OTE: Optimal Trustworthy EdgeAI solutions for smart cities

Vasileios Mygdalis[1], Lorenzo Carnevale[2,6], Jose Ramiro Martínez-de-Dios[3], Dmitriy Shutin[4], Giovanni Aiello[5], Massimo Villari[2], Ioannis Pitas[1]

[1]*Department of Informatics*, *Aristotle University of Thessaloniki*, Thessaloniki, Greece, {mygdalisv},{pitas}@csd.auth.gr

[2]*Department of Mathematical and Computer Science, Physics and Hearth Sciences*, *University of Messina*, Messina, Italy, {lcarnevale},{mvillari}@unime.it

[3]*Robotics, Vision and Control Group*, *University of Seville*, Seville, Spain, {jdedios}@us.es

[4]*Institute of Communications and Navigation*, *German Aerospace Center (DLR)*, Wessling, Germany, {Dmitriy.Shutin}@dlr.de

[5]*Research and Development Lab.*, *Engineering Ingegneria Informatica S.p.A*, Rome, Italy, {giovanni.aiello}@eng.it

[6]*Gruppo Nazionale per il Calcolo Scientifico (GNCS)*, *Istituto Nazionale di Alta Matematica (INdAM) "F. Severi"*, Rome, Italy, lcarnevale@unime.it

***Abstract*—This work studies and defines the problem of providing extensive and opportunistic Edge AI-based area coverage in smart city application scenarios, by researching and determining the optimal configuration of sensing and computational resources for minimizing the environmental/technology footprint of the solution. A typical smart city computing continuum consists of statically installed multimodal sensing Internet-of-Things (IoT) nodes at various city locations, accompanied by interconnected computational Cloud/Edge/IoT nodes. This paper presents Optimal Trustworthy EdgeAI (OTE), an entirely novel research pipeline, that complements existing smart city infrastructure with intelligent drone Edge/IoT nodes (in the form of modularly equipped unmanned aerial vehicles), capable of autonomous repositioning according to individual/collective sensing and coverage criteria. Thereby, we envisage the emerging cutting-edge technologies of trustworthy sensing, perceiving, modelling technologies for predicting the behavior of moving targets (e.g., citizens/vehicles/objects), understanding natural phenomena (e.g., sea wave motion, urban flora/fauna, biodiversity) in order to anticipate events (people's bad habits, environmental changes), by exploiting novel continuous data processing services across the whole span of the enhanced Cloud-Edge-IoT computing continuum.**

***Index Terms*—EdgeAI, Trustworthy-AI, Smart city, Cloud-Edge-IoT intelligence, UAVs**

I. Introduction

Optimal Trustworthy EdgeAI (OTE), is a research pipeline aiming to lay down the groundwork for research in technologies for complementing existing static Cloud/Edge/IoT infrastructures in smart city environments, for providing a) increased intelligence in the data acquisition phase, b) enhanced coverage, perception, cognition, and understanding of the dynamically changing city environment, and c) increased data acquisition and data processing efficiency. The research challenges are organised in three different elements.

First, it focuses on obtaining increased *IoT intelligence*. We assume that there are some existing statically installed IoT nodes at the Edge, mainly used as sensors and have some (but limited) computational capacity, that are utilized for semantic analysis by developing novel TinyML AI techniques. In addition, the Edge architecture includes advanced drone IoT nodes in the form of Unmanned Aerial Vehicles (UAV), having cognitive abilities for understanding and autonomously operating in the smart city environment, featuring automatic repositioning and perching in locations opportunistically, in order to save energy and provide enhanced coverage in areas not covered by the statically deployed sensor nodes.

Second, IoT intelligence is leveraged towards developing *Trustworthy AI* functionalities, ranging from collective semantic visual analysis and physics-informed machine learning processes, that can be used to analyse the inputs/outputs of all the available sensors. Smart sensor fusion technologies are studied in a two-fold purpose: a) to provide a robust understanding and modelling of the urban environment, and b) to optimally derive and propose the optimal drone IoT sensor locations, for enhanced and efficient area coverage. These technologies combined will lead to significant computational/memory reduction and huge energy savings.

Finally, these Trustworthy AI functionalities are orchestrated in a centralized fashion by increased *Cloud intelligence*, consisting of innovative data streaming and interoperability services at the Cloud layer. Artificial Intelligence will be the common denominator to harmonize the resource provisioning and services deployment, and distribute intelligence across the Cloud-Edge-IoT continuum. This methodology will lead to build fully adaptable and resilient intelligent ecosystems.

OTE envisions a complete Cloud/Edge/IoT system, summarized in Figure 1, that provides rich semantic analysis of static and moving targets and flows of items of interest in urban environments. The semantic analysis outputs can be opened up to the general public for supporting new innovative city-wide applications beyond OTE scope, for offering services to a) the municipality and b) to the citizens. The components and research challenges of the proposed pipeline are analysed in Sections II-IV. Conclusions are drawn in Section V.

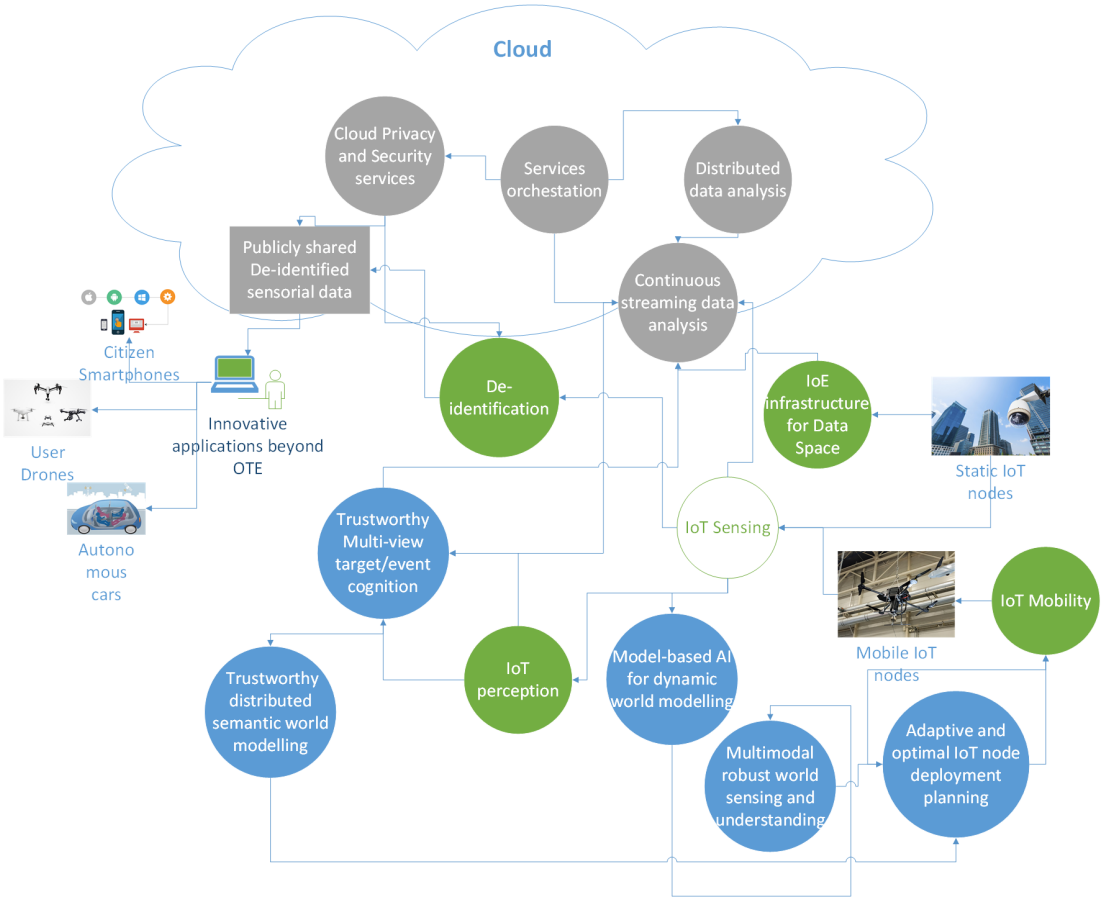

Fig. 1. OTE conceptual block diagram. Interplay between IoT intelligence (Green), Trustworthy Edge AI technologies (Blue) and Cloud Intelligence (Grey)

II. IoT INTELLIGENCE

A. Privacy-by-design EdgeAI

Smart city environments require the installation and operation of IoT nodes in public environments, thus IoT sensors unavoidably capture sensitive information that may be used to infringe an individual's privacy. Such sensitive data information might include human facial images or other biometric identifiers, including clothes, tattoos etc, or even people on vehicles wearing helmets, or the vehicle plates. Assuming that we have the outputs of human detection algorithms, the privacy preservation task requires the generation of gender-neutral image samples that will be used to replace the detected ones. This problem can be viewed as determining an iterative transformation $\mathbf{f}(\cdot) : \mathcal{X} \mapsto \mathcal{X}$, such that the images in the resulting (same) space are no longer identifiable, according to k-anonymity principles [1]. De-identification technologies in existing products (see Google Maps, media industry) apply face/plate detection and aggressively corrupt the images using blurring. In general, privacy and gender neutrality are not guaranteed. Generative Adversarial Networks (GAN)-based methods which employ style transfer [2] are the state-of-the-art approach for privacy protection, but have only been evaluated in experimental datasets, in limited viewing angles. Gender/Race/Body related [3], [4], de-identification have been considered by individual efforts. Data utility has only been evaluated qualitatively.

The main research challenge is to design novel k-anonymity-based deep Neural Network (DNN) approaches that will be trained by incorporating novel-joint optimization criteria for de-identification performance, gender neutrality and data utility. The technology must provide continuous on-the-fly video de-identification in real-world city-captured data and different viewing angles. Such methodology could be based on embedded generative DNN approaches optimized for creating aesthetically pleasing and utilizable results for the remainder analysis, carried out by the rest of the system (e.g., human detection will still work). Focus should be given on real-time operation. To this end, research efforts may consider employing only the outputs of a human detection algorithm (in the form of a human localized in the 2D spatial domain) as input, and will reduce the resolution until reaching some acceptable execution time levels. If that fails, alterations in the employed DNN architecture, DNN compression, should be considered as well.

B. IoT Node sensing for seamless and safe operation

We envision autonomous IoT nodes that will seamlessly and safely operate in urban environments, by combining

embedded knowledge about their assigned use-case task of interest (e.g., traffic monitoring) with the ability to sense and understand potential mobility constraints prior to/during (re-)deployment (e.g., flight is not permitted above humans). Such tasks require the localization of moving targets (e.g., human crowd, humans, cars) that might freely roam around the environment, thus appearing from different viewpoints, in various scales, different lighting conditions or perhaps even occluding each other. Additional challenges are introduced when considering the mobilization (e.g., flying) phase, related due to camera vibrations and and/or the parallax problem. The design of such techniques must also take into consideration the computational and memory constraints imposed by the embedded IoT computational units. Taking into account the above challenges, the sensing modules should be based on DNNs, similar to those applied on human crowd detection [5] and top-down person detection [6], safe-landing spot detection [7] and embedded DNN implementation. The research focus should be to further decrease execution times, by using e.g., DNN compression or knowledge-transfer methods to transfer knowledge from deep complex architectures to lighter ones [8].

C. IoE infrastructure for data space

Distributing IoT nodes all over the cities causes a daily production and storage of huge amounts of data coming from different stakeholders. Data format and shape is therefore different because based on non-standard protocols and data models, thus preventing knowledge exchange with external users, projects and systems. Interoperability and integration of data from various sources is a key challenge that has to be tackled to extract insight and give an understandable value with the purpose to build new and precious services. For this reason, it is necessary to have adequate technologies that allow breaking the silos where data stay, enabling the possibility to access, fuse, and consume cross-domain and multi-scale data, models, and observations independently from data sources, enabling data analysis and tools for data-driven decision making.

In the next few years, advances in technology should produce mechanisms specifically developed to provide cost-effective mechanisms for data management and AI also over the Edge in line with the distributed [9] and federated architecture paradigms [10]. Additional capabilities must be implemented to facilitate the discovery and evaluation of Edge AI data and any other useful data sources available online (as example, Earth Observation repositories, apps for citizen science initiatives), as well as their federation and easy "plug-in". Innovative modelling tools based on "no coding" approach will assist and make semi-automated the activities required to fuse, transform, refine, and harmonise raw and unstructured data in the standard data models, in order to be used by any data consumer. New data models for biodiversity information must be defined to contribute to the reference standardisation programs, such as FIWARE [11].

D. Drone IoT node mobility

OTE pipeline envisions a fleet of intelligent aerial robots that will continuously and non-invasively survey the city autonomously building accurate maps, detecting, localizing, and tracking target objects/events of interest [12]. In order to operate in a city environment, each aerial robot must be endowed with all required onboard sensors, computational resources, and functionalities for fully autonomous operation including take-off, perch on safe locations to save energy e.g., on traffic lights, safe navigation, trajectory planning, flight control, and advanced intelligent perception methods. In order to achieve safe and autonomous navigation of aerial robots, the platforms should include propeller protections and energy-absorption components. Also, they must be endowed with flight safety supervision that will detect mechanical, sensor, or software malfunctioning, to timely take safety actions. Aerial robots should also endowed with fully autonomous navigation capabilities including mission planning, trajectory planning, obstacle detection and avoidance, and autonomous takeoff/perching. Although all of these challenges have been studied individually, implementing everything in the same platform remains a challenge.

The second challenge is how to organise the fleet or aerial robots as a team. Centralized optimal zone partition robot fleet planning methods should be used to coordinate the aerial robots deployed in the site, each assigned with an area depending on the site geometrical or monitoring requirements. Multi-robot trajectory planning methods must be used to define optimal aerial robot trajectories. Obstacle detection and avoidance methods will monitor the pose and velocity of each robot to detect risks and timely change trajectories or command safety actions [13]. Perception-aware aerial robot planning methods must be developed to optimize the amount of information gathered during the flight for (a) target tracking and for (b) map building. During target tracking missions, the trajectories and planning of the aerial robots of the fleet should be performed using the real-time greedy optimization of a utility function that considers the target tracking uncertainty and also the energy consumption required by the aerial robots. During mapping missions, the aerial robot trajectories must be optimized using entirely novel cost-reward utility functions that trade between the time in building an accurate map and the energy consumed by the drone fleet. Finally, the robots must be endowed with autonomous takeoff/perching functionalities based on visual servoing.

III. Trustworthy Edge AI for world modelling

A. Distributed static 3D world modelling

Smart city missions require accurate real-time updated large-scale georeferenced maps with geometrical and photo-realistic content. Cities are complex, unstructured and highly dynamic environments with poor Global navigation satellite system (GNSS) reception which poses significant challenges for autonomous real-time mapping using aerial robots. Dynamic objects pose particularly challenging issues in mapping [14]. Differentiating between static and dynamic objects

for mapping requires providing semantic information to the objects in the map. Besides, since several drones will be used in smart cities, multi-drone mapping techniques are necessary, which also will allow faster mapping of large-scale environments and will enable sharing the maps between different drones to improve mapping accuracy.

The research challenge is to develop multi-drone mapping functionalities with advanced AI-based object recognition that will provide semantic content. The local mapping method on each IoT drone node should be based on a multisensor Simultaneous Localization and Mapping (SLAM) scheme capable of fusing robot 6-DoF location (e.g., computed by GNSS & inertial measurement unit (IMU)), 3D light detection and ranging (LIDAR) scans, and image features (e.g. SIFT, SURF, ORB) [15]. Multisensor optimal data registering methods may be used to generate minimum-drift georeferenced maps with geometrical and pixel information. Multisensor SLAM methods are naturally robust to lack of features, and may integrate GNSS measurements if available. In case of lack of GNSS, the SLAM must integrate LIDAR and cameras providing maps of sufficient accuracy. Next, the semantic information provided by the AI perception methods must be added to the map. Finally, local map sharing and merging techniques must be developed in order to enable creating an unique global map resulting from the entire drone fleet. The IoT nodes can be also localized and mapped by the drones using radio range information, see e.g., [16]. The identifiers and locations of these IoT nodes may be used as reference to improve the accuracy of local map merging.

B. Semantic video instance segmentation

The dynamic/static node operational environment will have to be on-the-fly-analyzed from a semantic standpoint, using dynamically acquired camera input and AI methods for computer vision. The research challenge includes the development of corresponding DNN modules that will perform video instance segmentation, in real-time. The state-of-the-art approach is to employ Transformer [17] or CNN neural architectures [18] and to exploit adversarial learning strategies to compensate for occlusions or distortions [19] and/or employ novel training goals that augment regular supervised training with unsupervised [20] or adversarial objectives [5], in order to increase accuracy in potential use-cases. Another challenge is to accelerate those algorithms for on-board execution without sacrificing accuracy, e.g., by combining multitask training on auxiliary tasks (such as scene geometry extraction by unsupervised depth map estimation).

C. Dynamic Multi-view Target detection/recognition

A global 3D map can be derived after fusing/merging information from the instance segmentation performed by each camera-equipped IoT node. It should contain the location of each tracked target in a common 3D coordinate system at all times. This can be exploited in repeated post-hoc steps of fine-tuning the instance segmentation models in the IoT nodes, using on-line continual learning [21], during mission execution. For example, the 3D location of each target can be prospectively projected using the camera parameters of each IoT node, while the spatial difference of the resulting target 2D location (in pixel coordinates) from the last instance segmentation map prediction can be used to form a loss value, to be back-propagated through the corresponding instance segmentation DNN. The end-result will be a gradual building of increased DNN robustness at each operating drone IoT node, through indirectly exploiting the collective fleet intelligence.

D. Robustness in edge node perception

One definition of robustness refers to the ability of a system to withstand or overcome input or parameter perturbation (hardware malfunction, data acquisition/transmission noise, adversarial attacks etc.). Assuming a system $\mathbf{y} = \mathbf{f}(\mathbf{x}; \boldsymbol{\theta})$ (model $\mathbf{f}$ with inputs $\mathbf{x}$, parameters $\boldsymbol{\theta}$ and outputs $\mathbf{y}$), robustness is quantified by determining its tolerance to perturbation $\|\mathbf{p}\| < \epsilon$ per se, i.e., $\mathbf{f}(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{f}(\mathbf{x} + \mathbf{p}; \boldsymbol{\theta})$ or $\mathbf{f}(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{f}(\mathbf{x}; \boldsymbol{\theta} + \mathbf{p})$. Recently, particular interest has been paid to the problem of adversarial robustness, that involves studying and addressing the inherent model weaknesses that allow adversaries to easily fool a neural network classifier by carefully crafting input perturbations, the so-called adversarial attacks.

In urban environments, there are also adversarial threats present in the physical world [22] (e.g., stickers, people hats/masks, dirty road signs etc.). The root causes of DNN model weaknesses that make them vulnerable to adversarial threats, have not been properly identified yet, however, they are somehow related with the unified deep learning optimization procedure that involves feature learning and classifier learning at the same time. Robustness to adversarial attacks in visual classification problems can be achieved both by detecting them by exploiting one-class classification models [1], or by robustifying the DNN learning process by incorporating geometrically-inspired optimization criteria in the training phase [23]. Nevertheless, the problem remains far to be solved, especially in relevant visual perception tasks (e.g., object detection, semantic segmentation), which lie in the core of OTE pipeline. In the near future, the problem variants in visual perception tasks must be addressed in order to devise entirely novel neural network training strategies, robust to adversarial threats.

E. Explainable event detection/recognition

In environmental monitoring scenarios there is a need to identify changes/events related to e.g., garbage disposal in a previously clean environment. The challenge of this task is that both the environment and the event are very difficult to model/predict, i.e., no easily identifiable features can be extracted in some cases (e.g., at the sea), or the exact opposite may happen in other cases (places at the peripheral districts of the city parks). Assuming that a distribution about the relevant environment has been captured using an appropriate neural architecture e.g., neural autoencoders based on 3D convolutional, Long short-term memory (LSTM) [24] and/or

Transformer-based architectures [17], events can be identified in the cases where the perceived distribution changes dramatically.

For the modelling part, two different classes of algorithmic approaches may be employed, i.e., (a) the unsupervised modelling case and (b), the supervised modelling case. In the former case, it is assumed that the instances of "normal" sensed data can be used to reconstruct a significantly longer temporal sequence (e.g., video), while an event can be detected when the sensed data no longer reconstructs the previous sequence with the same quality. This problem is typically known as the Out-of-distribution detection [25]. For the latter case, one-class classification methods [26], [27] may be employed to model the normal distribution of the sensed data. Events can thereby be identified by the output of the novelty detector (i.e., the one-class classifier). One of the most important challenges is that the analysis of the temporal axis introduces significant computational/memory burdens to the modelling problem. This can be addressed by many different approaches e.g., clipping similar sensing inputs, designing tailor-made lightweight modelling architectures and optimization options, as well as by offloading computations to stronger computational grid units (e.g., to the Cloud).

F. Model-based AI: modeling spatial dynamics

In many applications the processes of interest, such as city traffic density, or concentration of air pollutants over a residential area, vary both in time and space. A brute-force approach to capture these variations typically implies a high-resolution spatial and temporal sampling of such processes. Unfortunately, such an approach turns out to be quite impractical: while in some cases temporal sampling can be realized thanks to modern high-speed processing and acquisition, spatial sampling can be very costly. For instance, air quality monitoring in the city relies on sparsely distributed measurement stations equipped with chemical sensors that provide only in-situ measurements. Moreover, non-visual sensors, and in particular chemical sensors, such as those used for gas or particle measurements, are often characterized by a low information acquisition rate. Specifically, measurements with a frequency in fractions of Hz per sampling point are not uncommon. This is a challenge for a robotic system relying on such sensors for decision making and inference.

The solution investigated in OTE pipeline foresees augmentation of the available fixed sensing networks with intelligent drone-based IoT sensor carrying platforms. These can be dispatched to optimal measurement locations and augment fixed sensor networks. For a successful and efficient deployment of such drone IoT sensors for these purposes, deficiencies of sensors must be compensated by prior information in terms of process models describing the dynamics of the phenomenon of interest and related inverse modeling approaches. The choice of the model is essential for this purpose.

Two approaches can be typically used for this purpose. On the one hand, it is (a) data-driven approaches [28], where neural networks or other non-parametric techniques are used. These methods are quite versatile and powerful provided sufficient training data is available, which is not always the case. On the other hand, (b) physics-based models [29] can be used. For instance, for modeling a distribution of chemicals/particles in the air, a model based on convection-diffusion partial differential equation can be used. To be precise, the physical model for the gaseous material propagation in some region Ω of interest within the time frame $(0, T)$ can be represented with a time-dependent convection-diffusion equation in 3D as

$$\partial_t u - \epsilon \Delta u + (\beta_w + \beta_r) \cdot \nabla u = q\delta_s, \tag{1}$$

supplemented by appropriate boundary and initial conditions. Here, $\partial_t u$ denotes a partial time-derivative of a concentration $u \equiv u(x,t)$ of some material of interest, ϵ is diffusivity of the material, Δ is a Laplace operator, β_w and β_r are wind velocity fields due to wind and, e.g., a drone taking the measurements, respectively, and $q\delta_s$ is a function that models a spatial distribution of material sources. Although (1) has relatively few parameters and can describe the physics of the process quite accurately, estimation of these parameters from measurements is a numerically complex inverse problem. Therefore, researchers should combine data-driven modeling approaches and physics-based models in a unified framework known as model-based machine learning. Such a combination either incorporates data trainable elements into the equations of the spatial dynamics of the process, or trains the network subject to constraints imposed by the differential equation (which are also known as Physics-Informed Neural Networks [30]). For instance, to represent the influence of the drone on the local wind conditions we will rely on data-driven models instead of employing physical models, such as the compressible Navier-Stokes equations. The goal is to obtain numerical representations that are accurate enough, yet sufficiently tractable to be estimated and learned in real-time using computing resources on the drone. The latter is an essential element for deriving optimal sampling strategies for the robots. This can find application in the cases of air quality monitoring in a smart city environment.

G. Multimodal robust world sensing and perception

Given models of the process dynamics and sensor data, the next step is to apply algorithms to cooperatively learn model parameters and thus "understand" the world as "seen" by the robots. The intention is not to collect the measurement data at a central location, but instead use methods of distributed signal processing and estimation. In this way "locally" collected measurements or computations can propagate through the whole network without the need of a central server. The advantage of such an approach is the absence of single point of failure of the system. This failure does not need to be physical breakdown of a component. Network connectivity to a central computer can disrupt the system functionality in centralized architectures. Distributed systems are robust to such disruptions.

The two basic concepts of distributed signal processing that should be investigated are consensus-based [31] strategies and

diffusion-based strategies [32]. Consensus strategies are simple distributed protocols that allow achieving agreement between multiple entities in a network on some quantity of interest. It has been studied extensively in the literature and there are multiple efficient algorithms that implement, e.g., parameter estimation using consensus. While consensus is useful for processing batch data, diffusion strategies are handy for online or streaming data, when constant adaptation is needed. Both strategies may be used and compared in the project for solving inference and learning problems in a decentralized manner.

H. Adaptive and optimal IoT node deployment planning

Adaptive and optimal deployment of IoT and drones aims at finding sampling locations that (a) accelerate and improve the accuracy of learning under realistic time constraints, and (b) using as few measurements as possible. The latter is particularly important when a limited number of IoT or drones is available for measurements at a particular time. The most recent approach follows a so-called reactive strategy that responds in real time to measurements and adapts the decisions accordingly. One approach to design such an exploration strategy based on optimal sensor placement problem or optimal experimental design.

The key assumption underlying these approaches lies in the fact that due to measurement errors, the estimated physical parameters of the model, e.g., spatial concentrations of the material determined as solutions to the inverse problems, can only be approximations. In order to improve the quality of those estimates, an exploration strategy can be defined as a solution to a sequential optimal design problem with two-level structure, where the size of the confidence regions is minimized with the passing of time around the estimated parameters by optimizing the measurement or deployment positions. In this regard, the simpler, yet still realistic, situation is that the measurements are stationary and the process change very little during the measuring phase. As a consequence, determining for each measuring phase only requires a set of measuring positions in space (but not in time). This can be done in a sequential manner. Specifically, alternate measurements from mobile sensors and other fixed sensor networks can be employed, with online fusion of the measurements and solution to the inverse problem for estimating the parameters of interest and identifying the measuring points for the next measuring phase by online solving an optimal design problem. Hence, the optimal design problem in each step represents the upper-level problem compared to the inverse problem, which is the lower-level problem.

IV. Edge/Cloud systems Intelligence

A. Services organization, orchestration and provisioning

Production systems of IoT nodes will be often reconfigured in the future as part of the engineering processes. This aspect needs to achieve adequate orchestration and security levels in an automated way, reducing the current static procedures and manual efforts [33]. Indeed, even though several automated deployment of applications have been developed, the management of deployed applications in a multi-cloud and/or IoT-Edge environment is only partially covered by existing approaches [34]. For example, MEDAL [35] is an intelligent solution that facilitates building and managing data workflows on top of existing flexible and composable data services, seamlessly exploiting and federating IaaS/PaaS/SaaS resources across different Cloud and Edge environments. Another solution in literature is [36] establishing a dynamic network virtualization technique enabled Service Function Chain (SFC) orchestration framework. It operates maximizing the total utility and decomposing it into two sub-problems, i.e., SFC selection and dynamic SFC orchestration.

In OTE, we aim to harmonize the resource provisioning and services deployment over both Cloud and Edge, proposing a methodology based on cost functions driven by AI-models [37]–[39]. Microservice abstractions enable the support of a virtual environment that can be adapted on the basis of the available hardware equipment, where each microservice is autonomous from a development and deployment standpoint. Scaling and managing these types of systems, given the resource heterogeneity and the privacy and security constraints, is complex, so a novel orchestrator is necessary to leverage such dynamic and heterogeneous computing infrastructures. Therefore, one should extend the traditional "cloud-only" notion of run-time control and reconfiguration to resources that are deployed and available at the Cloud-Edge-IoT continuum. This requires the utilization of Machine Learning techniques for developing predictive models to forecast workload inputs and performance metrics across multiple, co-located micro-service on Cloud-Edge-IoT resources, in order to understand the nature of their composition and decide which micro-service can coexist and can be deployed together.

B. Distributed data analysis in the Cloud-Edge-IoT continuum

In a data siloed world, most of the infrastructures, services and applications adopted a self-centered design criterion addressing only their specific challenges and needs deferring any issues related to the interaction with other services. Consequently, the data silos can operate only within a predefined set of protocols, technologies, and data models. This lack of interoperability gets the solutions in a closed (or semi closed) ecosystem and hampers the effective interaction with new services and solutions that could tackle innovative operational purposes. In the continuum, the above-mentioned issues are furtherly stressed and made more complex to address by the dynamic and hybrid contexts that the continuum represents, where the focus is to make inter-operable entities (i.e., data producer nodes) forming Distributed Data Ecosystems (DDE). In this sense, we can identify local data interoperability issues at data producer node level and global data interoperability issues that require the gathering and the processing of data from different heterogeneous data producer nodes (i.e., different cloud nodes of a federation, edge nodes). Therefore, nowadays there is more and more the need to build solutions that "enable" the data interoperability among distributed heterogeneous data provider nodes (both edge and cloud nodes) in

the continuum exchanging harmonized and normalized data in a secure and fair manner for the common good, using standard data models, tools for data quality and data integration incorporating augmented data catalogues in multidomain trusted data spaces. In this regards, Reinforcement Learning (RL) is often used as main methodology. Indeed, it was used to build a distributed-learning-based vehicle routing decision algorithm to adaptively adjust vehicle routing online [40], as well as to make a distributed Multi-Armed Bandits (MAB) model for developing a dynamic network topology change [41].

In OTE, the DDE approach allows manipulating and processing the data (also applying AI) in several "mid-points" located between the data producer and the cloud. Some of these mid-points can be the edge, the fog or, still, the cloud. The DDE architecture allows exchanging harmonized and normalized data in the continuum in a secure and fair manner according to interoperable data structures compliant with common and standardised data models.

C. Security and privacy concerns over services and data

Nowadays, with the exponential growth of connected devices over the Internet, security is one of the major concerns for network communications between heterogeneous parties (i.e., people, devices, etc.). Security is not just about protecting the confidentiality of messages exchanged between parties, but it also involves integrity, and availability. Most secure communications rely, among others, on a centralized trusted server (or a group of servers) architecture. This carries an important security bottleneck: the server delegated to provide services represents a single point of failure (SPoF), being exposed to well-known Distributed Denial of Service (DDOS) attacks. To mitigate such a risk, multiple strategies have been proposed, but these cause an increase in costs for maintaining a security infrastructure. Central server architecture is limited in terms of security because it is susceptible to Man-in-the-middle (MITM) attacks. This type of attack can take place in different forms:

- as a malevolent user who takes the control over the communication channel between legitimate parties sending altered messages to them;
- as a service malfunction refusing to deliver data to one of the parties, causing the incomplete exchange of messages which can drastically alter the meaning of the communication.

Moreover, this configuration is a SPoF because it is sufficient to attack the centralized server (or the cluster of servers) through a DDOS attack, making the entire service unusable. Although this is a well-known attack and it has origin at the beginning of the IT era, it is still one of the most used and crucial types of attack. DDOS consists in obtaining the disruption of services by attempting to limit access to a machine, making a network incapable of providing normal service by targeting either the network bandwidth or its connectivity. These attacks achieve their goal by sending to the victim host a stream of packets that saturate his network or processing capacity, denying access to his consumers. Nowadays, using simple and lightweight tools, malevolent users can use malicious code into unaware victims to leverage a huge number of machines ready to run a distributed version of such an attack. These attacks consume some critical resource at the target and deny the service to legitimate clients as the attack volume can be larger than what the system can handle. There are multiple strategies to prevent such attacks, but these have an immediate correlation with the increase of cost to maintain the secure infrastructure. During a DDOS attack, the malevolent user can gain root permission to the system or database and add or remove data from the system making it difficult to identify what is legitimate and what is altered. DDOS is just an example of possible distributed attacks, but the same principle applies for almost all of them.

D. Streaming data analysis on Edge/IoT computing resources

Although data stream processing is a paradigm with a long tradition [42], traditional systems like Apache Storm and Flink, which have a wide popularity and support continuous streaming, target homogeneous clusters and clouds and are not designed for the Edge [43]. As a matter of fact, the shift to the Edge/IoT advocates new software engineering techniques to develop efficient streaming runtime systems, which should exhibit a high-degree of reconfigurability of the underlying implementation to leverage different kinds of resource-constrained hardware components in an efficient way and in face of dynamic workload, networking, and energy conditions. Recent attempts [44] enhance traditional systems to fit the constraints of edge resources, by re-implementing parts of their runtime system introducing explicit scheduling of streamed data analysis tasks using custom scheduling policies. However, they represent custom prototypes which require to be maintained together with the standard code base of the traditional systems. OTE aims to identify parallel/concurrent building blocks that can be composed to build complex streaming applications, and whose internal implementation can leverage different kinds of resources transparently to the end user. This idea percolates the consolidated approach of Parallel Patterns and Algorithmic Skeletons [45] at the implementation level of the runtime system design of a framework, where each block describes a recurrent computation or communication pattern, which can be implemented with efficient mechanisms and with special focus on the constraints of embedded devices.

V. Conclusion

This work described the main research and development challenges that will arise in the next few years, towards incorporating trustworthiness in smart city applications. The research pipeline prescribed technical solutions and all the components that need to be integrated along the whole span of Cloud-Edge-IoT computing continuum. Such include the most recent hardware and software breakthrough technologies in edge sensing, combined with novel smart city robotics. We have defined the components of IoT intelligence and how they can be accompanied with trustworthy AI in order to provide rich multimodal and collective intelligence, at a fleet level.

In addition, recent advances in cloud computing will allow the streamlining of the semantic metadata extraction, sensor processing, services organisation, system orchestration, and provisioning, considering security and privacy concerns.

The findings of this work can be exploited for designing entirely novel smart city solutions that tackle a wide range of important applications, notably in traffic monitoring and management, human flow monitoring, traffic flow optimization, and even for addressing environmental challenges such as air quality assessment, pollution monitoring, urban-flora health estimation, and many others. This paper may serve as a guideline for researchers for assessing their current research interests within a general purpose smart-city pipeline, that can also stimulate new ideas for innovators.

ACKNOWLEDGMENT

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 871479 (AERIAL-CORE) and 101004605 (DECIDO). This work was also supported by the "GNCS-INdAM", the Italian Project FISR "La rifunzionalizzazione del Contemporaneo" (CUP J42F16000600001) and the Spanish Project ROBMIND (Ref. PDC2021-121524-I00) from the Programa Estatal de I+D+i. This publication reflects the authors' views only. The European Commission is not responsible for any use that may be made of the information it contains.

REFERENCES

[1] V. Mygdalis, A. Tefas, and I. Pitas, "K-anonymity inspired adversarial attack and multiple one-class classification defense," *Neural Networks*, vol. 124, pp. 296–307, 2020.

[2] J. Lin, Y. Li, and G. Yang, "Fpgan: Face de-identification method with generative adversarial networks for social robots," *Neural Networks*, vol. 133, pp. 132–147, 2021.

[3] K. Brkic, T. Sikiric, Ivan rkac, and Z. Kalafatic, "I know that person: Generative full body and face de-identification of people in images," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1319–1328, IEEE, 2017.

[4] P. Nousi, S. Papadopoulos, A. Tefas, and I. Pitas, "Deep autoencoders for attribute preserving face de-identification," *Signal Processing: Image Communication*, vol. 81, p. 115699, 2020.

[5] C. Papaioannidis, I. Mademlis, and I. Pitas, "Autonomous uav safety by visual human crowd detection using multi-task deep neural networks," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 11074–11080, IEEE, 2021.

[6] C. Symeonidis, I. Mademlis, N. Nikolaidis, and I. Pitas, "Improving neural non-maximum suppression for object detection by exploiting interest-point detectors," in *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, IEEE, 2019.

[7] C. Symeonidis, E. Kakaletsis, I. Mademlis, N. Nikolaidis, A. Tefas, and I. Pitas, "Vision-based uav safe landing exploiting lightweight deep neural networks," in *2021 The 4th International Conference on Image and Graphics Processing*, pp. 13–19, 2021.

[8] N. Passalis and A. Tefas, "Learning deep representations with probabilistic knowledge transfer," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 268–284, 2018.

[9] O. Vermesan, R. John, P. Pype, G. Daalderop, M. Ashwathnarayan, R. Bahr, T. Karlsen, and H.-E. Sand, "Internet of vehicles – system of systems distributed intelligence for mobility applications," in *Internet of Things*, pp. 93–147, Springer International Publishing, 2021.

[10] A. Mourad, H. Tout, O. A. Wahab, H. Otrok, and T. Dbouk, "Ad hoc vehicular fog enabling cooperative low-latency intrusion detection," *IEEE Internet of Things Journal*, vol. 8, no. 2, pp. 829–843, 2021.

[11] L. Carnevale, A. Galletta, M. Fazio, A. Celesti, and M. Villari, "Designing a fiware cloud solution for making your travel smoother: The fliware experience," in *2018 IEEE 4th Int. Conf. on Collaboration and Internet Computing (CIC)*, pp. 392–398, October 2018.

[12] J. P. Rodríguez-Gómez, A. G. Eguíluz, J. R. Martínez-De Dios, and A. Ollero, "Auto-tuned event-based perception scheme for intrusion monitoring with uas," *IEEE Access*, vol. 9, pp. 44840–44854, 2021.

[13] F. J. Fernández-Jiménez and J. R. Martínez-de Dios, "A robot-sensor network security architecture for monitoring applications," *IEEE Internet of Things Journal*, 2021.

[14] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[15] J. L. Paneque, J. Martínez-de Dios, and A. Ollero, "Multi-sensor 6-dof localization for aerial robots in complex gnss-denied environments," in *2019 IEEE/RSJ Int. Conf on Intelligent Robots and Systems (IROS)*, pp. 1978–1984, IEEE, 2019.

[16] F. J. Perez-Grau, J. R. Martinez-de Dios, J. L. Paneque, J. J. Acevedo, A. Torres-González, A. Viguria, J. R. Astorga, and A. Ollero, "Introducing autonomous aerial robots in industrial manufacturing," *Journal of Manufacturing Systems*, vol. 60, pp. 312–324, 2021.

[17] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European conference on computer vision*, pp. 213–229, Springer, 2020.

[18] L. Deng, M. Yang, Y. Qian, C. Wang, and B. Wang, "Cnn based semantic segmentation for urban traffic scenes using fisheye camera," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, pp. 231–236, IEEE, 2017.

[19] C. Papaioannidis, V. Mygdalis, and I. Pitas, "Domain-translated 3d object pose estimation," *IEEE Transactions on Image Processing*, vol. 29, pp. 9279–9291, 2020.

[20] M. Kaseris, I. Mademlis, and I. Pitas, "Adversarial unsupervised video summarization augmented with dictionary loss," in *2021 IEEE Int. Conf. on Image Processing (ICIP)*, pp. 2683–2687, IEEE, 2021.

[21] R. Aljundi, M. Lin, B. Goujaud, and Y. Bengio, "Gradient based sample selection for online continual learning," *Advances in neural information processing systems*, vol. 32, 2019.

[22] A. Kurakin, I. Goodfellow, S. Bengio, *et al.*, "Adversarial examples in the physical world," 2016.

[23] V. Mygdalis and I. Pitas, "Hyperspherical class prototypes for adversarial robustness," *Pattern Recognition*, p. 108527, 2022.

[24] T. Akilan, Q. J. Wu, A. Safaei, J. Huo, and Y. Yang, "A 3d cnn-lstm-based image-to-image foreground segmentation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 959–971, 2019.

[25] J. Ren, P. J. Liu, E. Fertig, J. Snoek, R. Poplin, M. Depristo, J. Dillon, and B. Lakshminarayanan, "Likelihood ratios for out-of-distribution detection," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[26] V. Mygdalis, I. Alexandros, A. Tefas, and I. Pitas, "Large-scale classification by an approximate least squares one-class support vector machine ensemble," in *2015 IEEE Trustcom/BigDataSE/ISPA*, vol. 2, pp. 6–10, IEEE, 2015.

[27] V. Mygdalis, A. Tefas, and I. Pitas, "Exploiting multiplex data relationships in support vector machines," *Pattern Recognition*, vol. 85, pp. 70–77, 2019.

[28] S. Wang, B.-S. Shin, D. Shutin, and A. Dekorsy, "Diffusion field estimation using decentralized kernel kalman filter with parameter learning over hierarchical sensor networks," in *2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, IEEE, 2020.

[29] T. Wiedemann, D. Shutin, and A. J. Lilienthal, "Model-based gas source localization strategy for a cooperative multi-robot system-a probabilistic approach and experimental validation incorporating physical knowledge and model uncertainties," *Robotics and Autonomous Systems*, vol. 118, pp. 66–79, 2019.

[30] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational physics*, vol. 378, pp. 686–707, 2019.

[31] T. Buchgraber and D. Shutin, "Distributed variational sparse bayesian learning for sensor networks," in *2012 IEEE International Workshop on Machine Learning for Signal Processing*, pp. 1–6, IEEE, 2012.

[32] N. L. Pedersen, C. N. Manchón, M.-A. Badiu, D. Shutin, and B. H. Fleury, "Sparse estimation using bayesian hierarchical prior modeling

for real and complex linear models," *Signal processing*, vol. 115, pp. 94–109, 2015.

[33] M. Ehrlich, H. Trsek, M. Gergeleit, J. Paffrath, K. Simkin, and J. Jasperneite, "Secure and flexible deployment of industrial applications inside cloud-based environments," in *2019 IEEE Int. Conf. on Emerging Technologies and Factory Automation (ETFA)*, pp. 1269–1272, 2019.

[34] L. Harzenetter, U. Breitenbücher, F. Leymann, K. Saatkamp, B. Weder, and M. Wurster, "Automated generation of management workflows for applications based on deployment models," in *2019 IEEE 23rd International Enterprise Distributed Object Computing Conference (EDOC)*, pp. 216–225, 2019.

[35] V. Theodorou, I. Gerostathopoulos, I. Alshabani, A. Abelló, and D. Breitgand, "MEDAL: An AI-driven data fabric concept for elastic cloud-to-edge intelligence," in *Advanced Information Networking and Applications*, pp. 561–571, Springer International Publishing, 2021.

[36] H. Chen, S. Wang, G. Li, L. Nie, X. Wang, and Z. Ning, "Distributed orchestration of service function chains for edge intelligence in the industrial internet of things," *IEEE Transactions on Industrial Informatics*, pp. 1–1, 2021.

[37] M. Villari, M. Fazio, S. Dustdar, O. Rana, and R. Ranjan, "Osmotic computing: A new paradigm for edge/cloud integration," *IEEE Cloud Computing*, vol. 3, no. 6, pp. 76–83, 2016.

[38] M. Villari, A. Celesti, and M. Fazio, "Towards osmotic computing: Looking at basic principles and technologies," in *Advances in Intelligent Systems and Computing*, pp. 906–915, Springer International Publishing, July 2017.

[39] M. Villari, A. Galletta, A. Celesti, L. Carnevale, and M. Fazio, "Osmotic computing: Software defined membranes meet private/federated blockchains," in *2018 IEEE Symposium on Computers and Communications (ISCC)*, pp. 01292–01297, 2018.

[40] K. Lin, C. Li, Y. Li, C. Savaglio, and G. Fortino, "Distributed learning for vehicle routing decision in software defined internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3730–3741, 2021.

[41] S. Chen, Y. Tao, D. Yu, F. Li, and B. Gong, "Distributed learning dynamics of multi-armed bandits for edge intelligence," *Journal of Systems Architecture*, vol. 114, p. 101919, 2021.

[42] H. C. M. Andrade, B. Gedik, and D. S. Turaga, *Fundamentals of Stream Processing: Application Design, Systems, and Analytics*. Cambridge University Press, 2014.

[43] M. Dias de Assunção, A. da Silva Veith, and R. Buyya, "Distributed data stream processing and edge computing: A survey on resource elasticity and future directions," *Journal of Network and Computer Applications*, vol. 103, pp. 1–17, 2018.

[44] X. Fu, T. Ghaffar, J. C. Davis, and D. Lee, "EdgeWise: A better stream processing engine for the edge," in *2019 USENIX Annual Technical Conference (USENIX ATC 19)*, (Renton, WA), pp. 929–946, USENIX Association, July 2019.

[45] F. A. Rabhi and S. Gorlatch, eds., *Patterns and Skeletons for Parallel and Distributed Computing*. Springer London, 2003.