Optimizing Object Detection in Electro-Optical Systems with Snapshot Compressive Imaging

Yaping Zhao and Edmund Y. Lam

Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong SAR, China

ABSTRACT

This study introduces object detection within electro-optical (EO) systems utilizing snapshot compressive imaging (SCI). Traditional EO systems often suffer from extensive processing time and high computational demands due to the sequential processes of capture, compression, reconstruction, and detection. To address these challenges, our research leverages SCI integrated with artificial intelligence (AI) to streamline the detection process directly from compressed optical measurements, thereby omitting conventional intermediate steps. This methodology significantly reduces time, storage, and computational overhead, while simultaneously enhancing accuracy by exploiting motion-encoded information inherent in the compressed data. By satisfying stringent size, weight, and power (SWaP) requirements, our method holds promise for a variety of applications including autonomous driving, environmental monitoring, and public security. This paper represents not only a significant advancement in computational imaging, proposing a cost-effective, but also optimized solution suitable for a spectrum of EO systems.

Keywords: object detection, imaging systems, electro-optical systems, compressed sensing

1. INTRODUCTION

In the evolving landscape of electro-optical (EO) systems,[1] snapshot compressive imaging (SCI)[2] integrates compressive sensing with conventional imaging techniques to optimize optical signal capture. This integration aims to address the growing commercial demands for reducing costs and expediting development timelines. While SCI has advanced significantly, its potential for synergizing with artificial intelligence (AI) for tasks like object detection[3] in a variaty applications has been underexplored.[4]

Object detection in EO systems[5] is crucial for applications that require rapid responses under diverse and complex conditions, such as modern operations and civilian safety. Traditional imaging processes involve multiple stages: capture,[6] compression,[7] reconstruction,[8] and detection,[9] each adding to the time and computational overhead. The challenge lies in enhancing these systems to meet stringent size, weight, and power (SWaP) requirements while maintaining or improving detection performance in real-time scenarios.

Previously, approaches to improving object detection in EO systems primarily relied on separately optimizing individual stages of the imaging process. However, these methods often resulted in increased computational costs and extended processing times, which are not feasible for real-time applications.[10] Moreover, traditional methods typically do not utilize the motion-encoded information available in compressed measurements, limiting their effectiveness in dynamic environments.

Our research introduces a transformative AI-driven method for direct object detection from compressed optical measurements, tailored for modern EO systems. By leveraging SCI technology and eliminating the need for intermediate steps between capture and detection, our approach significantly reduces time, storage, and computational demands. Additionally, it enhances accuracy by incorporating motion information present in the compressed data. The efficacy of our method has been demonstrated by experiments on the dataset, indicating substantial improvements in real-time EO applications, including UAV and drone-operated reconnaissance and

Corresponding author: Edmund Y. Lam.

Yaping Zhao and Edmund Y. Lam are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong SAR, China. E-mail: {zhaoyp, elam}@eee.hku.hk

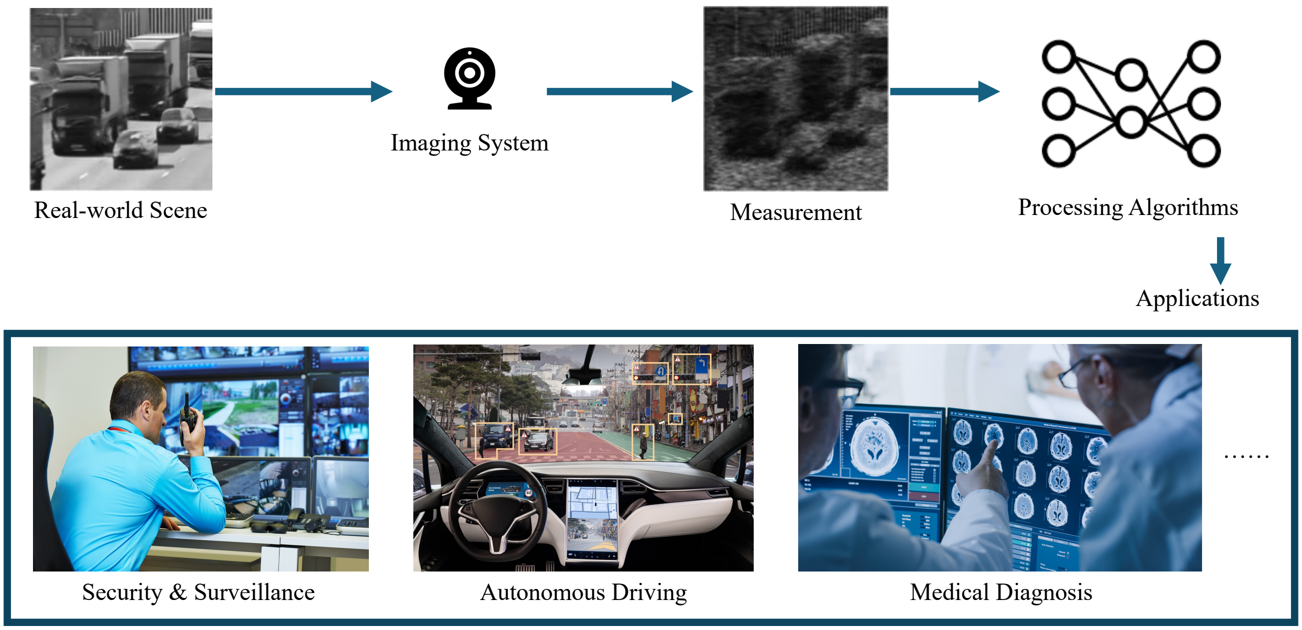

Figure 1. Object detection in EO systems with SCI boosts applications in a variety of scenarios including security (e.g., surveillance system and smart city), transportation (e.g., flying position control and automotive vehicles), healthcare (e.g., real-time patient monitoring and medical diagnosis) and etc

damage assessment. This pioneering approach not only meets the dual challenges of operational and commercial demands but also extends potential applications to fields like autonomous driving, environmental monitoring, and public security,[11] as Figure 1 shows.

2. METHODS

Our approach harnesses the potential of SCI to enhance object detection in electro-optical systems. This section delineates our methodological framework, emphasizing the optimization of mask patterns[12] and compression rates[13] to refine both the quality of imaging results and the efficacy of subsequent object detection tasks.

The core of our method involves adapting the mask patterns[12] and compression rate[13] dynamically, a departure from the traditional fixed or random settings utilized in prior compressive imaging studies. The adaptability is crucial for tailoring the imaging process to specific operational requirements, balancing between accuracy and efficiency of on-device processing.

For a practical understanding, we consider the scenario of video snapshot compressive imaging. As illustrated in Figure 2, high-speed frames from a video are captured at a rate exceeding the camera's typical capture speed. These frames undergo modulation using distinct mask patterns before being compressed into a single measurement output. Formally, suppose C_r denotes the number of video frames involved, with each frame $X_i \in \mathbb{R}^{w \times h}, \forall i = 1, \dots, C_r$, where w and h are the frame's width and height, respectively. Each frame X_i is modulated by a unique mask $M_i \in \mathbb{R}^{w \times h}, \forall i = 1, \dots, C_r$. The cumulative effect of these modulated frames is integrated within the camera's exposure time to produce a singular compressed measurement $Y \in \mathbb{R}^{w \times h}$, represented mathematically as:[14]

$$\mathbf{Y} = \sum_{i=1}^{C_r} \mathbf{M}_i \odot \mathbf{X}_i + \mathbf{E}, \tag{1}$$

where $\odot$ signifies the element-wise product and $\mathbf{E}$ represents the inherent measurement noise.

Figure 3 underscores the impact of varying mask patterns[12] and compression rates[13] on the resultant measurement quality and volume, which are critical for the subsequent processing and transmission phases. To optimize

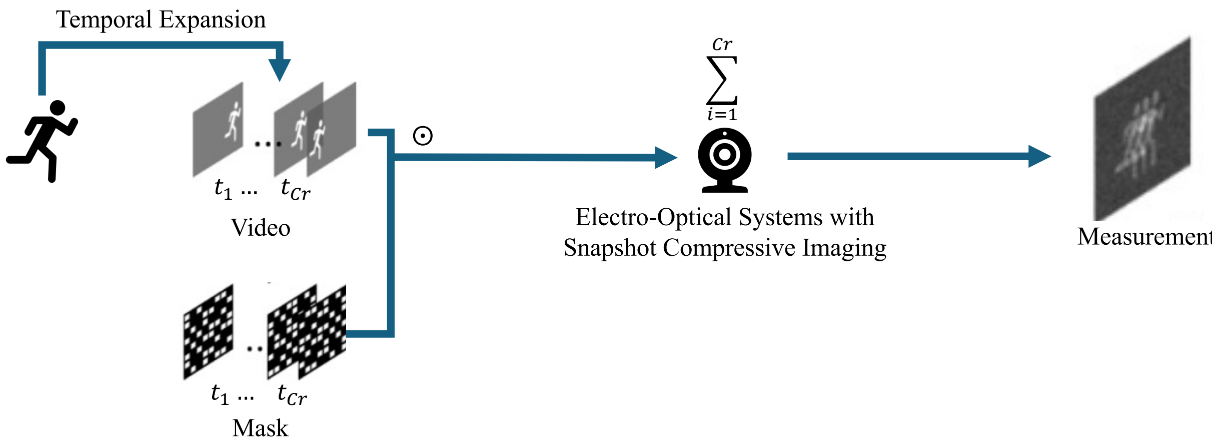

Figure 2. The pipeline of video compressive imaging, where $\odot$ denotes the element-wise product, C_r is the compression rate.

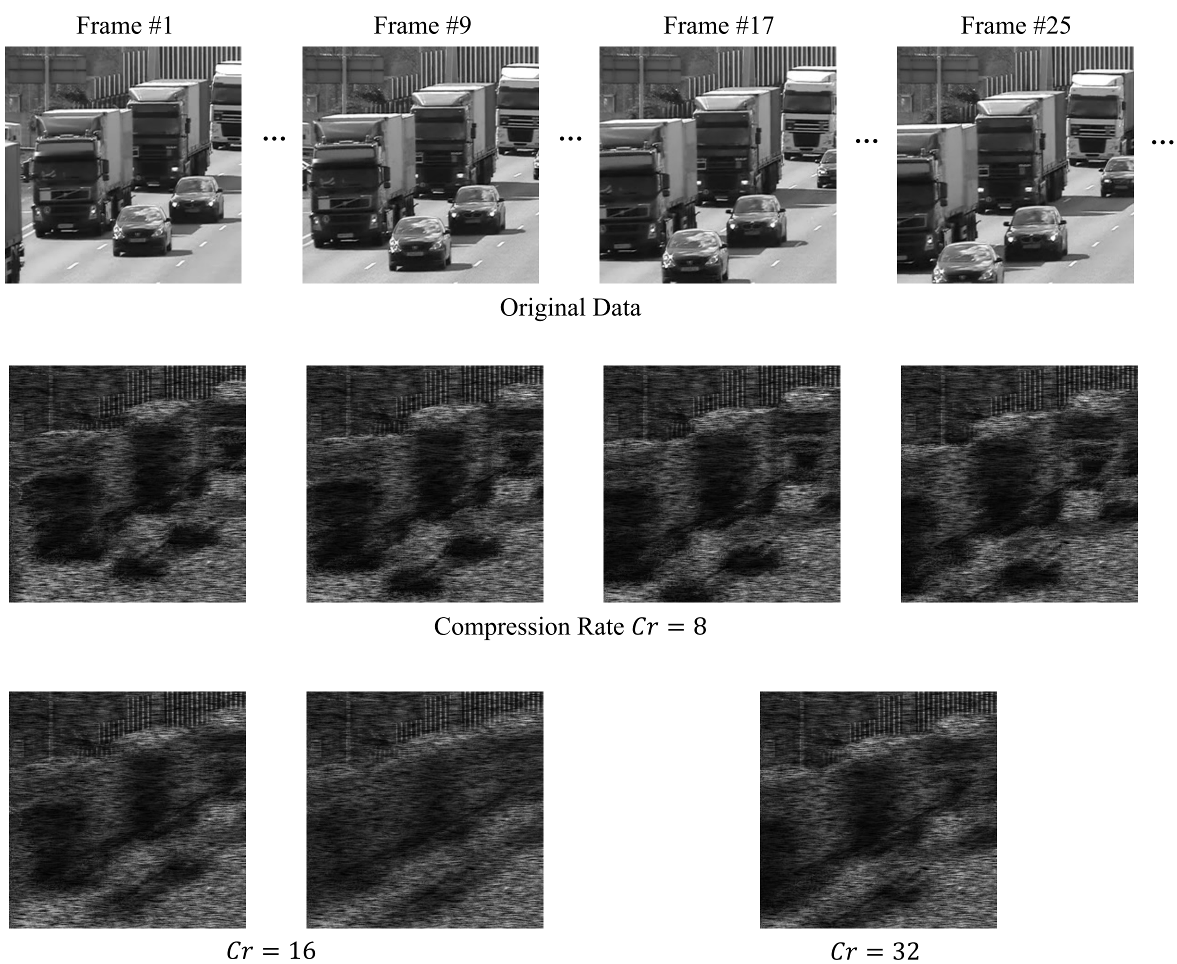

Figure 3. The different measurements (i.e., optical-domain compressed videos) under different compression rate settings, using the snapshot compressive imaging technique.

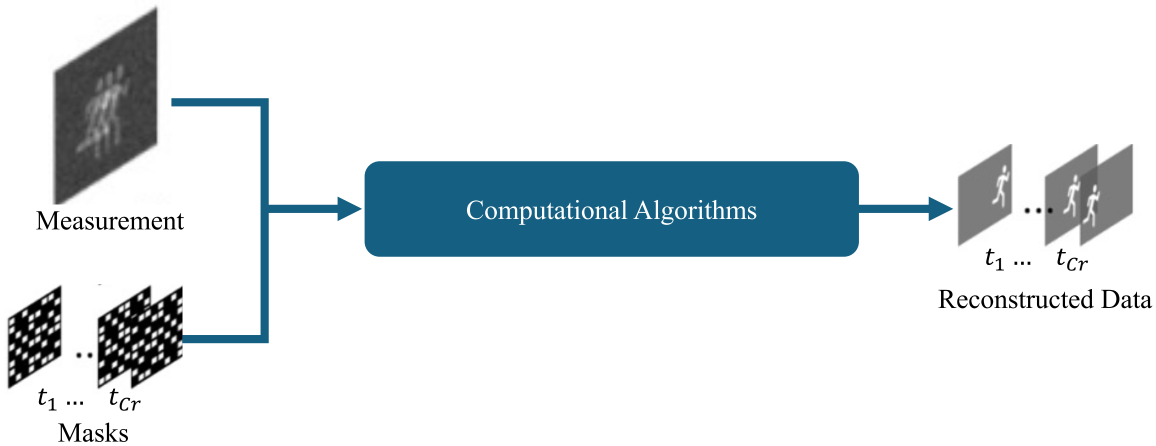

Figure 4. The decoding process of snapshot compressive imaging. The measurement captured by the snapshot compressive imaging technique along with the masks are input to recover the high-dimensional data using algorithms.[15–21]

these parameters, we have developed algorithms that adjust the mask patterns[12] and compression rates[13] in real-time based on feedback from the edge server and specific application scenarios. This adaptability allows for a seamless balance between data quality and operational efficiency.

In this study, rather than reconstructing the high-dimensional optical signal as depicted in Figure 4 traditionally required for object detection, our approach processes algorithms directly on the compressed measurements. This method not only conserves computational resources but also shortens the response time, crucial for real-time applications. The capability to perform direct object detection on compressed measurements is demonstrated in Section 3, highlighting the potential for efficient and accurate detection in resource-constrained environments.

3. RESULTS

Our experiments were conducted using the Vimeo-90K[22] dataset, a comprehensive high-quality video dataset designed for lower-level video processing tasks. Each video was compressed with a compression ratio $C_r = 8$ to generate optical compressed measurements. For object detection, we employed various configurations of the YOLO[23] model to benchmark our modified approach against traditional methods.

We implemented four different experimental setups: YOLO tested on the original uncompressed video, YOLO on the compressed and subsequently reconstructed video, YOLO directly on the compressed measurements as a baseline, and our optimized YOLO on the compressed measurements.

From the results depicted in Figure 5, the YOLO model on the original video detected a car with a confidence score of 0.96. However, on the reconstructed video, the same model misidentified the car as a truck. The baseline method, testing YOLO directly on compressed measurements, detected the car but with a reduced confidence of 0.92. Our optimized YOLO model on the compressed measurements not only maintained a high confidence score of 0.95 in detecting the car but also uniquely identified a potted plant, which was not detected by any other methods. This additional detection likely benefits from the rich temporal information embedded in the compressed optical measurements.

Figure 6 further illustrates our method's superiority. On the original video, the YOLO model detected a person with a confidence of 0.89 but erroneously marked a frisbee. The reconstructed video saw an incorrect detection of a skateboard, while the baseline detected the person with a lower confidence of 0.81. Our method, however, achieved a confidence of 0.92 in detecting the person without any erroneous detections of unrelated objects. This improvement underscores the advantage of leveraging compressed optical measurement for enhancing detection accuracy.

The experimental outcomes highlight several critical insights. While detection from original uncompressed videos generally showed high accuracy, it demands substantial storage and computational resources and occasionally provides inferior results compared to our method. Videos that were compressed and reconstructed typically

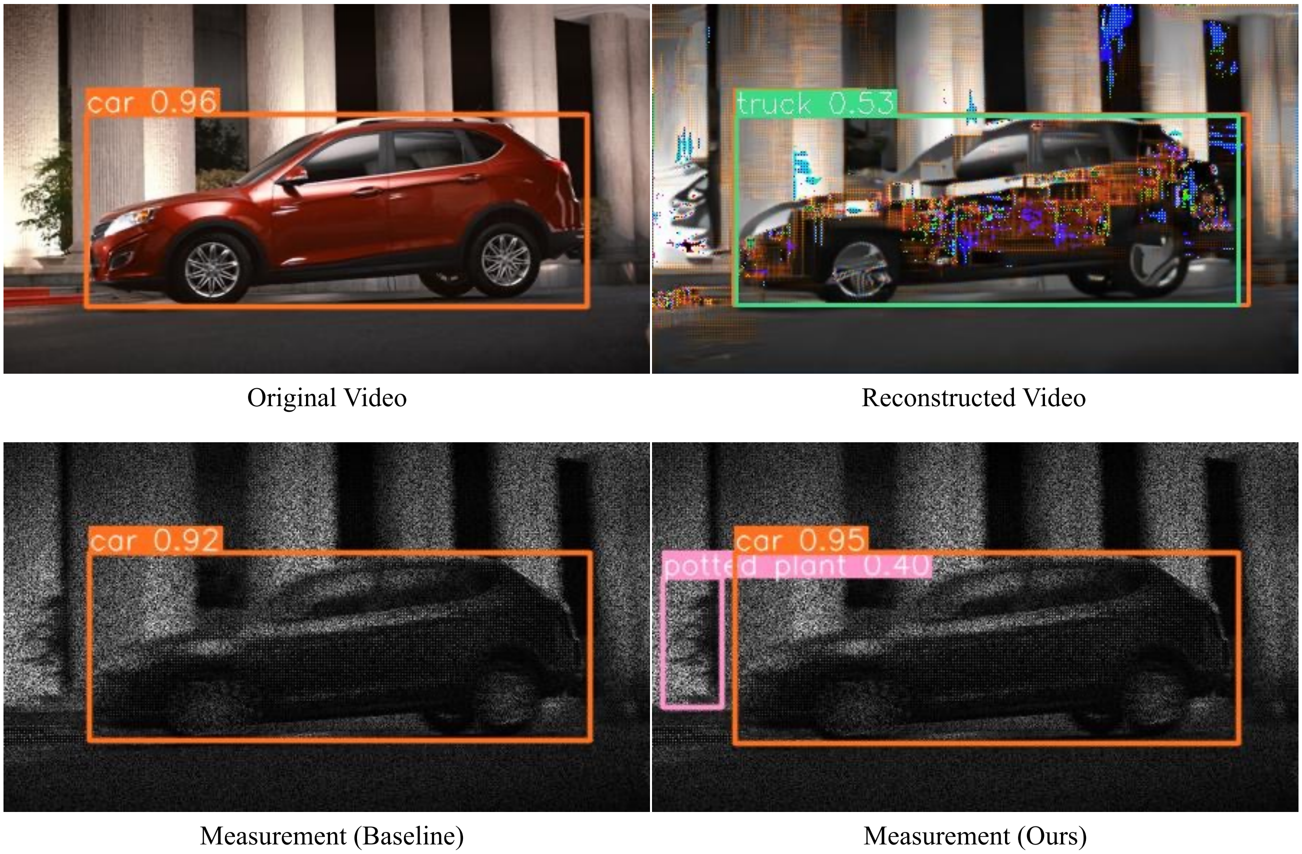

Figure 5. Comparison of object detection results using different configurations of the YOLO model: Original uncompressed video, compressed and reconstructed video, baseline compressed measurements, and optimized YOLO on compressed measurements. The figure illustrates detection confidence scores across setups, highlighting the original video's high accuracy in detecting a car and the unique identification of a potted plant by our optimized model, which was not detected in other configurations.

suffered from quality degradation, impacting the accuracy. The baseline method, without any adjustments, performed suboptimally. In contrast, our approach not only efficiently utilizes the high-dimensional information from compressed optical measurements but also incorporates AI-driven model adjustments to achieve the best results with optimized resource usage. These findings underscore the potential of our method in real-time EO applications, promising significant advancements in a variety of sectors.

4. CONCLUSION

In this study, we have demonstrated a transformative approach for optimizing object detection in EO systems through the integration of SCI and AI. Our method addresses the critical challenge of balancing accuracy and efficiency in EO systems by innovatively employing adaptive mask patterns[12] and compression rates,[13] tailored specifically to the needs of real-time operational environments.

The utilization of SCI significantly enhances the efficiency of optical signal capture by compressing high-speed video frames into single measurements without sacrificing the quality of the data required for accurate object detection. By directly processing these compressed measurements for object detection, our approach circumvents the traditional need for reconstructing high-dimensional optical signals, thereby reducing computational overhead and storage requirements.

Our experiments, conducted on the Vimeo-90K dataset, illustrate the effectiveness of our method. The results consistently show that our optimized model outperforms traditional object detection methods, not only in terms

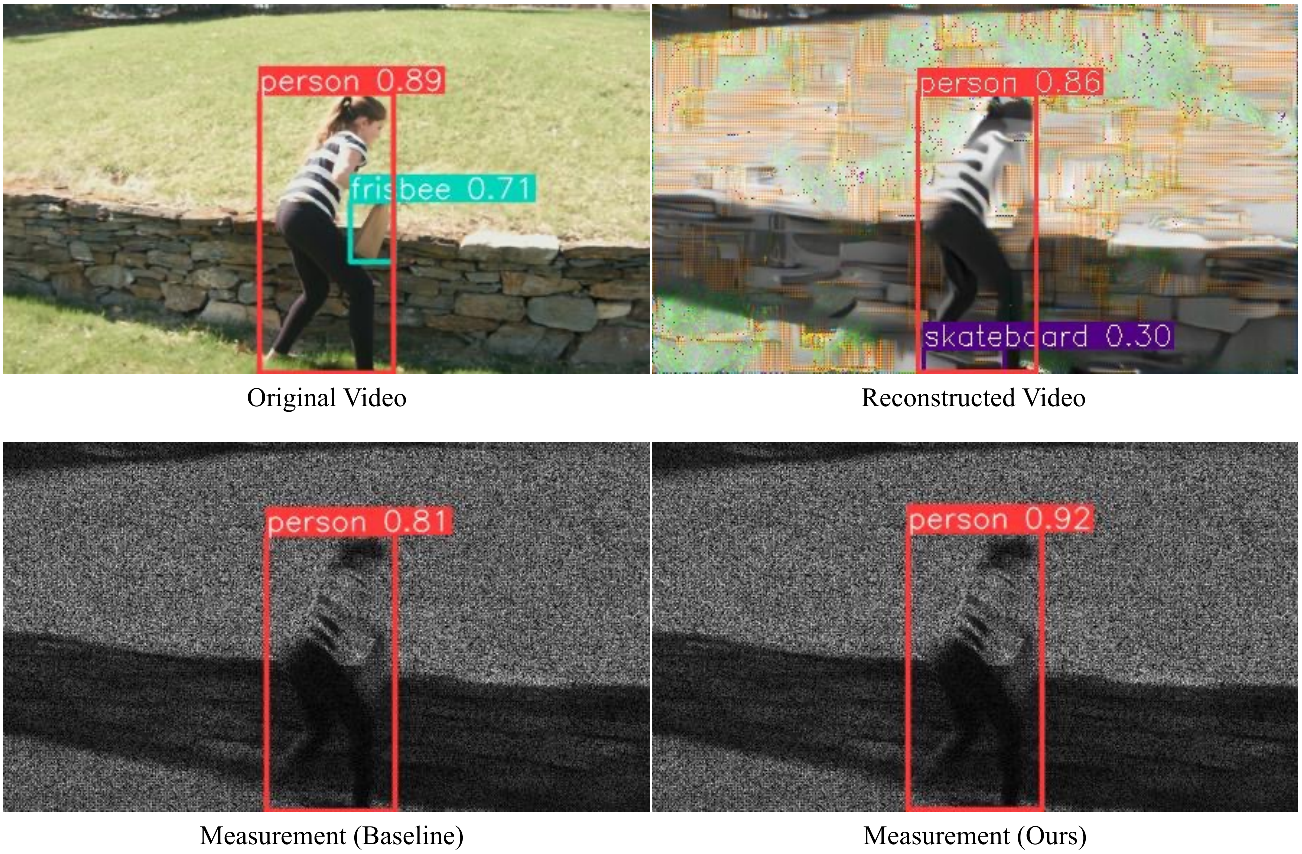

Figure 6. Comparison of object detection results using different configurations of the YOLO model. This figure showcases the original video incorrectly identifying a frisbee, the reconstructed video misidentifying a skateboard, and the baseline's lower confidence in detecting a person. Our optimized method not only improves person detection confidence but also eliminates erroneous detections, demonstrating the effectiveness of processing directly on compressed measurements.

of detection accuracy but also in the capability to detect additional objects that were previously undetected by conventional systems. This is particularly evident in our model's ability to accurately identify objects with high confidence under varying operational conditions, which is critical for a variety of applications where rapid and reliable object detection is paramount.

Furthermore, our approach's capability to adjust dynamically to different operational scenarios through real-time feedback enhances its applicability across a wide range of EO systems. This adaptability ensures that our method can meet the stringent requirements of modern operations as well as emerging civilian technologies such as autonomous driving and urban surveillance.

In conclusion, our research marks a significant advancement in the field of computational imaging within EO systems. By integrating SCI with AI-driven object detection, we provide a robust framework that not only meets but exceeds the current standards of operational efficiency and accuracy in real-time EO applications. We believe that our findings will pave the way for further innovations in the field and contribute to the evolution of smarter, more efficient EO systems.

ACKNOWLEDGMENTS

This work is supported in part by the Research Grants Council (GRF 17201822 and TRS T45-701/22-R) of Hong Kong SAR.

REFERENCES

[1] Hobbs, P. C., [*Building electro-optical systems: making it all work*], John Wiley & Sons (2022).

[2] Yuan, X., Brady, D. J., and Katsaggelos, A. K., "Snapshot compressive imaging: Theory, algorithms, and applications," *IEEE Signal Processing Magazine* **38**(2), 65–88 (2021).

[3] Zhao, Z.-Q., Zheng, P., Xu, S.-t., and Wu, X., "Object detection with deep learning: A review," *IEEE Transactions on Neural Networks and Learning Systems* **30**(11), 3212–3232 (2019).

[4] Zhang, Z., Zhang, B., Yuan, X., Zheng, S., Su, X., Suo, J., Brady, D. J., and Dai, Q., "From compressive sampling to compressive tasking: retrieving semantics in compressed domain with low bandwidth," *PhotoniX* **3**(1), 19 (2022).

[5] Prasad, D. K., Rajan, D., Rachmawati, L., Rajabally, E., and Quek, C., "Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey," *IEEE Transactions on Intelligent Transportation Systems* **18**(8), 1993–2016 (2017).

[6] Burt, P. J. and Kolczynski, R. J., "Enhanced image capture through fusion," in [*IEEE International Conference on Computer Vision*], 173–182 (1993).

[7] Bhaskaran, V. and Konstantinides, K., [*Image and video compression standards: algorithms and architectures*], Springer Science & Business Media (1997).

[8] Altunbasak, Y., Patti, A. J., and Mersereau, R. M., "Super-resolution still and video reconstruction from MPEG-coded video," *IEEE Transactions on Circuits and Systems for Video Technology* **12**(4), 217–226 (2002).

[9] Papageorgiou, C. P., Oren, M., and Poggio, T., "A general framework for object detection," in [*IEEE International Conference on Computer Vision*], 555–562 (1998).

[10] Lu, S., Yuan, X., and Shi, W., "Edge compression: An integrated framework for compressive imaging processing on cavs," in [*IEEE/ACM Symposium on Edge Computing (SEC)*], 125–138 (2020).

[11] Li, G., Zhao, Y., Ji, M., Yuan, X., and Fang, L., "Zoom in to the details of human-centric videos," in [*IEEE International Conference on Image Processing (ICIP)*], 3089–3093 (2020).

[12] Zhao, Y. and Lam, E. Y., "SASA: saliency-aware self-adaptive snapshot compressive imaging," in [*IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*], 2370–2374 (2024).

[13] Zhao, Y., Zeng, Q., and Lam, E. Y., "Adaptive compressed sensing for real-time video compression, transmission, and reconstruction," in [*IEEE International Conference on Data Science and Advanced Analytics (DSAA)*], 1–10 (2023).

[14] Zhao, Y., "Mathematical cookbook for snapshot compressive imaging," *arXiv preprint arXiv:2202.07437* (2022).

[15] Zhao, Y., Zheng, S., and Yuan, X., "Deep equilibrium models for video snapshot compressive imaging," *arXiv preprint arXiv:2201.06931* (2022).

[16] Yang, Q. and Zhao, Y., "Revisit dictionary learning for video compressive sensing under the plug-and-play framework," in [*SPIE Proceedings*], **12166**, 2018–2025 (2022).

[17] Chong, Z. Y., Zhao, Y., Wang, Z., and Lam, E. Y., "Solving inverse problems in compressive imaging with score-based generative models," in [*IEEE International Conference on Data Science and Advanced Analytics (DSAA)*], 1–10 (2023).

[18] Zhao, Y., Zheng, S., and Yuan, X., "Deep equilibrium models for snapshot compressive imaging," in [*AAAI Conference on Artificial Intelligence*], **37**(3), 3642–3650 (2023).

[19] Wang, L., Wu, Z., Zhong, Y., and Yuan, X., "Snapshot spectral compressive imaging reconstruction using convolution and contextual transformer," *Photonics Research* **10**(8), 1848–1858 (2022).

[20] Wang, L., Cao, M., Zhong, Y., and Yuan, X., "Spatial-temporal transformer for video snapshot compressive imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**(7), 9072–9089 (2023).

[21] Wang, L., Cao, M., and Yuan, X., "EfficientSCI: Densely connected network with space-time factorization for large-scale video snapshot compressive imaging," in [*IEEE/CVF Conference on Computer Vision and Pattern Recognition*], 18477–18486 (2023).

[22] Xue, T., Chen, B., Wu, J., Wei, D., and Freeman, W. T., "Video enhancement with task-oriented flow," *International Journal of Computer Vision* **127**, 1106–1125 (2019).

[23] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A., "You only look once: Unified, real-time object detection," in [*IEEE Conference on Computer Vision and Pattern Recognition*], 779–788 (2016).