HEIMDALL: an AI-based infrastructure for traffic monitoring and anomalies detection

Andrea Atzori*, Silvio Barra†, Salvatore Carta*, Gianni Fenu* and Alessandro Sebastian Podda*

*Department of Mathematics and Computer Science, University of Cagliari
Via Ospedale 72, 09124 Cagliari (Italy)
Email: a.atzori28@studenti.unica.it, {salvatore, fenu, sebastianpodda}@unica.it

†Dept. of Electric Engineering and Information Technology, University of Naples "Federico II"
Via Claudio 21, 80125 Napoli (Italy)
Email: silvio.barra@unina.it

Abstract—Since their appearance, Smart Cities have aimed at improving the daily life of people, helping to make public services smarter and more efficient. Several of these services are often intended to provide better security conditions for citizens and drivers. In this vein, we present HEIMDALL, an AI-based video surveillance system for traffic monitoring and anomalies detection. The proposed system features three main tiers: a ground level, consisting of a set of smart lampposts equipped with cameras and sensors, and an advanced AI unit for detecting accidents and traffic anomalies in real time; a territorial level, which integrates and combines the information collected from the different lampposts, and cross-correlates it with external data sources, in order to coordinate and handle warnings and alerts; a training level, in charge of continuously improving the accuracy of the modules that have to sense the environment. Finally, we propose and discuss an early experimental approach for the detection of anomalies, based on a Faster R-CNN, and adopted in the proposed infrastructure.

Index Terms—smart cities, artificial intelligence, anomalies detection

I. INTRODUCTION

The times in which we live are characterised by an ever-increasing adoption of the so-called *smart objects*. This definition arises from the fact that the use of such objects is able to confer "smartness" to a specific environment. Indeed, these devices are actually sensors, in charge of three main goals:

- *sense* the environment;
- *collect* information from the environment;
- *take actions* within the environment;

and their application to the wide range contexts of everyday life has given rise to concepts such as those of Smart City, Smart Home, Smart Car, to name a few.

Nowadays, scientific and industrial research are very active in this direction, since it involves a large number of topics, including, e.g., pattern recognition applications, cloud-based systems, cryptography and several others. In this regard, Smart Cities certainly represent one of the most studied environments, since the definition of Internet of Things (IoT) takes here a wider and more interesting meaning. In fact, many applications take advantages from the adoption of IoT networks in Smart Cities, including (but not limited to) biometrics [1], [2], traffic light management [3], [4], video surveillance [5], [6], crowd analysis [7] (especially in situations in which gathering of people is not allowed [8]) and anomalies detection.

However, given the *in-the-wild* nature of the urban scenarios, many of the aforementioned applications need to be tailored to the difficulties and challenges of such a typical unconstrained environment: here, visual sensors like cameras may present issues due to light variations, pose, occlusion and perspective [9]; microphones may suffer interference from nearby devices or distortions when loud noises are recorded [10]; wearable sensors are often affected by the sudden movements of the subject which may tend to cause artifact and noise during the acquisition stage [11].

Additional difficulties may also arise from the management of master-slave camera systems, where the direction, coordination, calibration and communication of cameras is fundamental for a proper video surveillance activity [12]. Nevertheless, data heterogeneity is another issue which designers need to deal with. The challenge, from this perspective, lies in the fact that different sources gather different data in different format. Fusion algorithms may therefore operate at different levels (sensor, feature vector, score and decision), hence resulting in different grades of precision [13].

In the light of the above, in this paper we present HEIMDALL, an AI-based video surveillance infrastructure for traffic monitoring in Smart Cities. The main contributions of this paper are summarized below:

- the design and description of an end-to-end IoT system for detecting and preventing traffic anomalies that may potentially cause danger to road users, within a urban scenario;
- a novel three-tier architecture, in which a ground layer consisting of a set of autonomous smart lamppost is coordinated by an advanced centralised unit that acts at a territorial level, both supported by a modelling layer that continuously trains and improve the involved AI modules;
- an early experimental approach for the automatic detection of anomalies, that leverages on Deep Neural Networks (DNNs), and in which the training is carried out
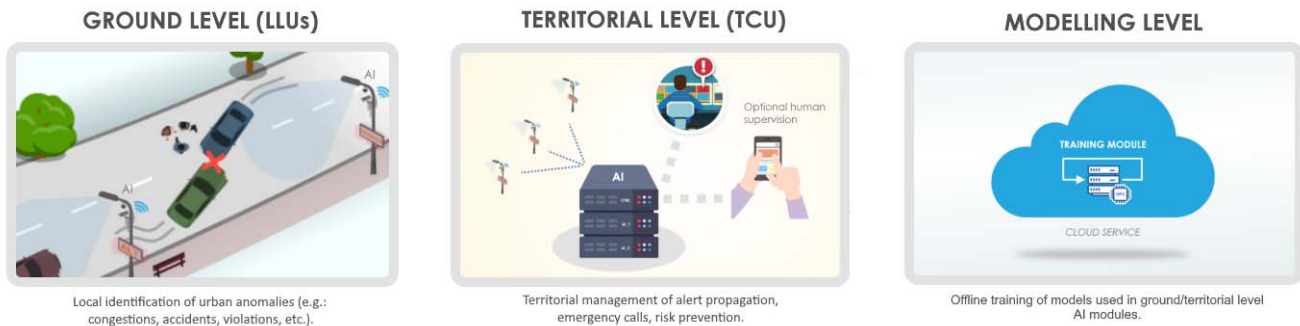
Fig. 1. The three main subsystems of Heimdall.

by combining synthetic and real-world datasets in order to achieve high effectiveness in detecting both static and dynamic anomalies;

- last generation NVIDIA Jetson AGX Xavier card is used as integrated hardware board embedded in the lampost.

The remaining of the paper is organized as it follows: Section II briefly describes the related work in the literature. Section III presents the architecture of HEIMDALL, while Section IV illustrates the preliminary proposed approach for the automatic detection of anomalies. Finally, in Section V, a timely discussion of the main problems and limitations related to the used techniques is inserted, with a particular emphasis on the proposed infrastructure.

II. STATE OF THE ART OF THE CURRENT SYSTEMS

The massive deployment of the video surveillance infrastructures in modern cities is mainly due to the request for security and safety for pedestrians and drivers [14], [15]. Systems aimed at ensuring road safety have been active in various parts of the world between the second half of the 90s and the beginning of the new century [16], [17], when video cameras were installed at road junctions in many metropolitan areas, to monitor road sections. In most cases, accident and anomalies detection were carried out using feature extraction techniques that based their operation on the study and generalisation of patterns within images and videos, to recover the same pattern from successive video streams, in times far removed from real time. Many of these techniques are still used today [18], but have given way to methodologies that allow for a better generalization and exploit more advanced tools. Indeed, in the last decade, the theoretical background on artificial intelligence, introduced and deepened in the 80s, have began to show the first practical applications, thanks to the adoption of new high-performance processing systems, thus leading to the fast emergence of modern Machine Learning techniques and, subsequently, of their more extensive branch, the Deep Learning. These methods have undermined canonical feature extraction techniques, which, although they do not require much data to work acceptably, do not always guarantee high reliability and accuracy at the generalisation stage. Deep Learning is now widely used in video and image analysis in

a variety of research areas, ranging from medical diagnostics [19], [20], financial forecasting [21], [22] and cybersecurity [23], and over the years has also had a positive impact on road safety [24]. The state of the art presents considerable scientific solutions to the problem of road safety, providing inspiration for the development of methods for pedestrian detection [25], vehicle tracking [26], and accident detection [27]. The solution proposed for this project, compared to the state-of-the-art, presents several interesting innovations, introduced with the aim of creating a safe and safeguarded environment for vehicles and pedestrians. First of all, its architecture composed of environmental sensors, hence strongly based on the Internet Of Things concept, brings the idea into the Smart Cities domain, within which so far only theoretical frameworks, with a limited development phase, have been proposed [28]. On the other hand, from a technical point of view, the proposed infrastructure employs video-summarization methodologies for the reduction of the computational load to be transmitted to central and geographically adjacent processing systems. This issue is very much in vogue nowadays, especially with regard to Edge Computing [29] and thus new research directions that aim to migrate part of the processing effort to systems on the edge of the cloud.

III. HEIMDALL: THE AL(S)IGHTY

The goal of HEIMDALL is to automatically detect traffic anomalies in urban or extra-urban scenarios, mainly in order to prevent critical situations and accidents. In this context, the system exploits the state of the art of AI technologies (specifically, those based on Machine Learning and Deep Learning techniques), in order to support the action of traffic wardens, as well as to implement appropriate warning measures to protect vehicles and pedestrians in the involved areas.

HEIMDALL consists of three main subsystems, strongly interconnected, whose coordinated action guarantees the effectiveness of detection and prevention. These subsystems, depicted in Figure 1, can be summarized as it follows:

- the *ground level*, composed by a large number of local AI-based units, embedded on roadside smart lampposts, for real-time detection of traffic anomalies and accidents;

- the *territorial level*, which collects data from the lampposts, combining them with information from external sources (e.g. weather, civil protection, traffic conditions), in order to globally manage the alerts, propagate them and assist the human operator responsible for supervising the territory;
- the *modelling level*, that consists of an offline training module to upgrade the neural networks according to the newly observed traffic data.

In next subsections, we now provide additional details on the aforementioned subsystems.

A. Ground level

The core of the ground level consists of a local Decisional Support System (DSS), embedded on a special-purpose processor installed on the HEIMDALL's lampposts. We defined such a DSS as the *Lamppost Local Unit* (LLU). It provides an AI module that acquires, in real-time through the on-board cameras, high-resolution images and videos involving the current traffic situation in the observed area. Then, the acquired data are automatically augmented with vehicles speeds and spatial correlations.
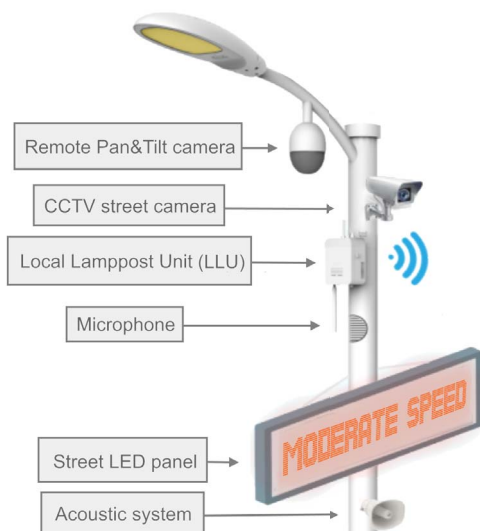

Fig. 2. Hardware scheme of the smart lamppost.

To this purpose, the module on the LLU leverages on a *YOLO (You Only Look Once)* system [30], a pretrained SSD (Single Shot Detector) based on a deep Convolutional Neural Network (CNN) that, by processing a video frame, returns information on the position and dimensions of any object inside the frame, denoting it with a bounding box. Hence, a second software component, consisting of a Faster R-CNN (detailed in Section IV-B) uses the output of YOLO to identify the presence of anomalies. In the event an anomaly is detected, the LLU reports its inferred typology and associated *criticality index* ϕ (with $\phi \in [0, N]$), together with the area-specific metadata (audio/video recordings, MSD, *etc.*), to the central unit that constitutes the Territorial level.

The lamppost is also equipped with sound and light signalling devices which, depending on the severity of the anomaly detected, may be activated in order to invite drivers to moderate their speed, or to signal the accident, as well as receive remote commands from the human supervisor. Finally, we should note that the complete software that powers the LLU, including the AI module that executes the deep neural networks, runs on a single integrated hardware board, embedded in the lamppost. In our configuration, this board is represented by an NVIDIA *Jetson AGX Xavier*[1], a compact and low-power system-on-module (SoM), which features a Volta GPU, two Deep Learning Accelerators (DLAs), an octa-core Carmel CPU and 32GB of memory, provided with a pre-installed Linux distribution and a dedicated developer kit. The full hardware scheme of the smart lamppost is summarized in Figure 2.

B. Territorial level

The territorial level represents a higher level of HEIMDALL's architecture: specifically, its core module is mainly composed by the *Territorial Control Unit* (TCU). With respect to the Ground level devices, this unit performs a wider range analysis, since: *i)* it gathers the local information and criticality index coming from the single LLUs, with the aim of validating and managing the local alerts, and *ii)* it is in charge of predicting possible risk situations which could affect the traffic circulation, in order to prevent accidents.

To achieve this goal, the TCU collects data coming from the LLUs installed on smart lampposts, combining them with the information retrieved from external sources (e.g., weather services, public utility information, civil protection alerts). By geo-locating such information, the TCU employs an additional level of AI, primarily based on heuristics and control thresholds, where each criticality index ϕ_i, sent by the i-th LLU, is reassessed on the basis of a *global risk index* λ (with $\lambda \in [0, M]$). In formulae:

$$\varphi_i = \max\{\lceil \phi_i + \alpha\lambda \rceil, N\}$$

where φ_i is the *reassessed criticality index* (associated to the local area monitored by the smart lamppost i), while $\alpha \in [0, 1]$ is the parameters for weighting the impact of the global index, tuned according to the specific properties of the considered territory. It should also be noted that this index is rounded up, and can never exceed N, which represents the maximum level of criticality expected for a certain area.

Given φ values, the TCU can therefore operate in several ways. First of all, it can automatically propagate alert signals between proximal smart lamppost (e.g. the active acoustic and light warnings). Secondly, it can activate a *preventive warning*, if it detects the possibility of imminent critical situations, such as rain risk.

However, although the TCU is set up to handle a large set of anomalies in a fully automatic way, the system is still set

up for supervision by a human operator. Through a responsive *management dashboard* (shown in Figure 3), both accessible from desktop and mobile devices, the operator is able to filter false positives/negatives, thus improving the accuracy of the different AI layers, as well as managing an intelligent queue of notifications related to active alerts (sorted by priority based on timestamp and criticality index), to deactivate or propagate further.
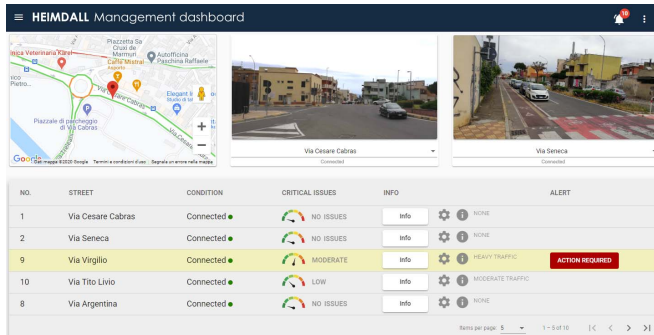

Fig. 3.   The management dashboard of HEIMDALL.

C. Modelling level

The last tier of HEIMDALL is the modelling level, through a dedicated Training Module. Its main goal is the definition and training of the models involved in the prediction and classification operations which rule the AI modules of the LLUs and TCU. This subsystem runs on an *ad-hoc* cloud service, and it is not required to work in real-time with the first two tiers. The activities performed by this level can be summarized in four distinct but still interconnected phases:

- **Phase 1**: collecting training data, merged with additional features from outer sources (e.g., weather forecasting databases, historical series providers, and/or further local traffic monitoring systems);
- **Phase 2**: training the involved networks;
- **Phase 3**: validating the models and tuning of the hyper-parameters; testing of the tuned networks (repeat Phase 2 when necessary);
- **Phase 4**: deploying of trained models on the destination units.

IV.   ANOMALIES DETECTION

The main issues faced by means of the HEIMDALL system are related to traffic monitoring tasks, with special attention to the detection of anomalous events in Smart Cities and urban scenarios. This class of problems can be handled in two different ways:

- the first one requires to build the model by learning from the scene which events can be considered as *normal*, in order to detect those which deviate from the regular behavior [31];
- the second one requires to address a supervised object detection problem: in this case, the model is trained over

the anomalous events, in order to generalise on the new videos.

We observe that there exist several pros and cons in relation to both the approaches: the first one could potentially be more accurate, but it needs a very considerable training phase, in order to let the system discriminate which scene contains an anomaly. Moreover, in this context, if a particular anomaly occurs with regularity, it can be missed by the model. The second approach, which is currently a preliminary proposal, is the one adopted for the HEIMDALL infrastructure: a set of anomalies is listed *a-priori*; then, all items in this set are labelled and passed as a training set to a Convolutional Neural Network (CNN), in charge of learning the pattern related to each type of anomaly. The adopted CNN architecture is a Faster R-CNN, which is discussed in details in the remainder of this section.

A. Dataset and annotated anomalies

In order to achieve a reliable training of the network, we have taken into consideration some preliminary factors: (i) the population of the dataset; indeed, in order to achieve a proper anomalies detection in urban scenarios, a variegated dataset needs to be used and labelled too. At the state-of-the-art, some datasets exist, but either proper annotations are missing or the camera view is not suitable for video surveillance systems [32]; (ii) the kind of anomalies to be detected; in most scenarios, only anomalies leading to extreme events are annotated, e.g. accidents or pedestrians hitting. However, since the anomalies we aim to detect cover wider aspects of the urban education, we decided to proceed with an *ad-hoc* acquired dataset, so to include a large number of specific events.

Fig. 4.   Five kinds of annotated anomalies of the collected dataset. From the top-right, clockwise: *illegally parked vehicle*, *risky overtaking*, *vehicle on pedestrian area*, *red light violation*, *vehicle-vehicle collision*.

Hence, for the training phase, we built and annotated a large set of videos (currently, they are ~ 500, but the number is still increasing), featuring several types of events, which can be considered anomalous or not depending on the kind of scenario application. Such videos are synthetically generated, and acquired from the *Grand Theft Auto V* videogame from Rockstar Games. This solution provided a good trade-off between the realism of the simulations and the configuration

of the acquisition, along with the simplicity and the velocity of reproducibility of different, even rare, events. We also observe that such technique has already shown to be very effective in the literature, as in [33], where the authors created the JTA (Joint Track Auto) dataset to estimate subject poses. In Figure 4, some examples of the annotated anomalies are shown.

B. Faster R-CNN

The Training module of the modelling level, detailed in the previous section, involves the training of a Faster R-CNN [34], with a backbone consisting of a VGG-16 net. This kind of network is a Region Proposal Network (RPN), since it involves a two step process for detecting objects within a scene:

- the first step is the region proposal one, in which a set of regions within the image is firstly selected for subsequent classification;
- the second step is in charge of classifying the single regions, in order to verify whether an object of interest is present.

The architecture of such a network is depicted in Figure 5, where the Region Proposal Network and the Faster R-CNN network respectively achieve the aforementioned steps. Please refer to the work in [34] for Faster R-CNN implementation and architecture details.
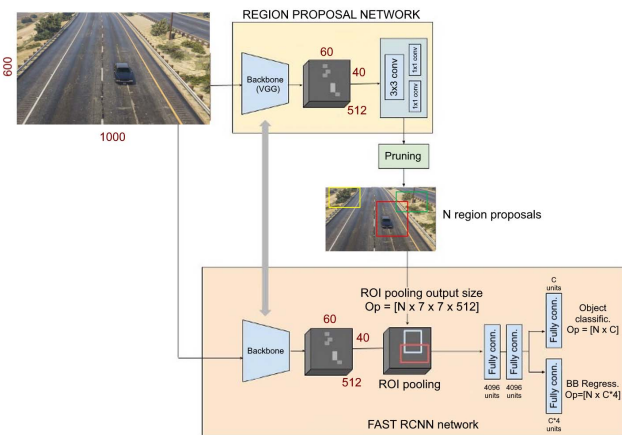
Fig. 5. The Faster R-CNN used in the Training Module for classifying events.

In order to fasten the deploy of the network, we exploited the `detecto` framework, to allow a quick training phase configuration, thus producing a trained network. In order to properly use the framework, the images have been resized to $600X1000$. The dataset has been labelled by means of the tool `LabelImg`, which facilitates the selection and annotation of the regions within the frames of a video. The preliminary qualitative results obtained, shown in Figure 6, reveal a twofold interesting insight:

- on synthetic videos, the network results very accurate, especially on *static* anomalies (which do not imply a time component); but, nonetheless, some dynamic anomalies are still well detected (e.g., *overtaking* or *U-turns*). Due to the adopted approach, all the anomalies related with

vehicle speed are not currently detected, since we adopted no method for relating consecutive frames at this stage;
- even when executed on real scenario videos, the Faster R-CNN is able to detect most of the anomalies, especially when the camera view is similar to that of dataset samples and no major occlusions are present, even under dense traffic conditions. Some examples are shown in Figure 7.

Fig. 6. Some examples of correct anomaly detection, when the trained Faster R-CNN is executed on synthetic test videos. From the top-right, clockwise: *traffic collision*, *vehicle on pedestrian area*, *car in wrong way*, *risky overtaking*, *car in no parking area* and *red light violation*.

Fig. 7. Some examples of correct anomaly detection, when the trained Faster R-CNN is executed on real-world videos. From the top-right, clockwise: *traffic congestion*, *car in wrong way*, *traffic collision* and *risky u-turn*.

V. DISCUSSION, CONCLUSIONS AND LIMITATIONS

We have presented HEIMDALL, an AI-based video surveillance infrastructure composed by three tier, respectively in charge of *i)* sensing the urban environment, *ii)* coordinating the communications and *iii)* training the AI models involved. Our proposed infrastructure employs a YOLO system, at the ground level, for tracking vehicles and pedestrians, and exploits a preliminary approach based on a Faster R-CNN, for the detection of potential anomalies in the scene. Specifically, the learning of such a CNN is in charge of the modelling level, so to separate the computing burden needed for the training phase, from the actual deployment step, which takes place on the Jetson AGX Xavier card embedded in the lamppost.

The anomalies detection module has shown interesting insights from a qualitative point of view: even though it was mainly trained on synthetic videos (generated by means of the Rockstar editing tool over the *GTA-V* videogame), it has proven to be capable of providing good results also in the real-world scenario. However, it still presents some limitations, mainly due to two aspects, which have become clearer by following the downstream analysis on the working

tools: (i) some anomalies which implies dynamic analysis are cumbersome to detect, since, natively, the Faster R-CNN does not take into consideration the relationships between the objects in consecutive frames; (ii) to improve the accuracy of the detection, particularly in the presence of widely differing scenes, it could be useful to exploit a domain adaptation approach [35] to keep unaltered the training phase, while modifying internally the learnt filters and adapting them to the new scenario. These elements represent significant milestones for further improvements, and they are currently planned for a future release of the discussed infrastructure.

REFERENCES

[1] S. Barra, A. Castiglione, M. De Marsico, M. Nappi, and K.-K. R. Choo, "Cloud-based biometrics (biometrics as a service) for smart cities, nations, and beyond," *IEEE Cloud Computing*, vol. 5, no. 5, pp. 92–100, 2018.

[2] M. De Marsico, A. Mecca, and S. Barra, "Walking in a smart city: Investigating the gait stabilization effect for biometric recognition via wearable sensors," *Computers & Electrical Engineering*, vol. 80, p. 106501, 2019.

[3] X. Liang, X. Du, G. Wang, and Z. Han, "A deep reinforcement learning network for traffic light cycle control," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1243–1253, 2019.

[4] S. You, Q. Bi, Y. Ji, S. Liu, Y. Feng, and F. Wu, "Traffic sign detection method based on improved ssd," *Information*, vol. 11, no. 10, p. 475, 2020.

[5] A. Rego, A. Canovas, J. M. Jiménez, and J. Lloret, "An intelligent system for video surveillance in iot environments," *IEEE Access*, vol. 6, pp. 31 580–31 598, 2018.

[6] J. C. Neves, J. C. Moreno, S. Barra, and H. Proença, "Acquiring high-resolution face images in outdoor environments: A master-slave calibration algorithm," in *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2015, pp. 1–8.

[7] W. Liu, M. Salzmann, and P. Fua, "Context-aware crowd counting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5099–5108.

[8] R. C. Chick, G. T. Clifton, K. M. Peace, B. W. Propper, D. F. Hale, A. A. Alseidi, and T. J. Vreeland, "Using technology to maintain the education of residents during the covid-19 pandemic," *Journal of Surgical Education*, 2020.

[9] J. C. Neves, G. Santos, S. Filipe, E. Grancho, S. Barra, F. Narducci, and H. Proença, "Quis-campi: Extending in the wild biometric recognition to surveillance environments," in *International Conference on Image Analysis and Processing*. Springer, 2015, pp. 59–68.

[10] T. Van Renterghem, P. Thomas, F. Dominguez, S. Dauwe, A. Touhafi, B. Dhoedt, and D. Botteldooren, "On the ability of consumer electronics microphones for environmental noise monitoring," *Journal of Environmental Monitoring*, vol. 13, no. 3, pp. 544–552, 2011.

[11] A. F. Abate, M. Nappi, S. Barra, and M. De Marsico, "What are you doing while answering your smartphone?" in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 3120–3125.

[12] J. C. Neves, J. C. Moreno, S. Barra, and H. Proença, "A calibration algorithm for multi-camera visual surveillance systems based on single-view metrology," in *Iberian Conference on Pattern Recognition and Image Analysis*. Springer, 2015, pp. 552–559.

[13] S. Barra, M. De Marsico, M. Nappi, and D. Riccio, "Complex numbers as a compact way to represent scores and their reliability in recognition by multi-biometric fusion," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 28, no. 07, p. 1460003, 2014.

[14] R. J. Franklin, Mohana, and V. Dabbagol, "Anomaly detection in videos for video surveillance applications using neural networks," in *2020 Fourth International Conference on Inventive Systems and Control (ICISC)*, 2020, pp. 632–637.

[15] J. Edinger, A. Hofmann, A. Wachner, C. Becker, V. Raychoudhury, and C. Krupitzer, "Wheelshare: Crowd-sensed surface classification for accessible routing," in *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 2019, pp. 584–589.

[16] C. S. Regazzoni, G. Fabri, and G. Vernazza, *Advanced video-based surveillance systems*. Springer Science & Business Media, 1998, vol. 488.

[17] R. T. Collins, A. J. Lipton, and T. Kanade, "Introduction to the special section on video surveillance," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 8, pp. 745–746, 2000.

[18] S. Kamijo, Y. Matsushita, K. Ikeuchi, and M. Sakauchi, "Traffic monitoring and accident detection at intersections," *IEEE transactions on Intelligent transportation systems*, vol. 1, no. 2, pp. 108–118, 2000.

[19] J.-E. Bibault, M. Bassenne, H. Ren, and L. Xing, "Deep learning prediction of cancer prevalence from satellite imagery," *Cancers*, vol. 12, no. 12, p. 3844, 2020.

[20] Z. Swiderska-Chadaj, J. Gallego, L. Gonzalez-Lopez, and G. Bueno, "Detection of ki67 hot-spots of invasive breast cancer based on convolutional neural networks applied to mutual information of h&e and ki67 whole slide images," *Applied Sciences*, vol. 10, no. 21, p. 7761, 2020.

[21] S. Barra, S. M. Carta, A. Corriga, A. S. Podda, and D. Reforgiato Recupero, "Deep learning and time series-to-image encoding for financial forecasting," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. JAS-2019-0392, p. 683, 2020.

[22] S. Carta, A. Corriga, A. Ferreira, A. S. Podda, and D. R. Recupero, "A multi-layer and multi-ensemble stock trader using deep learning and deep reinforcement learning," *Applied Intelligence*, pp. 1–17, 2020.

[23] M. Ismail, M. F. Shaaban, M. Naidu, and E. Serpedin, "Deep learning detection of electricity theft cyber-attacks in renewable distributed generation," *IEEE Transactions on Smart Grid*, 2020.

[24] A. B. Parsa, A. Movahedi, H. Taghipour, S. Derrible, and A. K. Mohammadian, "Toward safer highways, application of xgboost and shap for real-time accident detection and feature analysis," *Accident Analysis & Prevention*, vol. 136, p. 105405, 2020.

[25] A. Helali, H. Ameur, J. Górriz, J. Ramírez, and H. Maaref, "Hardware implementation of real-time pedestrian detection system," *Neural Computing and Applications*, pp. 1–13, 2020.

[26] T.-H. Do and M. Yoo, "Visible light communication-based vehicle-to-vehicle tracking using cmos camera," *IEEE Access*, vol. 7, pp. 7218–7227, 2019.

[27] H. Samia and D. Abdeslem, "A review of artificial intelligence techniques used for urban automatic incident detection systems," in *Proceedings of the 2020 9th International Conference on Software and Computer Applications*, 2020, pp. 281–286.

[28] V. Davydov and S. Bezzateev, "Accident detection in internet of vehicles using blockchain technology," in *2020 International Conference on Information Networking (ICOIN)*. IEEE, 2020, pp. 766–771.

[29] S. Barra, A. Castiglione, F. Narducci, M. De Marsico, and M. Nappi, "Biometric data on the edge for secure, smart and user tailored access to cloud services," *Future Generation Computer Systems*, vol. 101, pp. 534–541, 2019.

[30] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[31] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9592–9600.

[32] Y. Yao, X. Wang, M. Xu, Z. Pu, E. Atkins, and D. Crandall, "When, where, and what? a new dataset for anomaly detection in driving videos," *arXiv preprint arXiv:2004.03044*, 2020.

[33] M. Fabbri, F. Lanzi, S. Calderara, A. Palazzi, R. Vezzani, and R. Cucchiara, "Learning to detect and track visible and occluded body joints in a virtual world," in *European Conference on Computer Vision (ECCV)*, 2018.

[34] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., vol. 28. Curran Associates, Inc., 2015, pp. 91–99. [Online]. Available: https://proceedings.neurips.cc/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf

[35] H. Kashiani, A. A. H. Imani, S. B. Shokouhi, and A. Ayatollahi, "Online visual tracking with one-shot context-aware domain adaptation," *arXiv preprint arXiv:2008.09891*, 2020.