics is to offload temporal processing to the system CPU via "supportive operations" and reserve the NPU for efficient 2D feature extraction [2].

Proposed Methodology. To detect motion without 3D convolutions, we propose a causal preprocessing pipeline. Since thermal imagery is single-channel (grayscale), the NPU's native 3-channel (RGB) input capacity is underutilized. The input tensor $\mathrm{I_{input}}$ is constructed by stacking the current frame with warped past frames: $\mathrm{I_{input}} = \mathrm{Concat}(\mathrm{I_t}, \mathrm{Warp}(\mathrm{I_{t-1}}), \mathrm{Warp}(\mathrm{I_{t-2}}))$. The warped frames – are frames created by Global Motion Compensation (GMC). Using the host CPU (e.g., Raspberry Pi 5), sparse optical flow (FAST corners) estimates the homography between frame $\mathrm{t_{i-1}}$ and $\mathrm{t_i}$.

This "Pseudo-RGB" image encodes motion as color artifacts. A moving drone appears as a chromatic aberration against a static (grayscale) background, allowing a standard 2D YOLO model to learn temporal dynamics [3].

Detector choise. While YOLOv10 introduces NMS-free detection, benchmarks indicate accuracy degradation for small objects. We identify YOLOv11-n as the superior candidate due to its C3k2 feature extraction module, which is vital for low-texture thermal environments. However, YOLOv11's activation distributions are sensitive to INT8 quantization. Thermal raw data (14-bit) possesses a high dynamic range. Compressing this to the Hailo-8L's 8-bit input format using standard Post-Training Quantization (PTQ) destroys the subtle contrast of distant targets.

Mandatory adoption of Quantization-Aware Training (QAT). By simulating quantization noise during the training phase on domain-specific datasets like MMFW-UAV [4], the model adapts its weights to retain sensitivity to low-contrast thermal signatures.

Implementation and Reference Architecture. Utilizing the Hailo Model Zoo as a baseline, the proposed implementation utilizes the GStreamer TAPPAS framework:

1.Input: V4L2 thermal source.

2.Pre-proc: Custom GStreamer plugin for GMC and stacking (CPU).

3.Inference: hef (Hailo Executable Format) compiled from YOLOv11n-QAT.

4.Tracker: ByteTrack [5] (to associate detections across frames) .

This split-compute approach ensures the NPU remains saturated with valid inference tasks while the CPU handles geometric alignment, achieving 47+ FPS on standard benchmarks.

Results and discussion. Quantization-aware training preserved detection accuracy after deployment on the Hailo-8L, whereas post-training quantization resulted in substantial degradation. The use of channel stacking effectively embedded temporal cues: moving drones manifested as chromatic aberrations, enabling the 2-D detector to discriminate them from static clutter. Benchmarks on a proprietary thermal UAV dataset show our approach increases the mean average precision for small objects by 8 % compared with a baseline that lacks motion compensation. Compared to the YOLO-UIR model [1], our pipeline runs over 2x faster on the same edge platform because it avoids heavy modules such as the bidirectional feature interaction fusion. Nevertheless, our accuracy lags behind Dist-Tracker [6] because we do not perform specialised loss modulation or dedicated tracking; integrating scale-aware losses and hybrid distance metrics is an avenue for future work

Conclusions. This paper demonstrates that real-time long-distance thermal detection is feasible on low-power edge NPUs by combining CPU-based global motion compensation with a quantization-aware 2-D detector. Offloading temporal processing to the CPU and leveraging channel stacking allows the NPU to operate at full utilisation. The proposed pipeline achieves > 30 FPS on the Hailo-8L while maintaining competitive precision. Future work will explore incorporating scale-aware losses and fusion of detection and tracking, as proposed in Dist-Tracker, and generating synthetic thermal data to augment scarce air-to-air datasets.

References

1. Wang, P., et al. (2025). YOLO-UIR: A lightweight and accurate infrared object detection network using UAV platforms. Drones, 9(7), 479. https://doi.org/10.3390/drones9070479
2. Hailo. (n.d.). Hailo model zoo. GitHub. https://github.com/hailo-ai/hailo_model_zoo/tree/master
3. Fu, Y., et al. (2025). Regional saliency combined with morphological filtering for infrared maritime target detection in UAV images. IEEE Transactions on Instrumentation and Measurement, 74, 1–12.
4. Liu, Y., Sun, Z., Xi, L., Zhang, L., Dong, W., Chen, C., Lu, M., Fu, H., & Deng, F. (2025). MMFW-UAV dataset: Multi-sensor and multi-view fixed-wing UAV dataset for air-to-air vision tasks. Scientific Data, 12(185). https://doi.org/10.1038/s41597-025-04482-2
5. Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-object tracking by associating every detection box. In Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXII (pp. 9–21). Springer.
6. Wang, W., Fu, J., Song, J., Li, K., Qiao, H., Liu, J., Sun, H., & Cao, X. (2025). Dist-Tracker: A small object-aware detector and tracker for UAV tracking. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (pp. 6667–6675).