*Cognitive Computing and Internet of Things,* Vol. 43, 2022, 137–144
https://doi.org/10.54941/ahfe1002702

# Cyber-Physical Behavior Detection and Understanding using Artificial Intelligence

**Zoheir Sabeur, Alessandro Bruno, Liam Johnstone, Marouane Ferjani, Djamel Benaouda, Banafshe Arbab-Zavar, Deniz Cetinkaya, and Muntadher Sallal**

Bournemouth University, Department of Computing and Informatics, Talbot Campus, Fern Barrow, BH12 5BB, Poole, England, United Kingdom

## ABSTRACT

The advancement of cyber-physical behaviour detection and understanding in context of urban environment safety and security has been developed in the S4AllCities project (S4AllCities, 2020). Various concepts of fundamental artificial intelligence have been successfully implemented and subsequently tested in situ in some early S4AllCities pilot sites since early 2022. The detection of anomalies in cyber traffic communication protocols which take place in context of urban spaces have been investigated. But these were principally complemented with primary research on the intelligent detection of crowd physical behavior in the same urban spaces. The aim is to fuse both modes (cyber and physical) of detection for behavior deeper understanding. Indeed, this advances situation awareness for further native knowledge base reasoning as far as safety and security operations go across the urban space. Native knowledge concerns the evaluated risks and mitigation measures for responses to potential cyber-physical attacks on the urban space. In this study, the deployed artificial intelligence techniques established good benchmarks for classifying physical behavior under various scenarios of potential attacks. Our future work is to exercise the scalability, performance, evaluation, and validation of our intelligent algorithms which are fused together using in situ cyber and physical observation scenarios of the urban spaces under the final S4AllCities pilot sites in Bilbao, Spain.

**Keywords:** Machine learning, Artificial intelligence, Intelligent agents, Crowd behavior, Computer vision

## INTRODUCTION

Urban spaces around the world are potentially exposed to serious attacks of cyber, physical, or both natures (Szczepaniuk, et al., 2022). With the ever-increasing number of digital communication networks around urban spaces, cities have become much more vulnerable to cyber-attacks on these networks as they could put cities infrastructure operations to a standstill. Furthermore, these spaces safety can also be compromised to physical attacks, as this has already been occurring in various cities during the last decades. Nevertheless, the interesting element of this is that smart spaces do offer valuable access to

 

real-time observation data and context information for scientists to develop state of the art concepts of intelligent agents and digital twins. These technologies are enablers for virtualizing the urban spaces, while they augment situation awareness for us and deepen our understanding of crowd behavior in urban spaces. In the S4AllCities project, we have been able to experiment the development of a digital twin for monitoring crowd behavior in three major smart spaces, located in three cities in Europe. These include the City of Trikala, Greece; Pilsen, Czech Republic; and Bilbao, Spain.

### S4AllCities Project

S4AllCities focuses on the development of three major digital twins, which specialized in; 1- Distributed Edge Computing IoT (or *DEC-IoT*); 2) Malicious Actions Information Detection System (or *MAIDS*); and c) Augmented Context Management System (or *ACMS*). In this paper, we will discuss the MAIDS digital twin functions since we are focused on behavior detection and understanding for the advancement of situation awareness with safety and security in urban spaces.

## MAIDS DIGITAL TWIN ARCHITECTURE

The MAIDS digital twin is composed of three main core modules. These are listed below:

- Module 1: Spatial data fusion
- Module 2: Cyber behaviour detection
- Module 3: Physical behaviour detection

As shown, in Figure 1 below, the three core modules of the MAIDS have access to observation data and information through data brokers which respectively subscribe to the DEC-IoT digital twin. Through this mechanism, the core modules process observation data from heterogeneous sources intelligently under the JDL data fusion framework (lambert, 2009). They first module fuses data from multi-modal sensor platforms and provides spatial observation trends of the space under surveillance, such as background noise levels, meteorological conditions and more. The second module specializes in detecting behavior in cyber communication networks, while the third module detects behavior of crowds who adhere to the urban space using computer vision and deep learning (Arbab-Zavar and Sabeur, 2020). These core modules specifically operate under the medium levels of fusion under the JDL framework while they extract important levels of situation awareness using machine intelligence (Sabeur, 2021).

### Cyber-Physical Intelligence at High Level Fusion and Reasoning

The combined output of the MAIDS core modules is further exploited downstream for higher level fusion and situation awareness, using native knowledge and context information of the urban cyber-physical space under surveillance (see Figure 1). This final step is then transmitted through via a message broker to which the ACMS subscribe, for presenting it as a rich Common Operational Picture (COP) of the urban space.

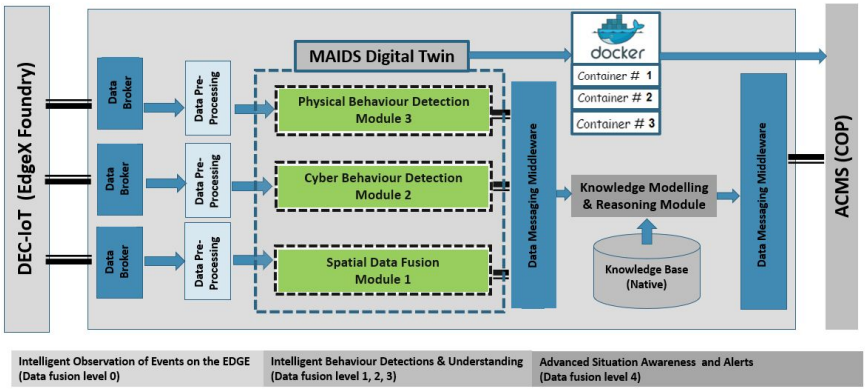


**Figure 1:** MAIDS Digital Twin overall modular architecture.

### MAIDS Dockerisation for Situation Awareness in the COP

There is also another option for the COP to present advanced situation awareness to security practitioners. This is usually for the purpose of performing training sessions by setting "what if" scenarios of cyber-physical attacks in the urban space. This is done by accessing the MAIDS core modules directly through their respective dockerization into containers within the S4AllCities system. Dockerizing is the process of creating, deploying, and running applications with all their necessary files dependencies in a container computer environment as their own image using Docker containers. Docker technology allows you to package any application (such as those coded in Python) with all the necessary code (e.g., libraries) as one container which can be uploaded, loaded and run on any other platform (such as cloud, remote VMwares servers, workstations, etc.), in our case ACMS and the COP. Docker is also an open-source tool, which assures interoperability with legacy city surveillance systems.

## CYBER BEHAVIOUR UNDERSTANDING

### Cyber Traffic Unusualnesss Detection

The threat landscape of computing and cybersecurity is everchanging and evolving, to aid in combating emerging and existing attacks on cities communication networks. The cyber behavior detection module uses several machine learning algorithms to detect unusualness in the network flow of data. The module has two main functions which are Detection and Integration. Under the detection function, named as CyberATDetect, several processes are conducted which all lead to behavior detection, as shown in Figure 2. These processes are as follows:

- Data ingestion and Harvesting: This stage refers to the collection and exploration of data used to design and test the network intrusion detection.
- Training: This stage uses datasets to train and test machine learning algorithms, and evaluates their performance under strictly defined metrics

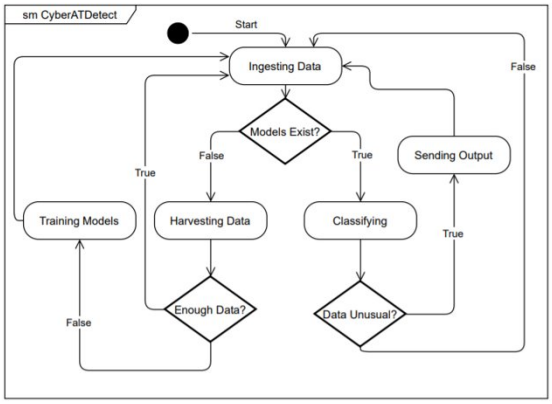


**Figure 2:** Cyber behavior detection module processes.

- Classification: Classifies an overall cyber behavior, as usual or unusual, while fusing the various classifiers, with associated weights for normalizing the resulting Correct Classification Rate (CCR)
- Messaging Output: Sending the resulting classification with CCR via a messaging broker to the ACMS digital twin for visualization in the COP environment.

## DISCUSSION

The cyber behavior module will be expanded to investigate the classification of various types of cyber behavior attacks, while it will be put under experimental tests for its scalability. Simultaneous attacks, including those with larger floods of cyber traffic will be subsequently investigated using our computer clusters and Apache Spark technologies for distributed processing and machine learning in the S4AllCities project.

## PHYSICAL BEHAVIOUR DETECTION

In light of Hou et al.'s (2019) work on vehicle tracking, the integration of state of the art deep learning approaches for object detection and tracking namely, YOLOv5 and DeepSort is realized (Jocher et al., 2022), (Wojke et al., 2017). YOLO's latest version architecture which extends YOLOv4 is considered as the most accurate since it is coupled with high inference speed (Bochkovskiy et al., 2020). Also, DeepSort tracks the movement of individuals in crowds over video sequences while using Kalman filters and target association. For a given video frame, having M pedestrians, $P\left(x,y\right)_{i\,=\,1,\ldots,M}$ signifies the spatial coordinates for an $i^{\text{th}}$ pedestrian. As seen in Figure 3, YOLOv5 detects pedestrians although it lacks the ability to re-identify what was previously detected. DeepSort is therefore added to overcome this issue and achieve the tracking of detected pedestrians in a video sequence. This is done by associating reference numbers on them. That is, across different time slices $(t_0,\ t_1,\ldots,t_N)$, DeepSort simply maintains records of $P\left(x,y\right)_{i\,=\,1,\ldots,M}$. As illustrated in Figure 4, the tracking of a pedestrian (ID = 1) between

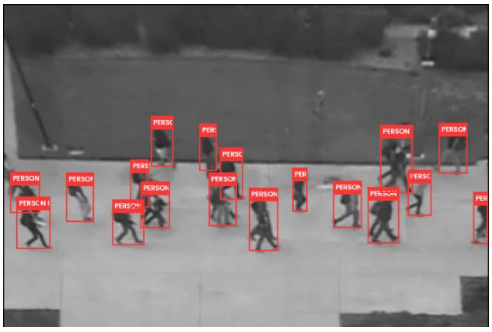

**Figure 3:** Pedestrians' detection using YOLOv5.

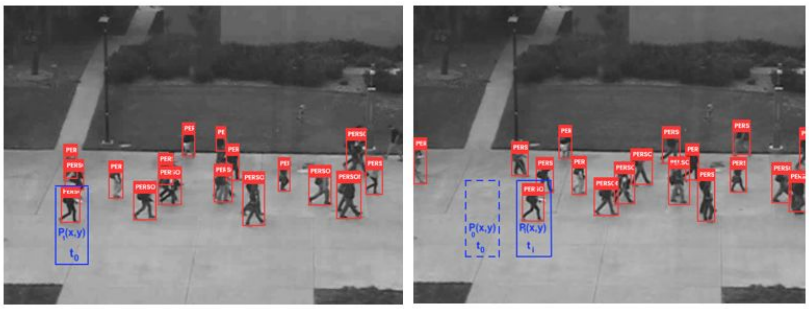

**Figure 4:** Pedestrians detection and tracking in time.

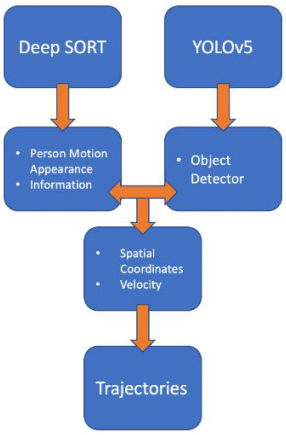


**Figure 5:** Flow chart of YOLO5 and DeepSort integration.

2 frames is highlighted. YOLOv5 yields a sequence of bounding boxes to depict the spatial coordinates of the detected pedestrians. As shown in Figure 5, a sequence of bounding boxes are depicted with the spatial coordinates of the detected pedestrians.Then, DeepSort determines the relationship between existing tracks and measurements under "measurement-to-track" associations using the Mahalanobis distance.

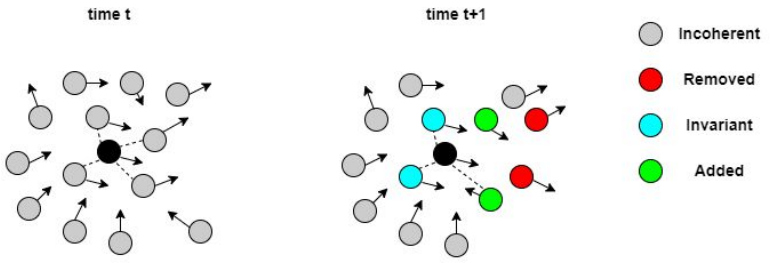


**Figure 6:** Exhibition of coherent neighbor invariance. Green dots are viewed as invariant neighbors of the centered black dot (for K = 4).

## Trajectories Clustering

The extraction of trajectories allows us understanding the detection of coherent physical behavior which represent crowd motions in urban spaces. For instance, with video surveillance, one captures coherent motions, which are exhibited by moving pedestrians. This leads us into acquiring high-level representations of crowd mechanics using computer vision (Dogbe, 2012). These representations combined with measurements of the environmental conditions using the MAIDS spatial fusion module can be utilized in completing crowd behavior, actions and situations understanding altogether.

Whilst coherent motions are regarded as macroscopic observations of pedestrian's congregational activities, these motions are distinguished through interaction among individuals in local neighborhoods. Inspired from (Zhou et al., 2012), Coherent Neighbor Invariance technique is deployed to capture coherent motion of crowd clutters. The key characteristics that establish the difference between cohesive and arbitrary movements are listed below:

- **Neighborship Invariance:** the spatial-temporal relationship among individuals is inclined to prevail overtime.
- **Velocity Correlations Invariance:** Neighboring individuals exhibiting coherent movements showcase high velocity correlations.

Conversely, incoherent individuals that showcase relative independence tend to lack the mentioned properties. To illustrate the Neighborship Invariance property, Figure 7 displays the use of K nearest neighbor to highlight the emergence of global coherence in local neighborships. Figure 7 shows the deployment of the coherent filter on the UCSD dataset. The equation below quantifies the velocity correlations between neighboring individuals, which allows discerning coherent motions.

$$g = \frac{1}{d+1} \sum_{\tau=t}^{t+d} \frac{v_\tau^i \,.\, v_\tau^{i_k}}{\left\| v_\tau^i \right\|^2 \,.\, \left\| v_\tau^{i_k} \right\|^2}$$

Where:

- $g$ : *velocity correlation between i and* $i_k$
- $v_\tau^i$ : *velocity of individual i at time* $\tau$


**Figure 7:** Coherent motion detection in action.

- $v_\tau^{i_k}$ : *velocity of individual* $i_k$ at time $\tau$
- $d$ : *duration of the experiment*

## CONCLUSION AND FUTURE WORK

The advancement of our work in the S4AllCities project has been steadily progressing since 2020 and we have now reached a position to further our experimentation on much larger volume and heterogeneous cyber traffic data as well as multi-modal sensor observation data. This will be explored in the final phase of the S4AllCities project where multi-level fusion and reasoning takes place. With the deployment of multiple specialized sensors and the streaming of background noise and meteorological condition signals in the urban space, monitoring crowds with optical and infra-red cameras and keeping attention to cyber traffic, one will lead to enriching the MAIDS digital twin with more intelligence for ACMS and the COP. This will be tested un situ during the last pilot of the S4AlCities project in Bilbao, Spain.

## ACKNOWLEDGMENT

The authors would like to thank the European Commission for partly funding our research work in the S4AllCities project, under the H2020 Programme, Grant Agreement No. 883522.

## REFERENCES

Arbab-Zavar, B., Sabeur, Z. (2020). Multi-scale crowd feature detection using vision sensing and statistical mechanics principles. Machine Vision and Applications 31:26 https://doi.org/10.1007/s00138-020-01075-4

Bochkovskiy, A., Wang, C.-Y. & Liao, H.-Y. M. (2020), 'Yolov4: Optimal speed and accuracy of object detection', arXiv preprint arXiv:2004.10934.

Dogbe, C. (2012), 'On the modelling of crowd dynamics by generalized kinetic models', Journal of Mathematical Analysis and Applications 387(2), 512–532.

Hou, X., Wang, Y. & Chau, L.-P. (2019), Vehicle tracking using deep sort with low confidence track filtering, in '2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)', IEEE, pp. 1–6.

Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., NanoCode012, Kwon, Y., TaoXie, Fang, J., imyhxy, Michael, K., Lorna, V, A., Montes, D., Nadar, J., Laughing, tkianai, yxNONG, Skalski, P., Wang, Z., Hogan, A., Fati, C., Mammana, L., AlexWang1900, Patel, D., Yiwei, D., You, F., Hajek, J., Diaconu, L. & Minh, M. T. (2022), 'ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference'. URL: https://doi.org/10.5281/zenodo.6222936

Kwon, D., Kim, H., Kim, J., Suh, S. C., Kim, I., Kim, K. J. (2019). A survey of deep learning-based network anomaly detection. Cluster Computing. https://doi.org/10.1007/s10586-017-1117-8

Lambert, D.A. (2009). A Blueprint for higher-level fusion systems. Information Fusion, Vol. 10, Issue 1. pp. 6–24.

S4AllCities: Smart Spaces Safety and Security in All Cities. (2020) Available from https://www.s4allcities.eu/project

Sabeur, Z. (2021) AI3SD Video: Artificial Intelligence for Safer Urban Space. Frey, J. G., Kanza, S. and Niranjan, M. (eds.) AI3SD Autumn Seminar Series 2021. 13 Oct - 15 Dec 2021. doi:10.5258/SOTON/AI3SD0173

Szczepaniuk, E.K. and Szczepaniuk, H., 2022. Analysis of cybersecurity competencies: Recommendations for telecommunications policy. Telecommunications Policy, 46(3), p.102282.

Wojke, N., Bewley, A. & Paulus, D. (2017), Simple online and realtime tracking with a deep association metric, in '2017 IEEE International Conference on Image Processing (ICIP)', IEEE, pp. 3645–3649.

Zhou, B., Tang, X. & Wang, X. (2012), Coherent filtering: Detecting coherent motions from crowd clutters, in A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato & C. Schmid, eds, 'Computer Vision – ECCV 2012', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 857–871.