

# MARVEL: Multimodal Extreme Scale Data Analytics for Smart Cities Environments

Dragana Bajovic<sup>1</sup>, Arian Bakhtiarnia<sup>2</sup>, George Bravos<sup>3</sup>, Alessio Brutti<sup>4</sup>, Felix Burkhardt<sup>5</sup>, Daniel Cauchi<sup>6</sup>, Antony Chazapis<sup>7</sup>, Claire Cianco<sup>6</sup>, Nicola Dall'Asen<sup>4</sup>, Vlado Delic<sup>1</sup>, Christos Dimou<sup>3</sup>, Djordje Djokic<sup>8</sup>, Antonio Escobar-Molero<sup>9</sup>, Lukas Esterle<sup>2</sup>, Florian Eyben<sup>5</sup>, Elisabetta Farella<sup>4</sup>, Thomas Festi<sup>10</sup>, Artemios Geromitsos<sup>11</sup>, Giannis Giakoumakis<sup>7</sup>, George Hatzivasilis<sup>12</sup>, Stiris Ioannidis<sup>7,13</sup>, Alexandros Iosifidis<sup>2</sup>, Theodora Kallipolitou<sup>14</sup>, Grigorios Kalogiannis<sup>12</sup>, Akrivi Kiouisi<sup>11</sup>, Despina Kopanaki<sup>7</sup>, Manolis Marazakis<sup>7</sup>, Stella Markopoulou<sup>14</sup>, Adrian Muscat<sup>6,15</sup>, Francesco Paissan<sup>4</sup>, Tomás Pariente Lobo<sup>16</sup>, Dusan Pavlovic<sup>8</sup>, Theofanis P. Raptis<sup>17</sup>, Elisa Ricci<sup>4,19</sup>, Borja Saez<sup>9</sup>, Farhan Sahito<sup>8</sup>, Kenneth Scerri<sup>6,15</sup>, Björn Schuller<sup>5,18</sup>, Nikola Simic<sup>1</sup>, George Spanoudakis<sup>12</sup>, Alex Tomasi<sup>10</sup>, Andreas Triantafyllopoulos<sup>5</sup>, Lorenzo Valerio<sup>17</sup>, Javier Villazán<sup>16</sup>, Yiming Wang<sup>4</sup>, André Xuereb<sup>6,15</sup>, and Johan Zammit<sup>6</sup>

<sup>1</sup>Faculty of Technical Sciences, University of Novi Sad, Serbia

<sup>2</sup>ECE, Aarhus University, Denmark

<sup>3</sup>ITML, Greece

<sup>4</sup>Fondazione Bruno Kessler, Italy

<sup>5</sup>audEERING GmbH, Germany

<sup>6</sup>Greenroads Ltd, Malta

<sup>7</sup>FORTH-ICS, Greece

<sup>8</sup>Privanova, France

<sup>9</sup>Infineon Technologies AG, Germany

<sup>10</sup>Municipality of Trento, Italy

<sup>11</sup>INTRASOFT, Greece

<sup>12</sup>Sphynx Technology Solutions AG, Switzerland

<sup>13</sup>Technical University of Crete, Greece

<sup>14</sup>ZELUS, Greece

<sup>15</sup>University of Malta, Malta

<sup>16</sup>Atos Spain SA, Spain

<sup>17</sup>IIT-CNR, Italy

<sup>18</sup>EIHW, University of Augsburg, Germany

<sup>19</sup>University of Trento, Italy

**Abstract**—A Smart City based on data acquisition, handling and intelligent analysis requires efficient design and implementation of the respective AI technologies and the underlying infrastructure for seamlessly analyzing the large amounts of data in real-time. The EU project MARVEL will research solutions that can improve the integration of multiple data sources in a Smart City environment for harnessing the advantages rooted in multimodal perception of the surrounding environment.

## I. INTRODUCTION

The amount of data generated every day in modern cities is vastly increasing and opens up new opportunities in improving various aspects of life and quality services, such as in transportation, safety and security. The underlying infrastructure plays a central role in a smart city environment, i.e., the sensors gathering information from various locations and different modalities (e.g., cameras and microphones) and computing devices distributed around the city and the cloud [1], the intelligent functionalities available for processing large and noisy data, and the data processing framework designed to seamlessly process all available information in real-time. The availability of multiple types of sensors allows for exploiting the benefits of multimodal perception, such as increased performance and robustness in the appearance of occlusions and noise that may occur on single modalities. Such multimodal perception has to be close-to-the-source, fast, and aligned with privacy preservation policies, rules, and laws of the state or country.

This paper presents the new EU project MARVEL<sup>1</sup> focusing on: i) employing a ubiquitous Edge-to-Fog-to-Cloud architecture, where data can be processed in various degrees of closeness to their source, ii) utilizing multi-modal perception for harnessing information from different modalities, and iii) exploiting platforms and technologies for data fusion, sharing,

and privacy. The proposed technological solutions will be showcased in real-life scenarios focusing on safer roads for pedestrians, monitoring of public spaces, and security and prevention. Analysis will be focused on visual and audio modalities for providing close-to-the-source analytics, facilitated by a novel Edge-to-Fog-to-Cloud (E2F2C) processing scheme for achieving (near) real-time operation.

## II. THE MARVEL ARCHITECTURE

The MARVEL conceptual architecture, shown in Figure 1 consists of three layers, the edge, the fog and the cloud, which jointly form a ubiquitous computing environment that enables seamless execution of computing tasks. Data will be collected in the edge layer through cameras, microphones, and other sensors, and also through mobile apps running on smartphones. In the cloud layer, the AI-based learning framework will perform machine learning (ML) and deep learning (DL) model training over audio-visual data. Once the models are trained, they will be distributed across the E2F2C MARVEL architectural elements. To harness the information contained in different datasets, while respecting their privacy, the MARVEL framework will adopt a personalized Federated Learning (FL) approach, where model training will be performed in parallel at different locations, at the respective edge and fog layers, whereas model merging will be achieved at the joint cloud server.

The MARVEL project is to adopt a container-based execution platform, which is spanning from the Centre (HPC/Cloud) to the Edge, including the intermediate layer of Fog. This environment will support i) life-cycle management of containers, at all execution sites in the Centre-to-Edge architectural framework, and ii) orchestration (based on the Kubernetes runtime environment) to support real-time, data-driven application workflows. Effective task-scheduling algorithms will be developed to ensure that tasks are completed on time and the

<sup>1</sup>[www.marvel-project.eu](http://www.marvel-project.eu)

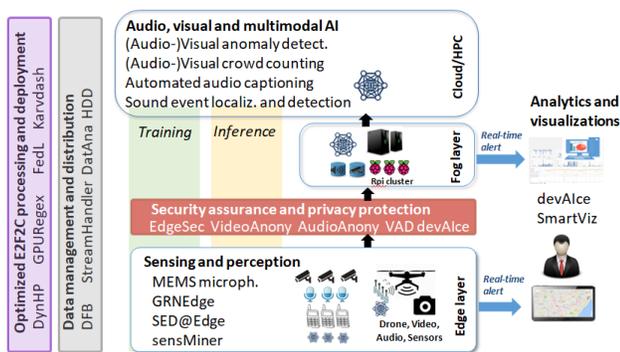


Fig. 1. MARVEL conceptual architecture

number of concurrent tasks at each available node is optimized. Furthermore, reallocation mechanisms will be used to reduce task delays in accordance with the container characteristics. Containers are easy-to-deploy software packages and containerized applications are easily distributable, making them a natural fit for edge computing solutions.

During the execution of the framework, data will be managed by four data management platforms of complementary functionalities, i.e., a platform for fusing data of different modalities, a platform for optimal distribution of data over the infrastructural elements, a platform for handling data streams in real-time, and a platform for managing data flows. The cloud layer will serve as a host for these applications, whereas at run time their different components will be distributed across all E2F2C MARVEL elements. The decision-making module will operate on both data-at-rest, for medium to long-term business decisions, and data in motion, for real-time alarms and real-time visualizations for situational awareness. Finally, GPU acceleration will be used both in the edge and in the fog/cloud for computations and for fast pattern matching.

Security and privacy mechanisms will be enforced at all layers of the architecture. With respect to privacy, MARVEL will adopt a privacy-by-design approach which is achieved in two ways. Privacy will be enforced at the very edge, where anonymization algorithms will be used to remove all personally identifiable information. With respect to security assurance, all dataflows will be end-to-end encrypted. Additionally, secure elements will be employed in an attempt to bulletproof traditionally-vulnerable edge devices.

### III. THE MARVEL E2F2C FRAMEWORK

MARVEL's E2F2C framework will serve as a ubiquitous computing environment where processing occurs across all architectural layers (edge, fog, cloud) and also in parallel across nodes belonging to a single layer. Based on historical data-at-rest, ML/DL models will be first trained in the cloud for the classification of AV events and detection of emerging entities via multimodal representations. Based on the trained models, the E2F2C framework will be infused over time by different AI tasks at various nodes of the MARVEL architecture. These tasks will be both of training and inference type.

To harness the power of datasets held by different entities (e.g., devices), MARVEL adopts a personalized FL approach where the models trained over locally augmented data are merged in the cloud layer, exploiting commonalities between the event occurrences, and thus increasing classification performance while preserving data privacy.

The deployment of the AI tasks (e.g., breaking the DL architecture in split learning) will be realized by a dedicated module, in a semi-automatic manner, guided by various performance metrics collected during the runtime of the framework. The deployments will run in containerized environments ensuring reliable communication between the dependent, consecutive AI tasks (e.g., different parts of a DL network), and reliable and smooth execution. To enable optimal data processing, i.e., "where needed" by the AI tasks, capabilities of MARVEL data management platforms will be exploited to achieve optimal data distribution such that the latency to the target processing point is minimized while keeping the power consumption at a desired level.

GPU stream processing for accelerated analytics will be implemented in both edge and cloud nodes with dynamic assignment of the execution parts and scheduling of the incoming data streams. Edge devices containing low-end GPUs will be responsible for light, preprocessing computations, allowing for partial accelerated analytics as early as possible. The remainder will be offloaded to the cloud pool of GPUs for sophisticated and advanced analytics. As the interface of the MARVEL platform with end-users, MARVEL will deliver a decision-making toolkit with both medium to long-term business analytics and real-time decision making. Real-time alerts will be enabled by edge-based deployments of AI tasks, and visualizations based on attention maps will be used to support human-based decision-making. Multimodal AV summaries will enable efficient forensics of recorded events and creation of appealing marketing materials for city authorities.

### IV. THE MARVEL COMPONENTS AND FUNCTIONALITIES

Within MARVEL, massive amounts of data generated by spatially distributed sensor sources need to be handled, coordinated, stored, and processed to perform appropriate alerting and notification. Several components need to be interlinked in order to achieve this challenging task. MARVEL will also include a number of AI functionalities implemented within the toolkits described in the following.

#### A. Data management and distribution toolkit

**The data acquisition framework DatAna** automates the coordination of streams and data flows between different sensors. It provides predefined processors and allows for extensions of data processors for smart cities that interface with other open data repositories and third party services.

**The StreamHandler platform** scales out and accommodates various sensor sources and Big Data from different domains. It interoperates with all modern data storage technologies as well as other persistence approaches.

**The data fusion bus DFB** fuses multiple modality data

streams. A modality processing pipeline can transform heterogeneous data in real-time by reducing the granularity from the signal level to a semantic level. Autonomous fusion agents are activated during the execution of the modality processing pipeline to perform modality fusion techniques on the data.

**The hierarchical data distribution system HDD** offers a set of distributed adaptive data delivery and access algorithmic schemes for guaranteeing real-time delay requirements while prolonging edge network lifetime [2]. It targets large-scale deployments characterised by an inherent computational intractability on the algorithmic management of the available data.

### B. Sensing and perception toolkit

**Advanced MEMS microphones.** Software for microphone array data acquisition will be developed for converting the data format from the microphones in USB-audio data format and optimal data collection. Data Analytics for early Decision-Support in the acoustic sensor unit will require pre-processing of data using edge ML.

**GRNEdge** [3] is a hardware and software edge setup for capturing audio and visual data in real-time. A waterproof housing metal box will be installed on site containing cameras, microphones, a Raspberry Pi 4 with 4GB RAM running Raspbian, an Arduino Mega256 for IO purposes, a 4G/5G modem, a WiFi router, a charging controller and battery pack, and streaming software.

**sensMiner** is a data collection application to record environmental acoustics and user annotations. It allows effective on site annotations aligned in real-time with the data collection through the user (“embedded annotator”) that records the environmental sounds and in-parallel annotates audio and stores the corresponding segment in the phone memory.

**SED@Edge kit**<sup>2</sup> consists of a low power STM32 PCB boards with embedded MEMS microphones. A tiny neural network, obtained via student-teacher compression of a larger one, is used for sound event classification using an extremely small memory footprint and few milliwatts [4].

### C. Audio, Visual and multimodal analytics

**(Audio-)Visual anomaly detection** aims to establish a representation of normal situation in a scene based on observations collected over time representing normal activity. The use of multi-modal data can increase performance under occlusion, and it enables the detection of anomalous activities that are not present in the visual stream [5].

**(Audio-)Visual crowd counting** aims at counting the total number of people present in an image when audio signal is available. The use of multi-modal data can increase performance, as the ambient audio can complement the visual information in situations where the quality of the image is low, for instance, low resolution, low illumination, severe occlusion, or presence of equipment noise [6], [7].

**Automated audio captioning** aims at producing a textual

descriptions for give audio segments [8]. In MARVEL, it offers the ability to extract high-level knowledge from audio signals such as spatiotemporal relationships of sound events and indication of abstract knowledge.

**Sound event detection** aims at recognizing and temporally locating sound events within a single-channel audio signal [9]. The output of the detection is the start and end timestamps of each sound event in the input audio stream.

**Sound event detection and localization** aims to recognize individual sound events of specific classes, detect their temporal activity and estimate their location during the activity [10]. The input of the system is multi-channel audio, captured with a microphone array, and the system output is the estimated sound event activity along with the spatial location of sound events (elevation and azimuth).

### D. Analytics and visualizations platforms

**devAIce**<sup>3</sup> is a software development kit for audio processing, including voice activity detection [11], feature extraction [12], acoustic scene classification, and can potentially be enriched with audio tagging and audio event detection.

**The advanced visualization toolkit SmartViz** offers a solution that can handle a large number of heterogeneous events and provide intuitive visualizations for IT and non-IT experts. It will offer temporal inspection of various metrics and combinations of relevant measurements in interactive visual representations, so as to facilitate investigation of captured events and help decision makers to gain real-time situational awareness of Smart Cities environments [13].

### E. Optimized E2F2C processing and deployment toolkit

**GPU Pattern matching. GPURex** [14] is a real-time high-speed pattern matching engine that leverages the parallelism properties of GPGPUs to accelerate the process of string and/or regular expression matching and can be deployed at any layer of the MARVEL E2F2C chain. It requires a fixed set of patterns (strings) to be loaded during the start-up phase and then compares them against the real-time data it digests.

**NN compression DynHP** aims at compressing a neural network during the training phase under a fixed memory budget [15]. In MARVEL it will be part of the pipeline enabling efficient DNN training and inference at Edge/Fog levels.

**FedL** is a federated learning [16] (FL) framework exploiting commonalities between event occurrences, and thus increasing classification performance, while keeping the data local. In MARVEL, the framework will be explored with FL clients situated both at edge devices (cross-devices) and fog layers (cross-silo). Personalization refers to adjusting the training losses according to the source of the data updating the FL model.

**Karvdash**<sup>4</sup> [17] (Kubernetes CARV dashboard) is a dashboard service for facilitating data science on Kubernetes. It supplies the landing page for working on a Kubernetes cluster, manages

<sup>3</sup><https://www.audeering.com/what-we-do/devaice/>

<sup>4</sup><https://github.com/CARV-ICS-FORTH/karvdash>

<sup>2</sup><https://gitlab.fbk.eu/gcerutti/urbansoundclassifier>

users, launches notebooks, and wires up relevant storage to the appropriate paths inside running containers. In MARVEL it will serve as the front end to manage data collection services/tools through a simple web-based interface. It will be deployed at the cloud layer and allow deployment and exposure of services at the cloud and fog layers through respective Kubernetes-based configuration and labels.

#### F. Security assurance and privacy protection toolkit

MARVEL will include anonymization-focused functionalities for audio and video modalities described below.

**VideoAnony** is designed to anonymize video content, i.e., removing identifiable information from the input video. This is done by processing a video in a frame per frame manner for detecting persons/faces and performing obfuscation, i.e., removing the identifiable information on the persons/faces [18], [19]. The tool also incorporates an advanced AI-based approach for removing the identifying characteristics of human faces and bodies, while still generating high-quality realistic images, thus not negatively impacting on subsequent processing tasks.

**AudioAnony** is designed to remove information about the speaker identity from an audio stream, for example via voice conversion [20], [21]. This could possibly replace other more invasive privacy preserving strategies (i.e., removing segments with speech content) or approaches that rely on extracting and anonymizing features.

**EdgeSec** is a security framework based on the open-source community version of the SCONE [22] framework which will be deployed in the edge layer and its goals are to encrypt data and network traffic to protect them from unauthorized access, and to attest programs ensuring that only the correct, unmodified binaries are executed inside a genuine SGX enclave.

### V. MARVEL PILOTS

MARVEL will showcase the potential to address civic challenges effectively against real-life experiments in the Smart Cities of Trento (Italy), Novi Sad (Serbia), and the Municipality of Malta. Multiple use-cases will be implemented under the umbrella of three Pilots described in the following.

#### A. Pilot I: Municipality of Trento

For the use cases included in Pilot I, data will be derived from real and staged recordings that include data acquired by the surveillance cameras and microphone arrays mounted near the cameras in the city of Trento.

**Use Case I - Crowd monitoring:** The goal of this use case is to increase safety in areas of particular activity (more crowded than others, particular crowd movements) to prevent dangerous situations, by alerting the operational center of the local police [23].

**Use Case II - Detecting criminal / anti-social behaviours:** The goal of this use case is to detect criminal or anti-social behaviours and trigger an alarm of a custom view. MARVEL aims to detect possible dangerous situations like robberies, aggression, and drug dealing especially during night-time [24].

**Use Case III - Monitoring of parking places:** The goal of this

use case is to conduct (audio-)visual monitoring of a parking place, including analysis of car trajectories, detection of cars out of the parking slots, car damages and obstructions [25].

**Use Case IV - Analysis of a specific area:** The goal of this use case is to monitor places of the city, to support long-term decision making by the Public Administration via data analytics, like number of persons, cars, trajectories and events [26].

#### B. Pilot II: Municipality of Malta

The Malta National Transport Strategy [27] recognises that planning needs to be carried out over an analytical and scientific basis, emphasizes the need to facilitate both walking and cycling and acknowledges the importance of education campaigns. In addition local studies strongly indicate that the lack of safe roads is the main barrier to cycling [28]. The use cases in Pilot II address the implementation of the transport vision through the processing of data sourced from audio and visual sensors mounted at the monitored streets.

**Use-case I - Safer roads:** The goal of this use case is to detect cyclists, pedestrians, e-bikes and possibly other micro-mobility modes exiting in a junction and inform the car and motorised-vehicle drivers of the presence of vulnerable road users.

**Use-case II - Driver Behaviour:** The goal of this use case is to monitor the behaviour of drivers near a junction, such as speeding, dangerous lane changes, etc. It involves the detection of situation useful for both planning and automatically evaluating education campaigns, which can take many forms such as paid adverts on the various media, including social-media, but also the installation of information boards at targeted zones.

**Use-case III - Traffic conditions/anomalous events:** The goal of this use case is to monitor traffic conditions (like traffic jam) to inform the drivers or to infer possible issues ahead in non-monitored areas.

**Use-case IV - Trajectory collection** for long-term data analytics about the traffic behaviour, which can be used for long-term planning.

#### C. Pilot III: City of Novi Sad

**The drone experiment** will evaluate the potential of drones in monitoring of large public events in a city center or in large open space areas outside of a city. Usually to perform surveillance and monitoring of such events the challenge is either the lack of infrastructure or the crowded frontal views. Mobility of a drone equipped with camera arrays, microphones and computing resources can allow closer inspection and verification of possible alarms, which is not possible with the fixed infrastructure already in place.

### VI. LEGAL AND ETHICAL COMPLIANCE

MARVEL is grounded on strong ethical and legal foundations. Tending to ultimately result in deployment and use of trustworthy AI, the project will focus on ensuring all requirements rooted in fundamental rights are respected. Particular attention will be given to preserving personal privacy and ensuring the secure processing of personal data. Pro-privacy

measures will be applied at the very edge using anonymization algorithms that will protect personal data. Also, all dataflows will be end-to-end encrypted and secure elements will be employed in edge devices. Finally, GPU acceleration will be used both in the edge and in the fog for optimized computations and for fast pattern matching. Therefore, application of technical and organizational measures should ensure the application of the privacy-by-design approach.

Considering that MARVEL should improve individual and collective wellbeing, the respect for fundamental rights including protection of personal privacy are also ethical imperatives. Therefore, the project will respect ethical requirements to prevent algorithmic bias and ensure the preservation of diversity, non-discrimination, and principle of fairness.

#### ACKNOWLEDGMENT

This work was funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 957337. 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] Cisco, "Cisco Annual Internet Report (2018–2023) White Paper," <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>, March 2020.
- [2] T. P. Raptis, A. Passarella, and M. Conti, "Distributed data access in industrial edge networks," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 5, pp. 915–927, 2020.
- [3] D. Bajovic and N. Simic, "MARVEL - D1.3: Architecture definition for MARVEL framework," Tech. Rep., 2021. [Online]. Available: <https://doi.org/10.5281/zenodo.5463897>
- [4] G. Cerutti, R. Prasad, A. Brutti, and E. Farella, "Compact recurrent neural networks for acoustic event detection on low-energy low-complexity platforms," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 4, pp. 654–664, 2020.
- [5] B. Ramachandra, M. J. Jones, and R. R. Vatsavai, "A survey of single-scene video anomaly detection," *arXiv:2004.05993*, 2020.
- [6] G. Gao, J. Gao, Q. Liu, Q. Wang, and Y. Wang, "Cnn-based density estimation and crowd counting: A survey," *arXiv:2003.12783*, 2020.
- [7] A. Bakhtiarnia, Q. Zhang, and A. Iosifidis, "Single-layer vision transformers for more accurate early exits with less overhead," *arXiv:2105.09121*, 2021.
- [8] K. Drossos, S. Adavanne, and T. Virtanen, "Automated audio captioning with recurrent neural networks," in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2017, pp. 374–378.
- [9] A. Mesaros, T. Heittola, T. Virtanen, and M. D. Plumbley, "Sound event detection: A tutorial," *IEEE Signal Processing Magazine*, vol. 38, no. 5, pp. 67–83, 2021.
- [10] A. Politis, A. Mesaros, S. Adavanne, T. Heittola, and T. Virtanen, "Overview and evaluation of sound event localization and detection in dcase 2019," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 684–698, 2020.
- [11] G. Hagerer, V. Pandit, F. Eyben, and B. Schuller, "Enhancing lstm rnn-based speech overlap detection by artificially mixed data," in *Audio Engineering Society Conference: 2017 AES International Conference on Semantic Audio*. Audio Engineering Society, 2017.
- [12] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 1459–1462.
- [13] A. T. Jebb, S. Parrigon, and S. E. Woo, "Exploratory data analysis as a foundation of inductive research," *Human Resource Management Review*, vol. 27, no. 2, pp. 265–276, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1053482216300353>
- [14] G. Vasiliadis, M. Polychronakis, S. Antonatos, E. P. Markatos, and S. Ioannidis, "Regular expression matching on graphics hardware for intrusion detection," in *Recent Advances in Intrusion Detection*, E. Kirda, S. Jha, and D. Balzarotti, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 265–283.
- [15] L. Valerio, F. M. Nardini, A. Passarella, and R. Perego, "Dynamic hard pruning of neural networks at the edge of the internet," *arXiv preprint arXiv:2011.08545*, 2020.
- [16] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.
- [17] A. Chazapis, C. Pinto, Y. Gkoufas, C. Kozanitis, and A. Bilas, "A unified storage layer for supporting distributed workflows in kubernetes," in *Proceedings of the Workshop on Challenges and Opportunities of Efficient and Performant Storage Systems (CHEOPS '21)*. ACM, Apr. 2021.
- [18] H. Hukkelås, R. Mester, and F. Lindseth, "Deepprivacy: A generative adversarial network for face anonymization," in *International Symposium on Visual Computing*. Springer, 2019, pp. 565–578.
- [19] R. Chen, X. Chen, B. Ni, and Y. Ge, "Simsmap: An efficient framework for high fidelity face swapping," in *ACM International Conference on Multimedia*. ACM, 2020, pp. 2003–2011. [Online]. Available: <https://doi.org/10.1145/3394171.3413630>
- [20] Y. Y. Lin, C.-M. Chien, J.-H. Lin, H.-y. Lee, and L.-s. Lee, "Fragmentvc: Any-to-any voice conversion by end-to-end extracting and fusing fine-grained voice fragments with attention," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 5939–5943.
- [21] Y.-H. Chen, D.-Y. Wu, T.-H. Wu, and H.-y. Lee, "Again-vc: A one-shot voice conversion using activation guidance and adaptive instance normalization," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 5954–5958.
- [22] S. Arnautov, B. Trach, F. Gregor, T. Knauth, A. Martin, C. Priebe, J. Lind, D. Muthukumaran, D. O'Keefe, M. L. Stillwell, D. Goltzsche, D. Eyers, R. Kapitza, P. Pietzuch, and C. Fetzer, "SCONE: Secure linux containers with intel SGX," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*. Savannah, GA: USENIX Association, Nov. 2016, pp. 689–703. [Online]. Available: <https://www.usenix.org/conference/osdi16/technical-sessions/presentation/arnautov>
- [23] "Ladri in azione ai Mercatini di Natale di Trento: donna derubata del portafoglio," *La Voce Del Trentino, Quotidiano online indipendente*, available at: <https://www.lavocedeltrentino.it/2019/12/03/ladri-in-azione-ai-mercatini-di-natale-di-trento-donna-derubata-del-portafoglio/>.
- [24] "Piazza Santa Maria Maggiore incrocio dello spaccio. i video realizzati con le telecamere nascoste," *La Voce Del Trentino, Quotidiano online indipendente*. [Online]. Available: <https://www.lavocedeltrentino.it/2019/03/18/piazza-santa-maria-maggiore-incrocio-dello-spaccio-i-video-realizzati-con-le-telecamere-nascoste/>
- [25] "Parcheggio ex Zuffo, torna il rischio degrado." *La Voce Del Trentino, Quotidiano online indipendente*. [Online]. Available: <https://www.lavocedeltrentino.it/2018/06/04/parcheggio-ex-zuffo-torna-il-rischio-degrado/>
- [26] "Entro 15 anni l'interramento della Ferrovia a Trento." *TGR Trento - Testata Giornalistica Regionale Trento della RAI - Radio e Televisione Italiana*. [Online]. Available: <https://www.rainews.it/tgr/trento/video/2020/12/tnt-ferrovia-interramento-stazione-trento-brennero-treni-a5bc4c7f-89b7-431c-bdfa-329b52ba0aa4.html>
- [27] Transport Malta, "National transport strategy 2050," 2016. [Online]. Available: <https://www.transport.gov.mt/strategies/strategies-policies-actions/national-transport-strategy-and-transport-master-plan-1343>
- [28] S. Maas, "Bicycle sharing systems and their role in the promotion of cycling as a mode of transport in southern european island cities," Ph.D. dissertation, University of Malta, 6 2021.