



IJAS-26-147

ACG-YOLO for Real-Time Public Safety Surveillance: A Short Commentary

Hairui Wang¹, Xiaoqing Wang¹, Guifu Zhu² and Ya Li^{1*}

¹Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650504, China

²Information Technology Construction Management Center, Kunming University of Science and Technology, Kunming 650504, China

Corresponding Author: Ya Li, Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650504, China, E-mail: hrwang88@163.com

Received date: 07 August, 2026, **Accepted date:** 21 August, 2026, **Published date:** 28 August, 2026

Citation: Wang H, Wang X, Zhu G, Li Y (2026) ACG-YOLO for Real-Time Public Safety Surveillance: A Short Commentary. *Innov J Appl Sci* 3(4): 54.

Abstract

Quantum computing is often described as evaluating many answers simultaneously. That metaphor confuses a large mathematical state space with accessible classical output and encourages a second mistake: treating an asymptotic speedup as an engineered or economic result. This commentary revisits Wright (2025) against experimental and resource-estimation evidence through 2026. The central criticism survives, but two claims require revision. Fault tolerance is no longer merely hypothetical after below-threshold logical-memory experiments and cryptanalytic resource estimates have fallen substantially under specified hardware assumptions. The defensible position therefore rests on four distinctions: computability, algorithmic complexity, fault-tolerant execution and practical value. A credible claim of quantum advantage must identify the task, compare against the strongest current classical method, include the complete system cost and explain verification. This framework supports continued research and post-quantum migration while rejecting claims that outrun demonstrated performance.

Keyword: Quantum computing, Quantum advantage, Quantum error correction, Computational complexity, Postquantum cryptography, Technology assessment

Introduction

Real-time detection of violent behavior and hazardous weapons is an important component of intelligent public safety surveillance. In practical surveillance environments, however, detection models must operate under considerable resource constraints while handling small objects, occlusion, multiple people and complex backgrounds. These requirements make the balance between detection accuracy and computational efficiency particularly important for edge-deployed systems.

The recent work on ACG-YOLO presents a lightweight detection framework designed for this setting. Built upon YOLOv11n, the framework combines three complementary components: ADown for feature-preserving down sampling, Coordinate Attention (CoordAtt) for spatial localization and a lightweight Global Context Enhancement Module (GCEM) for global semantic modeling. Rather than relying on a single architectural modification, ACG-YOLO distributes different representation requirements across these components. This commentary discusses the motivation, technical characteristics, experimental findings and limitations of this design,

with particular attention to what can be learned from its collaborative architecture.

Technical Motivation and Design Perspective

Violence and weapon detection involves several representation challenges that are closely related but not identical. Small weapons such as knives and firearms may occupy only a limited number of pixels, making them vulnerable to information loss during re-peated down sampling. At the same time, violent interactions often involve multiple people, body-part occlusion and changing spatial relationships. A detector therefore needs to preserve local details while also maintaining sufficient contextual information for scene interpretation.

ACG-YOLO addresses these requirements through a three-level design. ADown modifies the down sampling process by combining two feature-processing branches, aiming to retain useful local information while reducing computational cost. CoordAtt introduces directional positional information and is intended to improve the localization of spatially distributed features. GCEM operates on globally aggregated information and performs lightweight channel

recalibration, providing a global semantic signal without introducing a full spatial self-attention mechanism.

An important characteristic of this design is that the three components are not in-tended to perform the same function. ADown primarily concerns information preservation during resolution reduction, CoordAtt emphasizes spatial localization and GCEM provides global contextual information. This functional separation provides a useful perspective for lightweight detector design: reducing computational cost does not necessarily require removing representational capabilities if different operations are assigned to the stages where they are most useful.

What the Experimental Results Indicate

The experimental results provide evidence for the effectiveness of this collaborative design. On the Violence and Weapon Detection dataset, ACG-YOLO achieves 89.2% mAP@0.5 compared with 86.3% for the YOLOv11n baseline. At the same time, the model contains 2.07 M parameters and requires 5.1 GFLOPs, compared with 2.58 M parameters and 6.3 GFLOPs for YOLOv11n. These results indicate that the proposed modifications improve the measured detection accuracy while reducing the nominal model complexity.

The class-wise results are also informative. For Gun detection, the mAP@0.5 remains essentially unchanged, increasing from 96.8% to 96.8%, whereas Knife mAP@0.5 increases from 66.6% to 77.1%. The larger improvement for knives is consistent with the greater difficulty of detecting slender and relatively small targets. This result suggests that the benefit of lightweight architectural modifications may depend strongly on target characteristics rather than being uniformly distributed across object categories.

Another notable observation comes from the comparison of attention mechanisms. In the standalone attention experiment, CBAM and ECA obtain higher mAP@0.5 values than GCEM. However, after integration with ADown and CoordAtt, the complete ACG configuration achieves 89.2% mAP@0.5, compared with 88.5% for the ACC configuration and 88.6% for ACE. This difference is important because it shows that the performance of an individual module cannot necessarily be used to predict its behavior within a complete detector.

The result highlights the role of architectural compatibility. An attention mechanism that performs well in isolation may not provide the same benefit when combined with other feature-processing operations. In ACG-YOLO, CoordAtt already provides directional spatial information, while GCEM focuses on global channel-level contextual information. Their different functional roles may help avoid excessive redundancy between the two mechanisms. From this perspective, the main lesson is not that GCEM is universally superior to other attention mechanisms, but that the interaction among lightweight components should be considered when designing an efficient detector.

Generalization and Practical Relevance

The cross-dataset experiment on COCO2017 provides additional evidence regarding the generality of the architecture. ACG-YOLO achieves 89.6% mAP@0.5 on COCO2017, compared with 86.3% for the corresponding baseline, giving a 3.3 percentage-point improvement. Although cross-dataset results alone cannot establish that the improvement is independent of all dataset-specific factors, the consistent performance difference suggests that the architectural modifications are not restricted to the original violence and weapon detection scenario.

The reduction-ratio sensitivity experiment also provides a practical perspective. When the reduction ratio is varied among 8, 16 and 32, the model maintains relatively stable performance, with the reported mAP@0.5:0.95 values varying within a limited range. Such stability is useful for lightweight model development because it indicates that the framework does not depend on an extremely narrow hyperparameter setting.

From a deployment perspective, the reported 2.07 M parameters and 5.1 GFLOPs demonstrate a relatively compact architecture. However, theoretical complexity and desk-top or laboratory FPS measurements should not be regarded as direct substitutes for actual edge-device evaluation. Factors such as memory access, hardware-specific optimization, power consumption and inference latency can affect practical deployment performance. Therefore, benchmarking on representative platforms such as embedded GPUs would provide a more complete assessment of the framework's deployment potential.

Limitations and Future Directions

Several limitations remain to be addressed. First, the current framework focuses on frame-level object detection, whereas violence is inherently temporal. Information from consecutive frames could help distinguish transient interactions from sustained violent behavior and may reduce ambiguity caused by individual frames.

Second, the current efficiency evaluation mainly relies on parameter count, GFLOPs and FPS. These metrics are useful for comparing model complexity but do not fully characterize deployment efficiency. Future work should include hardware-specific latency, memory consumption and power measurements on representative edge platforms.

Third, although the collaborative architecture provides a lightweight solution, further simplification may be possible through quantization, pruning, knowledge distillation, or inference-engine optimization. Such techniques could be investigated together with Tensor RT or other deployment-oriented acceleration frameworks.

Finally, public safety surveillance involves challenging conditions such as illumination changes, severe occlusion, camera motion and crowded scenes. More diverse datasets and long-duration

video benchmarks would help determine whether the observed advantages remain stable under these conditions.

Conclusion

ACG-YOLO provides an example of how lightweight detection can be approached through functional collaboration rather than relying solely on global network compression. ADown, CoordAtt and GCEM address feature preservation, spatial localization and global contextual modeling, respectively. The experimental results show improved overall detection performance together with reduced parameter count and computational cost, while the class-wise results indicate a particularly noticeable improvement for knife detection.

More broadly, the value of the framework lies in the observation that lightweight modules should be evaluated not only individually

but also according to their compatibility within the complete architecture. The comparison between standalone attention performance and full-framework performance illustrates this point clearly. At the same time, further validation on actual edge devices and temporal surveillance data will be important for assessing the practical applicability of this approach. These directions provide opportunities to extend lightweight detection from model-level efficiency toward more comprehensive real-world surveillance systems.

Conflict of interest

The author declares no conflict of interest.

Copyright: © 2026 The Author(s). Published by Innovative Journal of Applied Science. This is an open-access article under the terms of the Creative Commons Attribution License (CC BY). (<https://creativecommons.org/licenses/by/4.0/>).