AIMH Research Activities 2023

Nicola Aloia, Giuseppe Amato, Valentina Bartalesi, Lorenzo Bianchi, Paolo Bolettieri, Catherine Bosio, Michele Carraglia, Fabio Carrara, Vittore Casarosa, Luca Ciampi, et al.

Annual Report, No. 490372, DOI: 10.32079/ISTI-AR-2023/001, December 2023
Consiglio Nazionale delle Ricerche - Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"

AIMH Research Activities 2023

Nicola Aloia, Giuseppe Amato, Valentina Bartalesi, Lorenzo Bianchi, Paolo Bolettieri, Catherine Bosio, Michele Carraglia, Fabio Carrara, Vittore Casarosa, Luca Ciampi, Davide Alessandro Coccomini, Cesare Concordia, Silvia Corbara, Claudio De Martino, Marco Di Benedetto, Andrea Esuli, Fabrizio Falchi, Edoardo Fazzari, Claudio Gennaro, Gabriele Lagani, Emanuele Lenzi, Carlo Meghini, Nicola Messina, Alessio Molinari, Alejandro Moreo, Alessandro Nardi, Andrea Pedrotti, Nicolò Pratelli, Giovanni Puccetti, Fausto Rabitti, Pasquale Savino, Fabrizio Sebastiani, Gianluca Sperduti, Costantino Thanos, Luca Trupiano, Lucia Vadicamo, Claudio Vairo, Loredana Versienti.

Abstract
The AIMH (Artificial Intelligence for Media and Humanities) laboratory is dedicated to exploring and pushing the boundaries in the field of Artificial Intelligence, with a particular focus on its application in digital media and humanities. This lab's objective is to enhance the current state of AI technology particularly on deep learning, text analysis, computer vision, multimedia information retrieval, multimedia content analysis, recognition, and retrieval. This report encapsulates the laboratory's progress and activities throughout the year 2023.

Keywords
Multimedia Information Retrieval – Artificial Intelligence — Computer Vision — Similarity Search – Machine Learning – Text Classification – Deep Learning – Transfer learning – Representation Learning

[1] *AIMH Lab, ISTI-CNR, via Giuseppe Moruzzi, 1 - 56124 Pisa, Italy*
***Corresponding author**: giuseppe.amato@isti.cnr.it

Contents

AIMH
ARTIFICIAL INTELLIGENCE FOR MEDIA AND HUMANITIES

http://aimh.isti.cnr.it

Introduction

The Artificial Intelligence for Media and Humanities laboratory (AIMH) of the Information Science and Technologies Institute "Alessandro Faedo" (ISTI) of the Italian National Research Council (CNR) located in Pisa, has the mission to investigate and advance the state-of-the-art in the Artificial Intelligence field, specifically addressing applications to digital media and digital humanities, and also taking into account issues related to scalability.

The laboratory is composed of four research groups for a total of 37 people:

- 14 Researchers
  - *2 Directors of Research*
  - *3 Senior Researchers*
  - *7 Researchers*
  - *2 Temp. Researchers*
- 2 Technologists
  - *1 Temp. Senior Technologist*
  - *1 Technologists*
- 2 Technicians
- 2 Post Doc Fellows
- 10 PhD Students
- 3 Graduate Fellows
- 6 Research Associates

AI4Text

The AI4Text group is active in the area at the crossroads of machine learning and text analysis; it investigates novel algorithms and methodologies, and novel applications of these to different realms of text analysis. Topics within the above-mentioned area that are actively researched within the group include representation learning for text classification, transfer learning for cross-lingual and cross-domain text classification, sentiment classification, sequence learning for information extraction, learning to quantify, transductive text classification, cost-sensitive text classification, and applications of the above to domains such as authorship analysis, technology-assisted review, and native language identification. During this year, the group consisted of Fabrizio Sebastiani (Director of Research), Andrea Esuli (Senior Researcher), Alejandro Moreo (Researcher), Giovanni Puccetti (Post Doc Fellow), Silvia Corbara, Alessio Molinari, Andrea Pedrotti, and Gianluca Sperduti (PhD Students), and is led by Fabrizio Sebastiani.

Digital Humanities

Investigating AI-based solutions to represent, access, archive, and manage tangible and intangible cultural heritage data. This includes solutions based on Semantic Web technologies and ontologies, with a special focus on narratives and geospatial data, and solutions based on data analysis, recognition, and retrieval. During this year, the group consisted of Valentina Bartalesi, Cesare Concordia (Researchers), Michele Carraglia (Senior Technologist) Luca Trupiano (Technologist), Claudio De Martino(Graduate Fellows), Emanuele Lenzi, Nicolò Pratelli (PhD Students), Nicola Aloia, Vittore Casarosa, Carlo Meghini, and Costantino Thanos (Research Associates), and is led by Valentina Bartalesi.

Large-scale IR

Investigating efficient, effective, and scalable AI-based solutions for searching multimedia content in large datasets of non-annotated data. This includes techniques for multimedia content extraction and representation, scalable access methods for similarity search, multimedia database management. During this year, the group consisted of Claudio Gennaro, Pasquale Savino (Senior Researchers), Nicola Messina, Lucia Vadicamo (Researcher), Claudio Vairo (Researchers), Paolo Bolettieri (Technician), Fausto Rabitti (Research Associate), Lorenzo Bianchi, Giulio Federico (PhD Students), and is led by Claudio Gennaro.

Vision and Deep Learning

Investigating novel AI-based solutions to image and video content analysis, understanding, and classification. This includes techniques for detection, recognition (object, pedestrian, face, etc), classification, counting, feature extraction (low- and high-level, relational, cross-media, etc), anomaly detection also considering adversarial machine learning threats). We also have specific AI research fields such as Bio-Inspired Deep Learning. The group consists of Giuseppe Amato (Director of Research), Fabrizio Falchi (Senior Researcher), Marco Di Benedetto, Fabio Carrara, Luca Ciampi (Researchers), Alessandro Nardi (Technician), Davide Alessandro Coccomini, Edoardo Fazzari (PhD Students), and is led by Fabrizio Falchi.

The rest of the report is organized as follows. In Section 1, we summarize the research conducted on our main research fields. In Section 2, we describe the projects in which we were involved during the year. We report the complete list of papers we published in 2023, together with their abstract, in Section 3. The list of theses on which we were involved can be found in Section 4. In Section 5.2 we highlight the datasets we created and made publicly available during 2023.

1. Research Topics

In the following, we report a list of active research topics and subtopics at AIMH in 2023.

1.1 Bio-Inspired AI

We aim at leveraging our understanding of biological systems and structures in order to enhance current Deep Learning (DL) technologies. Research directions in this field involve, for example, biologically-grounded models of synaptic plasticity (*Hebbian* learning) and neural computation (Spiking Neural Networks – SNNs), respectively for enhancing the learning

abilities and the processing efficiency of state-of-the-art systems. We produced two surveys providing a comprehensive coverage of these domains [47, 48], and several conference contributions showing that taking advantage of biologically inspired models can help improving the learning capabilities of Deep Neural Networks (DNNs) in scenarios of data scarcity [43, 45, 46]. We also developed a novel solution for Hebbian learning that is able to speed-up training up to 70x by leveraging GPU acceleration [49]. Results are also summarized in the PhD dissertation [44].

1.2 AI for Text

1.2.1 Learning to quantify

We have presented the first book about quantification [32], wherein we describe, among other things, several quantification methods that have been proposed so far, discuss the main applications of quantification, and explain the main evaluation measures and protocols used in quantification research. The book also explores advanced topics and hints at potential directions in the quantification landscape.

While quantification has mostly been investigated in the context of prior probability shift, other types of shift have received less attention in the literature. In [40] we carried out many experiments aimed at confronting most representative quantification algorithms with different types of dataset shift. Our results reveal none of these methods behave robustly against all types of dataset shift.

We also investigated the case of Ordinal Quantifcation (OQ), i.e., multiclass quantification problems in which there is a total order among the classes. In [15] we introduced two new datasets for OQ research, compared existing algorithms across diverse fields, and proposed new OQ methods that prevent, via a regularization factor, non-smooth solutions. These new algorithms outperform existing ones in our experiments.

We studied the application of quantification in the context of evaluating fairness [33], i.e., measuring the bias of a classifier toward sensible attributes of a group, such as race of sex. We shown how the use quantification is effective on this task and also it has the benefit of not requiring to infer the sensible attribute for each element of the population.

1.2.2 Learning to classify text

In [58] we present Generalized Funneling (gFun), an extension of our previous Funnelling (Fun) method for cross-lingual text classification. While Fun employs a two-tier learning ensemble for heterogeneous transfer learning, gFun expands on this by allowing diverse view-generating functions as components. We present a specific gFun instance that outperforms Fun and other state-of-the-art baselines on multilingual multi-label text classification datasets.

We explored the use of Large Language Model for the detection of fake news in Italian [59], participating to the MULTI-Fake-DetectIVE EVALITA shared task.

1.2.3 Technology-assisted review

Technology-assisted review (TAR) is the task of supporting the work of human annotators who need to "review" automatically labelled data items, i.e., check the correctness of the labels assigned to these items by automatic classifiers. Since only a subset of such items can be feasibly reviewed, the goal of these algorithms is to exactly identify the items whose review is expected to be cost-effective. We have been working on this task since 2018, proposing TAR risk minimization algorithms that attempt to strike an optimal tradeoff between the contrasting goals of minimizing the cost of human intervention and maximizing the accuracy of the resulting labelled data. We worked [56] on improving an already well-performing method, i.e., MINECORE, by combining it with an original active learning method that combines the traditional relevance sampling and uncertainty sampling strategies. We are also continuing studying the unexpected negative impact of the SLD algorithm when used in TAR applications.

1.2.4 Authorship analysis

In [30] we explore the use of syllabic quantity, a metric scheme prevalent in Latin written text, for computational authorship attribution of Latin prose, on the grounds that authors often displayed preferences for specific metric patterns based on syllabic length. We investigate the effectiveness of incorporating rhythmic features derived from syllabic quantity in the authorship attribution task, alongside other topic-agnostic features. Our experiments reveal that these rhythmic features indeed enhance the discrimination of Latin prose authors.

In [29] we explore a new approach to authorship identification by using Diff-Vectors, where feature vectors represent unordered pairs of documents, indicating the absolute difference in relative frequencies. This method proves advantageous, especially with limited training data, bringing systematic improvements in authorship identification tasks. The experiments we carried out demonstrate its effectiveness specially in scenarios with sparse training data.

In [65] we investigated explainable techniques for authorship identification (AId), an area that remained, despite its practical utility for scholars, unexplored until now. Our research concentrated on the application of existing eXplainable Artificial Intelligence (XAI) techniques, including feature ranking, probing, factual, and counterfactual selection. We evaluated these techniques on three AId tasks related to cultural heritage. Our findings revealed further refinement is still needed for seamless integration into scholarly workflows.

1.3 Computer Vision

1.3.1 Visual Counting

The counting task aims to estimate the number of objects instances, like people or vehicles, in still images or videos. State-of-the-art current solutions are formulated as supervised deep learning-based problems belonging to one of two main categories: counting by detection and counting by regression. Detection-based approaches necessitate the prior identification of individual instances of objects. Conversely, regression-

based methods learn to establish a direct correlation between the features of an image and the count of objects in the scene. This connection can be established either directly or by estimating a target map, such as a density map (i.e., a continuous-valued map). Regression-based techniques prove to be more effective in scenarios with high occlusion. In [24], we proposed a modular model-agnostic deep learning-based counting pipeline for estimating the number of insects present in pictures of chromotropic sticky traps, reducing the need for manual trap inspections and minimizing human effort. Pest monitoring is one of the cornerstones underlying integrated pest management (IPM) solutions and is often employed in smart agriculture. Our solution generates a set of raw positions of the counted insects and associated confidence scores expressing their reliability, allowing practitioners to filter out unreliable predictions. We trained and assessed our technique by exploiting PST - Pest Sticky Traps, a new collection of dot-annotated images we created on purpose and publicly released [25], suitable for counting whiteflies.

1.3.2 Learning from Virtual Worlds

Amidst the burgeoning era of artificial intelligence, particularly within the sub-field of machine learning, a noteworthy sequence of significant outcomes has redirected the attention of industrial and research communities toward generating valuable data. This data serves as the foundation for training learning algorithms. While there are existing annotated datasets that have proven effective in yielding notable academic achievements and commercially viable products, numerous scenarios still demand labor-intensive human involvement to create high-quality training sets. A compelling solution lies in acquiring synthetic data from virtual environments that closely mimic the real world, wherein the labels are gathered automatically through interactions with the graphical engine. In [37] and [36], we introduced and made publicly available CrowdSim2 [66], a novel synthetic dataset derived from a simulator built on the Unity graphical engine. This data collection is designed for the purposes of people and vehicle detection and tracking, and it comprises thousands of images generated across diverse synthetic scenarios resembling real-world conditions. The labels are automatically generated and include precise bounding boxes that localize objects belonging to the two specified classes. We leveraged this new benchmark to assess the performance of some state-of-the-art detectors and trackers. Our findings demonstrate that these simulated scenarios serve as a valuable tool for evaluating the effectiveness of these algorithms in a controlled environment. Furthermore, in [22], we presented MC-GTA, which stands for Multi-Camera Grand Tracking Auto. It consists of a synthetic image dataset sourced from the immersive virtual environment offered by the highly realistic Grand Theft Auto 5 (GTA) video game, and it is suitable as a benchmark for multi-camera vehicle tracking (MCVT). Specifically, our dataset captures scenes from urban settings through multiple cameras strategically placed at various intersections. The annotations, featuring bounding boxes that precisely delineate vehicles and include unique IDs consistent across different video sources, have been automatically generated through interactions with the game engine. To assess this new dataset, we performed a performance evaluation exploiting a SOTA MCVT deep learning methodology using MC-GTA as a testing ground, showing that it can be a valuable benchmark that alleviates the need for real-world data.

1.3.3 Unsupervised Domain Adaptation

The automated identification of violent actions within videos poses a challenging yet vital task in numerous real-world scenarios since it is paramount for investigating the harmful abnormal contents from vast amounts of surveillance video data. Current SOTA video violence detectors rely on deep learning models trained in a supervised fashion. However, the success of these approaches hinges on two assumptions: (i) the existence of large collections of labeled data required for accurate model fitting during the training phase, and (ii) training (or source) and test (or target) datasets are independent and identically distributed. While plentiful annotated data is accessible for certain predefined domains, like ImageNet for image classification or COCO for object detection, obtaining manual annotations can be impractical for ad-hoc target domains or tasks. Consequently, models trained using pre-existing labeled data are often deployed in domains not encountered during training, leading to challenges arising from shifts in data distributions, commonly referred to as domain shifts between source and target domains. One potential approach to address this challenge is through unsupervised domain adaptation (UDA). This technique is designed to alleviate domain shifts between distinct domains by utilizing labeled data from the source domain and unlabeled data from the target domain. In essence, UDA methods leverage annotated data from the source domain and non-annotated data from the target domain, which is easily obtainable without the need for human labeling efforts. The key challenge lies in automatically extracting knowledge from this unlabeled data flow to narrow the gap between the two domains. In [**?**], we proposed a UDA scheme for detecting violent/non-violent actions present in trimmed videos. Our proposed solution is based on single image classification, randomly sampled from the frames making up the clips. The feature representations generated by the target images have been hooked and fed to a UDA module responsible for making them indiscriminative concerning the shift between the domains. We conducted experiments considering three datasets as source domains composed of videos of violent/non-violent scenes in general contexts and, as the target domain, a collection of clips of violent/non-violent actions in public transport. Preliminary results showed that our UDA scheme can help to improve the generalization capabilities of the considered models, mitigating the domain gap. This work was awarded as the best paper at the 3rd International Conference on Image Processing and Vision Engineering (IMPROVE 2023)[1].

[1] https://improve.scitevents.org/PreviousAwards.aspx#2023

1.3.4 Deep Fake Detection

We continued our research in the field of image and video deepfake detection by continuing to explore the challenges of generalizing various deep learning architectures in this field and extending our previous work in the field. In [27], we compared architectures based on Convolutional Neural Networks and Vision Transformer on video deepfake detection, showing that those based on the latter seem to be much more robust in identifying deepfakes created by methods not used for training set creation. Furthermore, in [28] we trained different classifiers to distinguish between pristine and synthetically generated images via text-to-image systems. From our experiments, it can be seen that it is rather easy to carry out this kind of classification, but only if the generation method on which the test is carried out is the same as the one used to create the training set, otherwise the classifiers tend not to be able to identify synthetic images.

1.3.5 Camera-based Smart Parking

The rise of the Internet of Things (IoT) has led to a variety of computer vision applications based on deep learning. These applications extract valuable information from the vast amount of data generated by edge devices, such as smart cameras. However, this promising approach presents new challenges, including limitations imposed by the constrained computational resources of these devices. Additionally, there are difficulties related to the generalization capabilities of AI-based models when faced with unforeseen scenarios not encountered during supervised training—a situation commonly experienced in this context. In [57], we proposed an effective deep learning (DL) approach utilizing knowledge distillation (KD) to localize vehicles in parking areas observed by multiple smart cameras. KD is a technique introduced to derive compressed models suitable for resource-constrained devices by transferring knowledge from complex, large models. In our specific scenario, we introduced a framework consisting of a robust detector serving as a teacher and multiple shallow models serving as students. This framework is designed explicitly for devices with limited computational resources and is intended for direct deployment on smart cameras. The teacher is pre-trained on diverse generic datasets and functions as an oracle, transferring its knowledge to the smaller nodes. On the other hand, the students rely solely on the distilled loss obtained from the oracle. They learn to detect vehicles in new scenarios they monitor without requiring additional labeled data. Preliminary results from experimental evaluations conducted under various settings indicate that the student models, trained exclusively with the distillation loss from the teacher, improve their performance, sometimes even surpassing the results achieved by the same models supervised with ground truth.

1.3.6 Life sciences

In the course of the year, we applied our vision-based AI system expertise to conduct collaborative research and development projects in healthcare and life sciences. These initiatives were carried out in conjunction with the Institute of Neuroscience at the CNR of Pisa. Specifically, in [21], we developed a deep learning-based counting system for biological structures that takes as input a microscopy image and produces as output the localization of the objects to be counted; furthermore, it also produces associated scores indicating the reliability of the detections that practitioners can use to exclude or include from the total count. This AI-based tool was exploited to create the first comprehensive atlas of Perineuronal nets (PNNs) [51], extracellular matrix aggregates surrounding the cell body of specific neurons in the brain that are involved in several forms of plasticity and clinical conditions. Their alterations are associated with several physiological processes and pathological conditions, e.g., psychiatric disorders such as schizophrenia, and they have attracted increasing interest in the scientific community that provides new insights into their role in different conditions and various animal models, including rodents, primates, and even human brain samples. However, our understanding of the PNN role in these phenomena is limited by the lack of highly quantitative maps of PNN distribution and association with specific cell types. To this end, we hope that our atlas will offer novel resources for understanding the organizational principles of the brain's extracellular matrix.

1.4 Multimedia Information Retrieval

1.4.1 Video Browsing

The prevalence of high-definition video cameras has led to an exponential growth in video data, making it the fastest-growing type of data on the Internet. This surge in video content has prompted extensive research into the development of large-scale video retrieval systems that are not only efficient and fast but also user-friendly for content search scenarios.

Within this context, we created a content-based video retrieval system called VISIONE, which participated in the Video Browser Showdown (VBS), an international competition evaluating the performance of interactive video retrieval systems. The evaluated tasks included visual Known-Item-Search (KIS), textual KIS, and Ad-hoc Video Search (AVS). The KIS tasks involved locating a specific video clip within a known collection, either visually or through textual descriptions. On the other hand, AVS required participants to find video shots matching a given textual description.

VISIONE took part in VBS in 2019 and from 2021 to 2023. In a joint publication [50], we detailed the competition settings, tasks, and results for the 2022 edition, offering insights into state-of-the-art methods employed by competing systems. Additionally, in [3, 2], we presented the system's latest release, which achieved first place in the KIS visual task and second place overall in the 2023 competition. Furthermore, we revisited the interface to be more user-friendly [4]. This novel version participated in a special session organized within CBMI where only non-expert users were allowed to interact with the systems. This served to factor out the role of expert users in solving challenging VBS tasks, in turn better

understanding the systems' potential.

Overall, VISIONE boasts various search functionalities enabling users to search for specific video segments using textual and visual queries, along with temporal search options. These functionalities include free text search, spatial color and object search, visual similarity search, and semantic similarity search.

The system leverages cutting-edge deep learning approaches for visual content analysis and incorporates highly efficient indexing techniques for scalability. Notably, it utilizes specially designed textual encodings for indexing and searching video content, allowing seamless integration with state-of-the-art text search engine technologies without the need for dedicated data structures or implementation concerns.

1.4.2 Similarity Search

The two papers presented contribute to the field of similarity search in distinct but complementary ways, each addressing challenges in large-scale information retrieval and metric search.

The first paper [70] focuses on enhancing Permutation-based Indexing (PBI) for approximate metric searching. The novel aspect of this work lies in its approach to permutation representations. Traditional PBI relies on object distances and their order relative to a set of anchors or pivots. However, this paper generalizes this concept by introducing permutations induced through space transformations and sorting functions. This innovation not only broadens the class of permutation representations usable in PBI but also introduces a new type of permutation representation calculated using distances from pairs of pivots. The significance of this method is its ability to produce longer permutations without increasing the number of object-pivot distance calculations, thereby enhancing efficiency in the search phase through the use of inverted files built on permutation prefixes.

The second paper [19] addresses a different aspect of similarity search. It recognizes the impact of deep learning on information retrieval (IR), particularly in how IR problems are now predominantly approached through the similarity of dense vectors derived from deep neural networks' latent spaces. Given the challenges in searching large-scale databases containing billions of such vectors, this paper proposes Vec2Doc, a method that converts dense vectors into sparse integer vectors. This conversion is critical as it allows the use of inverted indices, a traditional IR approach that works efficiently with sparse vectors. Preliminary results indicate that Vec2Doc offers a promising solution for large-scale vector-based IR problems.

In summary, both papers present innovative approaches to improve efficiency and effectiveness in similarity search. The first paper enhances PBI methods for metric search by introducing a more flexible and efficient permutation representation. In contrast, the second paper proposes a novel method for transforming the dense vectors, commonly used in modern IR systems, into a format compatible with traditional sparse vector methodologies. These contributions represent significant strides in adapting similarity search techniques to the evolving needs of large-scale information retrieval in the era of deep learning and artificial intelligence.

1.4.3 Cross-modal retrieval

We extended our previous studies on cross-modal image-text models for retrieval by tackling a novel, interesting modality, the human motion modality. Human motion is a sequence of 3D skeletal poses and can be acquired with ad-hoc hardware like motion capture devices or pose estimation methods directly from videos. In particular, we explored the latent interaction between human motions and their textual descriptions in a short-paper publication at SIGIR [55]. The work studies the possibility of employing the promising contrastive learning framework between text and motion data for searching large motion databases using textual prompts as queries. The previously employed streamlined methods primarily relied on query-by-example paradigms, which required the creation of an example motion query to search for similar ones – a quite unpractical and time-consuming operation. We proposed a set of model baselines based on recurrent neural networks and motion transformers, and we introduced suitable metrics for quantitatively evaluating the ability of our models to retrieve relevant motions from textual descriptions. This study paved the way for a more in-depth exploration of such an interesting yet unexplored task. The proposed system could be employed in a plethora of downstream applications, from the medical domain – where we may be interested in searching for gate or posture anomalies using motion-captured data – to game and film industries – where animators may need to search for specific 3D animation assets in large motion capture databases. This work deserved the best short-paper honorable mention at SIGIR 2023.

2. Projects & Activities

2.1 EU Projects

AI4media

Artificial Intelligence for the Society and the Media Industry (AI4Media) is a network of research excellence centres delivering advances in AI technology in the media sector. Funded under H2020-EU.2.1.1., AI4Media started in September 2020 and will end in August 2024.

Motivated by the challenges, risks and opportunities that the wide use of AI brings to media, society and politics, AI4Media aspires to become a centre of excellence and a wide network of researchers across Europe and beyond, with a focus on delivering the next generation of core AI advances to serve the key sector of Media, to make sure that the European values of ethical and trustworthy AI are embedded in future AI deployments, and to reimagine AI as a crucial beneficial enabling technology in the service of Society and

Media.

The leader of the AIMH team participating in AI4Media is Fabrizio Sebastiani.

The Craeft project aims to advance our understanding of the various aspects of crafts as a living and developing heritage, a sustainable source of income, and a means of expressing the mind through "imagery, technology, and sedimented knowledge". Drawing on disciplines such as Anthropology, Knowledge Representation, Cognitive Science, Art History, Advanced Digitisation, Audiovisual & Haptic Immersivity, and Computational Intelligence, the project will take a generative approach that can accommodate digital conservation, reenactable preservation, and scaling of approaches for different materials and techniques.

The leader of the AIMH team participating in Craeft is Valentina Bartalesi.

SoBigData++ is a project funded by the European Commission under the H2020 Programme INFRAIA-2019-1, started Jan 1 2020 and ending Dec 31, 2023. SoBigData++ proposes to create the Social Mining and Big Data Ecosystem: a research infrastructure (RI) providing an integrated ecosystem for ethic-sensitive scientific discoveries and advanced applications of social data mining on the various dimensions of social life, as recorded by “big data”. SoBigData plans to open up new research avenues in multiple research fields, including mathematics, ICT, and human, social and economic sciences, by enabling easy comparison, re-use and integration of state-of-the-art big social data, methods, and services, into new research. It plans to not only strengthen the existing clusters of excellence in social data mining research, but also create a pan-European, inter-disciplinary community of social data scientists, fostered by extensive training, networking, and innovation activities.

The leader of the AIMH team participating in SoBigData++ is Alejandro Moreo.

Social and hUman ceNtered XR (SUN) is a project funded by the European Commission under the H2020 Programme HORIZON-CL4-2022-HUMAN-01-14, started Dec 1 2022 and ending Nov 30 2025. SUN aims at investigating and developing extended reality (XR) solutions that integrate the physical and the virtual world in a convincing way, from a human and social perspective. The virtual world will be a means to augment the physical world with new opportunities for social and human interaction.

Our institute is the leading partner of the project and the coordinator is Giuseppe Amato.

2.2 NextGenerationEU PNRR National Projects

Then extended partnership titled Future Artificial Intelligence Research (hereafter FAIR) is the response of the Italian AI scientific community to the National Strategic Program. FAIR takes on the challenge to set the agenda of frontier research for the AI methodologies and techniques of tomorrow.

Well beyond currently available technologies, we need AI systems capable of interacting and collaborating with humans, of perceiving and acting within evolving contexts, of being aware of their own limitations and able to adapt to new situations, and interact appropriately in complex social settings, of being aware of their perimeters of security and trust, and of being attentive to the environmental and social impact that their implementation and execution may entail. In short, we need an AI that does not yet exist.

ITSERR

ITSERR is a project designed according to the needs of the Religious Studies scientific community to support the existing national infrastructure and bring it to a higher level of maturity, in terms of involvement of technology and ability to increase the innovation, quality and variety of the knowledge produced by the community of Religious Studies. The research aims at the development of Digital Maktaba (in Arabic "maktaba", "library") whose aim is to establish procedures for the extraction, management of libraries and archives and to develop virtuous models in the field of cataloguing that can accommodate texts written in non-Latin alphabets, starting with the case of the Arabic alphabet and testing it also with other alphabets.

MOST - Centro Nazionale per la mobilità sostenibile through collaboration with 24 universities, CNR and 24 large companies, has the mission of implementing modern, sustainable and inclusive solutions for the entire national territory.

The areas and technological fields of greatest interest in the project are: air mobility, sustainable road vehicles, water transport, rail transport, light vehicles and active mobility. The National Center will take care of making the mobility system more "green" as a whole and more "digital" in its management. It will do so through lightweight solutions and electric and hydrogen propulsion systems; digital systems for the reduction of accidents; more effective solutions for public

transport and logistics; a new model of mobility, as a service, accessible and inclusive.

MUCES

In this project ("A MUltimedia platform for Content Enrichment and Search in audiovisual archives"), we aim to develop advanced visual analysis and retrieval methodologies that can make unlabeled Italian audiovisual archives searchable through natural language and exemplar queries in a personalized manner. These methodologies will be multi-modal, adaptable to long-tail concepts, and efficient for large-scale archives. The project aims to bridge cutting-edge research in Computer Vision and Content-based Image Retrieval to enhance the accessibility of audiovisual cultural heritage in Italy. At the core of the project lies a new unifying synergy between cutting-edge research in Computer Vision, Machine Learning, and large-scale Content-Based Retrieval. The project brings together the research experiences and expertise of two internationally-recognized research teams: the AImageLab research group at UNIMORE and the Artificial Intelligence for Media and Humanities laboratory at ISTI CNR, encompassing years of expertise in Multimedia, Similarity Search, and Computer Vision.

PAPYRI

The PRIN PNRR Reconstructing Fragmentary Papyri through Human-Machine Interaction project investigates the application of Artificial Intelligence to the reconstruction of specific lots of papyrus fragments from two Italian papyrological collections: the Papiri collection of the Società Italiana, stored at the Istituto Papirologico "G. Vitelli" (University of Florence), and the Papiri collection of the University of Genova. Following an innovative and interdisciplinary approach, the two papyrological teams will work closely together with ISTI in implementing an already prototyped interactive software aimed to assist papyrologists in the screening phase and, mainly, in the matching of fragments, allowing the user to evaluate and revise multiple hypotheses of reconstruction. The system will take advantage of visual information of both the front and the back of the fragments by exploiting the continuity of the fibre patterns and by taking into account positional information and additional constraints supplied by the expert.

SERICS
SECURITY AND RIGHTS IN THE CYBERSPACE

AIMH collaborates to SEcurity and Rights in the CyberSpace (SERICS)

AIMH is involved in particular in Spoke 1 and Spoke 2:

Spoke 1. Human, Social, and Legal Aspects - Spoke 1 aims at protecting social rights and values in the Cyberspace. The research will involve public and private stakeholders in the implementation of innovative technological, legal, ethical and organizational solutions, in order to strengthen the resilience and digital sovereignty of the public and private sectors and, therefore, of the country system. The Spoke 1 has two main projects: Cyberights and DiSe; the former is devoted to study and promote legal and ethical aspects for cyberspace while the latter to digital sovereignty aspects that affect also the underlying digital technologies."

Spoke 2. Misinformation and Fakes The Spoke will establish an excellent multidisciplinary structure that, leveraging intelligence analysis, artificial intelligence, political analysis, data science, and web intelligence capabilities, employs suitable tools and methods to support information disorder awareness. The Spoke 2 has four main projects: DETERRENCE, FF4LL, HUMANE, IDA.

THE
Tuscany Health Ecosystem

The objective of Spoke 8 is to leverage the existing critical mass of neuroscientists in Tuscany, which has a long and well-established tradition and strength and a vast breadth of highly specialized expertise and tools, integrating it with the recruitment of a diverse set of interdisciplinary knowledge, ranging from chemistry to computational sciences, data sciences, synthetic biology, bioinformatics, high-throughput analysis of gene expression (-omics), imaging and others. Due to its intrinsic multidisciplinary nature, the spoke will naturally have strong links with other spokes in the overall project.

ISTI is active in particular Sub-project 8 - Patient-derived stem cells and "brain-in-a-dish" cultures: a cellular platform for target validation and drug screening. The aim of this subproject is to establish cell cultures of neurons from patients of Retinite Pigmentosa (RP) and Autism Spectrum Diseases (ASD) aimed at their diagnosis by Artificial Intelligence (AI) and at their pharmacological and cell therapy. Moreover, solutions will also be developed to use Cultured Neuronal Networks (CultNN), that is biological cultures of real neuronal cells or brain-in-a-dish, to execute Artificial Intelligence (AI) tasks. CultNN will be directly used for AI, rather than bioinspired devices or software implementations of neural networks. To do so, CultNN of neurons derived from patient reprogrammed cells (hiPSCs) will be established, and computational models of the neural activity-designed AI training methods will be developed to be applied to CultNN. Finally, imaging solutions will be developed, based on AI, to analyze the pupil area and the retina, and CultNN used to screen for drug panels and validate targets.

2.3 Other National Projects

AIsafety

In the industrial realm, it's well known that many accidents involving machine operators in production processes are

somehow linked to the operator's behavior and various types of errors. No matter how sophisticated, traditional control and supervision systems have been unable to entirely eliminate or satisfactorily manage these risks. This issue becomes particularly acute when machines malfunction, such as product jams that need to be cleared or other breakdowns, requiring operators to access dangerous areas of the machine and actively intervene. Predicting these operational conditions linked to faults, malfunctions, and operator errors is challenging during the design phase. The required activities may not be easily definable and hence can lead to accidents or near-miss incidents. In this context, equipping machines with networks of sensors and Artificial Intelligence (AI) systems capable of interpreting even new operational situations, recognizing potential risk conditions for operators, and generating appropriate commands for the machines is seen as an effective path towards enhancing user safety. AISAFETY steps into this challenging scenario with a groundbreaking approach. By integrating AI, RFID technology, and a network of intelligent cameras, it offers a proactive solution to enhance workplace safety. The system's AI brain analyzes data from the cameras and RFID tags worn by operators, understanding the workspace dynamics in real-time. If it detects any danger, it can instantly instruct the machines to stop or adjust their operation, often before any human can react. Moreover, AISAFETY respects the indispensable role of human judgment. Supervisors monitor the system, ensuring that the balance between automated safety measures and human oversight is maintained. Compliant with strict safety and data protection regulations, AISAFETY is not just about employing advanced technology; it's about responsibly creating a safer industrial environment where technology and human expertise collaborate to prevent accidents. Claudio Gennaro is the scientific responsible for ISTI.

HDN

Hypermedia Dante Network (HDN) is a three year (2020-2023) Italian National Research Project (PRIN) which aims to extend the ontology and tools developed by AIMH team to represent the sources of Dante Alighieri's minor works to the more complex world of the Divine Comedy. In particular, HDN aims to enrich the functionalities of the DanteSources Web application (https://dantesources.dantenetwork.it/) in order to efficiently recover knowledge about the Divine Comedy. Relying on some of the most important scientific institutions for Dante studies, such as the Italian Dante Society of Florence, HDN makes use of specialized skills, essential for the population of ontology and the consequent creation of a complete and reliable knowledge base. Knowledge will be published on the Web as Linked Open Data and will be access through a user-friendly Web application.

IMAGO

The IMAGO (Index Medii Aevi Geographiae Operum) is a three year (2020-2023) Italian National Research Project (PRIN) that aims at creating a knowledge base of the critical editions of Medieval and Humanistic Latin geographical works (VI-XV centuries). Up to now, this knowledge has been collected in many paper books or several databases, making it difficult for scholars to retrieve it easily and to produce a complete overview of these data. The goal of the project is to develop new tools that satisfy the needs of the academic research community, especially for scholars interested in Medieval and Renaissance Humanism geography. Using Semantic Web technologies, AIMH team will develop an ontology providing the terms to represent this knowledge in a machine-readable form. A semi-automatic tool will help the scholars to populate the ontology with the data included in authoritative critical editions. Afterwards, the tool will automatically save the resulting graph into a triple store. On top of this graph, a Web application will be developed, which will allow users to extract and display the information stored in the knowledge base in the form of maps, charts, and tables.

The leader of the AIMH team participating in IMAGO is Valentina Bartalesi.

INAROS

INtelligenza ARtificiale per il mOnitoraggio e Supporto agli anziani (INAROS) is a 2-year project funded by Regione Toscana, Istituto di Scienza e Tecnologie dell'Informazione "A.Faedo" (ISTI) del CNR, Visual Engines srl. The main goal of the INAROS project is to build solutions for monitoring and surveillance of the elderly based on the use of autonomous smart cameras. Computer vision algorithms will be developed by leveraging artificial intelligence, in particular deep learning to automatically and in real time analyze video streams from smart cameras positioned in the home environment. To achieve these results, techniques will be developed for the tracking and detection of the elderly person's activity in the home environment and for the discovery of new activities and abnormalities of the elderly through off-line analysis of temporal patterns of learned events. Claudio Gennaro is the scientific coordinator of the project.

MIGHT

Gut Microbiota as a bioremediator for gut-health in infants (MIGHT) is a 2 years project funded by the 'Progetti@CNR' (Area: Tecnologie a supporto delle fasce più fragili: giovani e anziani) program. The understanding of the relationship among diet, metabolites, and host/microbiota is a key challenge to investigate personalized nutrition for the most fragile segments of the population, and the modulation of the gut microbiota through dietary interventions is one of the most promising approaches. MIGHT project has the ambition to disentangle key research questions behind food proteins modifications and the effects on the host microbiota. MIGHT is interlinked with a recently granted project from the EU (MAMMAL, EIT Food, Innovation). While the MAMMAL proposal focuses on biochemical aspects, MIGHT enlightens IT solutions for the organization, management, and access to the data produced, and for their exploration to translate the experimental evidence from newborns to the general pop-

ulation, in particular to elderly. Partners of the project are: Istituto di Biologia e Biotecnologia Agraria (IBBA) - CNR, Istituto Sistema di Produzione Animale in Ambiente Mediterraneo (ISPAAM) - CNR, Istituto di Scienza e Tecnologie dell'Informazione (ISTI), CNR. Cesare Concordia is the scientific responsible for ISTI.

CY4Gate/SWOAD

The project involves the analysis, study, and implementation of a system for searching and recognizing images depicting works of art. The project includes the development of advanced solutions for the analysis and extraction of information from images, image search and recognition, and the indexing of information extracted from images. Specifically, the image analysis component extracts information (visual features) that can be used for image search and recognition based on visual content, using state-of-the-art artificial intelligence and computer vision techniques, including cutting-edge deep neural networks. The image search and recognition component receives images used as queries, sends them to the image analysis component, compares them with those in the dataset, and returns images whose visual content is most similar to the query. Finally, the image indexing component operates incrementally, coordinating with the image analysis component to extract information from images to be included in the system. It creates a database of visual features for comparison and recognition, allowing for fast and efficient search. The project is carried out within a collaboration between CNR-ISTI and the Comando Carabinieri per la Tutela del Patrimonio Culturale. It is funded through a sub-contract given to CNR-ISTI by the company CY4Gate, as part of the SWOAD (Stolen Work Of Art Detection) project.

3. Publications

In this section, we report the complete list of papers we published in 2023 organized in four categories: journals, proceedings, magazines, others, and pre-prints.

3.1 Journals

In this section, we report the paper we published (or accepted for publication) in journals during 2023, in alphabetic order of the first author. Our works were published in the following journals (ordered by Impact Factor):

- **Cell Reports**
  Cell Press, IF 8.8: [51]
- **Journal of Big Data**
  Springer, IF 8.1: [67]
- **Neural Networks**
  Elsevier, IF 7.8: [39]
- **Neural Computing and Applications**
  Springer, IF 6.0: [60]
- **Transactions on Information Systems**
  ACM, IF 5.6: [58]
- **International Journal of Machine Learning and Cybernetics**
  Springer, IF 5.6: [34]
- **Journal of the Association for Information Systems**
  ASIS, IF 5.6: [30]
- **Ecological Informatics**
  Elsevier, IF 5.1: [24]
- **International Journal of Digital Earth**
  Taylor and Francis, IF 5.1: [7]
- **Journal of Artificial Intelligence Research**
  AI Access Foundation, IF 5.0; [33]
- **Access**
  IEEE, I.F. 3.9: [5]
- **Multimedia Systems**
  Springer, I.F. 3.9: [50]
- **PeerJ Computer Science**
  PeerJ, IF 3.8: [14]
- **Information Systems**
  Elsevier, IF 3.7: [70]
- **Multimedia Tools and Applications**
  Springer, IF 3.6: [41]
- **Journal of Imaging**
  MDPI, IF 3.2: [27]
- **Semantic Web**
  IOS Press, IF 3.0: [9]
- **Journal on Computing and Cultural Heritage**
  ACM, IF 2.4: [11]
- **Heritage**
  MDPI, IF 1.7: [72]
- **Intelligent Systems with Applications**
  Elsevier, IF n/a: [56]

3.1.1 A Comprehensive Atlas of Perineuronal Net Distribution and Colocalization with Parvalbumin in the Adult Mouse Brain

L. Lupori, V. Totaro, S. Cornuti, L. Ciampi, F. Carrara, E. Grilli, A. Viglione, F. Tozzi, E. Putignano, R. Mazziotti, G. Amato, C. Gennaro, P. Tognini, T. Pizzorusso. Cell Reports, Cell Press. [51]

Perineuronal nets (PNNs) surround specific neurons in the brain and are involved in various forms of plasticity and clinical conditions. However, our understanding of the PNN role in these phenomena is limited by the lack of highly quantitative maps of PNN distribution and association with specific cell types. Here, we present a comprehensive atlas of Wisteria floribunda agglutinin (WFA)-positive PNNs and colocalization with parvalbumin (PV) cells for over 600 regions of the adult mouse brain. Data analysis shows that PV expression is a good predictor of PNN aggregation. In the cortex, PNNs are dramatically enriched in layer 4 of all primary sensory areas in correlation with thalamocortical input density, and their distribution mirrors intracortical connectivity patterns. Gene expression analysis identifies many PNN-correlated genes. Strikingly, PNN-anticorrelated transcripts are enriched in synaptic plasticity genes, generalizing PNNs' role as circuit stability factors.

3.1.2

A deep learning-based pipeline for whitefly pest abundance estimation on chromotropic sticky traps

L. Ciampi, V. Zeni, L. Incrocci, A. Canale, G. Benelli, F. Falchi, G. Amato, S. Chessa. Ecological Informatics, Elsevier. [24]

Integrated Pest Management (IPM) is an essential approach used in smart agriculture to manage pest populations and sustainably optimize crop production. One of the cornerstones underlying IPM solutions is pest monitoring, a practice often performed by farm owners by using chromotropic sticky traps placed on insect hot spots to gauge pest population densities. In this paper, we propose a modular model-agnostic deep learning-based counting pipeline for estimating the number of insects present in pictures of chromotropic sticky traps, thus reducing the need for manual trap inspections and minimizing human effort. Additionally, our solution generates a set of raw positions of the counted insects and confidence scores expressing their reliability, allowing practitioners to filter out unreliable predictions. We train and assess our technique by exploiting PST - Pest Sticky Traps, a new collection of dot-annotated images we created on purpose and we publicly release, suitable for counting whiteflies. Experimental evaluation shows that our proposed counting strategy can be a valuable Artificial Intelligence-based tool to help farm owners to control pest outbreaks and prevent crop damages effectively. Specifically, our solution achieves an average counting error of approximately compared to human capabilities requiring a matter of seconds, a large improvement respecting the time-intensive process of manual human inspections, which often take hours or even days.

3.1.3

An exploratory approach to data driven knowledge creation

C. Thanos, C. Meghini, V. Bartalesi, G. Coro. Journal of Big Data, Springer. [67]

This paper describes a new approach to knowledge creation that is instrumental for the emerging paradigm of data-intensive science. The proposed approach enables the acquisition of new insights from the data by exploiting existing relationships between diverse types of datasets acquired through various modalities. The value of data consistently improves when it can be linked to other data because linking multiple types of datasets allows creating novel data patterns within a scientific data space. These patterns enable the exploratory data analysis, an analysis strategy that emphasizes incremental and adaptive access to the datasets constituting a scientific data space while maintaining an open mind to alternative possibilities of data interconnectivity. A technology, the Linked Open data (LOD), was developed to enable the linking of datasets. We argue that the LOD technology presents several limitations that prevent the full exploitation of this technology to acquire new insights. In this paper, we outline a new approach that enables researchers to dynamically create data patterns in a research data space by exploiting explicit and implicit/hidden relationships between distributed research datasets. This dynamic creation of data patterns enables the exploratory data analysis strategy.

3.1.4

A roadmap for craft understanding, education, training, and preservation

X. Zabulis, N. Partarakis, I. Demeridou, P. Doulgeraki, E. Zidianakis, A. Argyros, M. Theodoridou, Y. Marketakis, C. Meghini, V. Bartalesi, N. Pratelli, C. Holz, P. Streli, M. Meier, M.K. Seidler, L. Werup, P. F. Sichani, S. Manitsaris, G. Senteri, A. Dubois, C. Ringas, A. Ziova, E. Tasiopoulou, D. Kaplanidi, D. Arnaud, P. Hee, G. Canavate, M.A. Benvenuti, J. Krivokapic Heritage [72]

A roadmap is proposed that defines a systematic approach for craft preservation and its evaluation. The proposed roadmap aims to deepen craft understanding so that blueprints of appropriate tools that support craft documentation, education, and training can be designed while achieving preservation through the stimulation and diversification of practitioner income. In addition to this roadmap, an evaluation strategy is proposed to validate the efficacy of the developed results and provide a benchmark for the efficacy of craft preservation approaches. The proposed contribution aims at the catalyzation of craft education and training with digital aids, widening access and engagement to crafts, economizing learning, increasing exercisability, and relaxing remoteness constraints in craft learning.

3.1.5

Blind bleed-through removal in color ancient manuscripts

M. Hanif, A. Tonazzini, S.F. Hussain, U. Habib, E. Salerno, P. Savino, Z. Halim. Multimedia Tools and Applications, Springer [41] *Archaic manuscripts are an important part of ancient civilization. Unfortunately, such documents are often affected by various age related degradations, which impinge their legibility and information contents, and destroy their original look. In general, these documents are composed of three layers of information: foreground text, background, and unwanted degradation in the form of patterns interfering with the main text. In this work, we are presenting a color space based image segmentation technique to separate and remove the bleed-through degradation in digital ancient manuscripts. The main theme is to improve their readability and restore their original aesthetic look. For each pixel, a feature vector is created using color spectral and spatial location information. A pixel based segmentation method using Gaussian Mixture Model (GMM) is employed, assuming that each feature vector corresponds to a Gaussian distribution. Based on this assumption, each pixel is supposed to be drawn from a mixture of Gaussian distribution, with unknown parameters. The Expectation-Maximization (EM) approach is then used to estimate the unknown GMM parameters. The appropriate class label for each pixel is then estimated using posterior probability and GMM parameters. Unlike other binarization based document restoration method where the focus is on text extraction, we are more interested in restoring the aesthetically pleasing look of the ancient documents.The experimental results validate the usefulness of proposed method in terms of successful bleed-through identification and removal, while preserving foreground-text and background information.*

3.1.6
Conditioned Cooperative training for semi-supervised weapon detection

J.L. Salazar-González, J.A. Álvarez-García, F.J. Rendón-Segador, F. Carrara. Neural Networks, Elsevier [39] *Violent assaults and homicides occur daily, and the number of victims of mass shootings increases every year. However, this number can be reduced with the help of Closed Circuit Television (CCTV) and weapon detection models, as generic object detectors have become increasingly accurate with more data for training. We present a new semi-supervised learning methodology based on conditioned cooperative student-teacher training with optimal pseudo-label generation using a novel confidence threshold search method and improving both models by conditional knowledge transfer. Furthermore, a novel firearms image dataset of 458,599 images was collected using Instagram hashtags to evaluate our approach and compare the improvements obtained using a specific unsupervised dataset instead of a general one such as ImageNet. We compared our methodology with supervised, semi-supervised and self-supervised learning techniques, outperforming approaches such as YOLOv5 m (up to +19.86), YOLOv5l (up to +6.52) Unbiased Teacher (up to +10.5 AP), DETReg (up to +2.8 AP) and UP-DETR (up to +1.22 AP).*

3.1.7
From unstructured texts to semantic story maps

V. Bartalesi, G. Coro, E. Lenzi, P. Pagano, N. Pratelli. International Journal of Digital Earth, Taylor and Francis. [7]

Digital maps greatly support storytelling about territories, especially when enriched with data describing cultural, societal, and ecological aspects, conveying emotional messages that describe the territory as a whole. Story maps are interactive online digital narratives that can describe a territory beyond its map by enriching the map with text, pictures, videos, and other multimedia information. This paper presents a semi-automatic workflow to produce story maps from textual documents containing territory data. An expert first assembles one territory-contextual document containing text and images. Then, automatic processes use natural language processing and Wikidata services to (i) extract key concepts (entities) and geospatial coordinates associated with the territory, (ii) assemble a logically-ordered sequence of enriched story-map events, and (iii) openly publish online story maps and an interoperable Linked Open Data semantic knowledge base for event exploration and inter-story correlation analyses. Our workflow uses an Open Science-oriented methodology to publish all processes and data. Through our workflow, we produced story maps for the value chains and territories of 23 rural European areas of 16 countries. Through numerical evaluation, we demonstrated that territory experts considered the story maps effective in describing their territories, and appropriate for communicating with citizens and stakeholders.

3.1.8
Generalized Funnelling: Ensemble Learning and Heterogeneous Document Embeddings for Cross-Lingual Text Classification.

A. Moreo, A. Pedrotti, F. Sebastiani. Transactions on Information Systems, ACM] [58]

Funnelling (Fun) is a recently proposed method for cross-lingual text classification (CLTC) based on a two-tier learning ensemble for heterogeneous transfer learning (HTL). In this ensemble method, 1st-tier classifiers, each working on a different and language-dependent feature space, return a vector of calibrated posterior probabilities (with one dimension for each class) for each document, and the final classification decision is taken by a meta-classifier that uses this vector as its input. The meta-classifier can thus exploit class-class correlations, and this (among other things) gives Fun an edge over CLTC systems in which these correlations cannot be brought to bear. In this article, we describe Generalized Funnelling (gFun), a generalization of Fun consisting of an HTL architecture in which 1st-tier components can be arbitrary view-generating functions, i.e., language-dependent functions that each produce a language-independent representation ("view") of the (monolingual) document. We describe an instance of gFun in which the meta-classifier receives as input a vector of calibrated posterior probabilities (as in Fun) aggregated to other embedded representations that embody other types of correlations, such as word-class correlations (as encoded by Word-Class Embeddings), word-word correlations (as encoded by Multilingual Unsupervised or Supervised Embeddings), and word-context correlations (as encoded by multilingual BERT). We show that this instance of gFun substantially improves over Fun and over state-of-the-art baselines by reporting experimental results obtained on two large, standard datasets for multilingual multilabel text classification. Our code that implements gFun is publicly available.

3.1.9
Graph-based methods for Author Name Disambiguation: a survey

M. De Bonis, F. Falchi, P. Manghi. PeerJ Computer Science [14]

Scholarly knowledge graphs (SKG) are knowledge graphs representing research-related information, powering discovery and statistics about research impact and trends. Author name disambiguation (AND) is required to produce high-quality SKGs, as a disambiguated set of authors is fundamental to ensure a coherent view of researchers' activity. Various issues, such as homonymy, scarcity of contextual information, and cardinality of the SKG, make simple name string matching insufficient or computationally complex. Many AND deep learning methods have been developed, and interesting surveys exist in the literature, comparing the approaches in terms of techniques, complexity, performance, etc. However, none of them specifically addresses AND methods in the context of SKGs, where the entity-relationship structure can be exploited. In this paper, we discuss recent graph-based methods for AND, define a framework through which such methods can be confronted, and catalog the most popular datasets and benchmarks used to test such methods. Finally, we outline possible directions for future work on this topic.

3.1.10
Improved risk minimization algorithms for technology-assisted review

A. Molinari, A. Esuli, F. Sebastiani. Intelligent Systems with Applications, Elsevier] [56]

MINECORE is a recently proposed decision-theoretic algorithm for technology-assisted review that attempts to minimise the expected costs of review for responsiveness and privilege in e-discovery. In MINECORE, two probabilistic classifiers that classify documents by responsiveness and by privilege, respectively, generate posterior probabilities. These latter are fed to an algorithm that returns as output, after applying risk minimization, two ranked lists, which indicate exactly which documents the annotators should review for responsiveness and which documents they should review for privilege. In this paper we attempt to find out if the performance of MINECORE can be improved (a) by using, for the purpose of training the two classifiers, active learning (implemented either via relevance sampling, or via uncertainty sampling, or via a combination of them) instead of passive learning, and (b) by using the Saerens-Latinne-Decaestecker algorithm to improve the quality of the posterior probabilities that MINECORE receives as input. We address these two research questions by carrying out extensive experiments on the RCV1-v2 benchmark. We make publicly available the code and data for reproducing all our experiments.

3.1.11
Induced permutations for approximate metric search

L. Vadicamo, G. Amato, C. Gennaro. Information Systems [70]

Permutation-based Indexing (PBI) approaches have been proven to be particularly effective for conducting large-scale approximate metric searching. These methods rely on the idea of transforming the original metric objects into permutation representations, which can be efficiently indexed using data structures such as inverted files. The standard conceptualization of permutation associated with a metric object involves only the use of object distances and their relative orders from a set of anchors called pivots. In this paper, we generalized this definition in order to enlarge the class of permutation representations that can be used by PBI approaches. In particular, we introduced the concept of permutation induced by a space transformation and a sorting function, and we investigated which properties these transformations should possess to produce permutations that are effective for metric search. Furthermore, as a practical outcome, we defined a new type of permutation representation that is calculated using distances from pairs of pivots. This proposed technique allowed us to produce longer permutations than traditional ones for the same number of object-pivot distance calculations. The advantage lies in the fact that when longer permutations are employed, the use of inverted files built on permutation prefixes leads to greater efficiency in the search phase.

3.1.12
Interactive video retrieval in the age of effective joint embedding deep models: lessons from the 11th VBS

J. Lokoč, S. Andreadis, W. Bailer, A. Duane, C. Gurrin, Z. Ma, N. Messina, T.N. Nguyen, L. Peška, L. Rossetto, L. Sauter, K. Schall, K. Schoeffmann, O. S. Khan, F. Spiess, L. Vadicamo, S. Vrochidis. Multimedia Systems [50]

This paper presents findings of the eleventh Video Browser Showdown competition, where sixteen teams competed in known-item and ad-hoc search tasks. Many of the teams utilized state-of-the-art video retrieval approaches that demonstrated high effectiveness in challenging search scenarios. In this paper, a broad survey of all utilized approaches is presented in connection with an analysis of the performance of participating teams. Specifically, both high-level performance indicators are presented with overall statistics as well as in-depth analysis of the performance of selected tools implementing result set logging. The analysis reveals evidence that the CLIP model represents a versatile tool for cross-modal video retrieval when combined with interactive search capabilities. Furthermore, the analysis investigates the effect of different users and text query properties on the performance in search tasks. Last but not least, lessons learned from search task preparation are presented, and a new direction for ad-hoc search based tasks at Video Browser Showdown is introduced.

3.1.13
Measuring Fairness Under Unawareness of Sensitive Attributes: A Quantification-Based Approach.

A. Fabris, A. Esuli, A. Moreo, F. Sebastiani. Journal of Artificial Intelligence Research, AI Access Foundation [33]

Algorithms and models are increasingly deployed to inform decisions about people, inevitably affecting their lives. As a consequence, those in charge of developing these models must carefully evaluate their impact on different groups of people and favour group fairness, that is, ensure that groups determined by sensitive demographic attributes, such as race or sex, are not treated unjustly. To achieve this goal, the availability (awareness) of these demographic attributes to those evaluating the impact of these models is fundamental. Unfortunately, collecting and storing these attributes is often in conflict with industry practices and legislation on data minimisation and privacy. For this reason, it can be hard to measure the group fairness of trained models, even from within the companies developing them. In this work, we tackle the problem of measuring group fairness under unawareness of sensitive attributes, by using techniques from quantification, a supervised learning task concerned with directly providing group-level prevalence estimates (rather than individual-level class labels). We show that quantification approaches are particularly suited to tackle the fairness-under-unawareness problem, as they are robust to inevitable distribution shifts while at the same time decoupling the (desirable) objective of measuring group fairness from the (undesirable) side effect of allowing the inference of sensitive attributes of individuals. More in detail, we show that fairness under unawareness can be cast as a quantification problem and solved with proven methods from the quantification literature. We show that these methods outperform previous approaches to measure demographic parity in five experimental protocols, corresponding to important challenges that complicate the estimation of classifier fairness under unawareness.

3.1.14
NoR-VDPNet++: Real-Time No-Reference Image Quality Metrics

F. Banterle, A. Artusi, A. Moreo, F. Carrara, P. Cignoni. IEEE Access [5]

Efficiency and efficacy are desirable properties for any evaluation metric having to do with Standard Dynamic Range (SDR)

imaging or with High Dynamic Range (HDR) imaging. However, it is a daunting task to satisfy both properties simultaneously. On the one side, existing evaluation metrics like HDR-VDP 2.2 can accurately mimic the Human Visual System (HVS), but this typically comes at a very high computational cost. On the other side, computationally cheaper alternatives (e.g., PSNR, MSE, etc.) fail to capture many crucial aspects of the HVS. In this work, we present NoR-VDPNet++, a deep learning architecture for converting full-reference accurate metrics into no-reference metrics thus reducing the computational burden. We show NoR-VDPNet++ can be successfully employed in different application scenarios.

3.1.15
On the Generalization of Deep Learning Models in Video Deepfake Detection

D.A. Coccomini, R. Caldelli, F. Falchi, C. Gennaro. Journal of Imaging, MDPI [27]

The increasing use of deep learning techniques to manipulate images and videos, commonly referred to as "deepfakes", is making it more challenging to differentiate between real and fake content, while various deepfake detection systems have been developed, they often struggle to detect deepfakes in real-world situations. In particular, these methods are often unable to effectively distinguish images or videos when these are modified using novel techniques which have not been used in the training set. In this study, we carry out an analysis of different deep learning architectures in an attempt to understand which is more capable of better generalizing the concept of deepfake. According to our results, it appears that Convolutional Neural Networks (CNNs) seem to be more capable of storing specific anomalies and thus excel in cases of datasets with a limited number of elements and manipulation methodologies. The Vision Transformer, conversely, is more effective when trained with more varied datasets, achieving more outstanding generalization capabilities than the other methods analysed. Finally, the Swin Transformer appears to be a good alternative for using an attention-based method in a more limited data regime and performs very well in cross-dataset scenarios. All the analysed architectures seem to have a different way to look at deepfakes, but since in a real-world environment the generalization capability is essential, based on the experiments carried out, the attention-based architectures seem to provide superior performances.

3.1.16
Syllabic quantity patterns as rhythmic features for Latin authorship attribution

S. Corbara, A. Moreo, F. Sebastiani. Journal of the Association for Information Systems, ASIS [30]

It is well known that, within the Latin production of written text, peculiar metric schemes were followed not only in poetic compositions, but also in many prose works. Such metric patterns were based on so-called syllabic quantity, that is, on the length of the involved syllables, and there is substantial evidence suggesting that certain authors had a preference for certain metric patterns over others. In this research we investigate the possibility to employ syllabic quantity as a base for deriving rhythmic features for the task of computational authorship attribution of Latin prose texts. We test the impact of these features on the authorship attribution task when combined with other topic-agnostic features. Our experiments, carried out on three different datasets using support vector machines (SVMs) show that rhythmic features based on syllabic quantity are beneficial in discriminating among Latin prose authors.

3.1.17
Training a shallow NN to erase ink seepage in historical manuscripts based on a degradation model

P. Savino, A. Tonazzini. Neural Computing and Applications, Elsevier. [60]

In historical recto-verso manuscripts, very often the text written on the opposite page of the folio penetrates through the fiber of the paper, so that the texts on the two sides appear mixed. This is a very impairing damage that cannot be physically removed, and hinders both the work of philologists and palaeographers and the automatic analysis of linguistic contents. A procedure based on neural networks (NN) is proposed here to clean up the complex background of the manuscripts from this interference. We adopt a very simple shallow NN whose learning phase employs a training set generated from the data itself using a theoretical blending model that takes into account ink diffusion and saturation. By virtue of the parametric nature of the model, various levels of damage can be simulated in the training set, favoring a generalization capability of the NN. More explicitly, the network can be trained without the need for a large class of other similar manuscripts, but is still able, at least to some extent, to classify manuscripts with varying degrees of corruption. We compare the performance of this NN and other methods both qualitatively and quantitatively on a reference dataset and heavily damaged historical manuscripts.

3.1.18
Using AI to decode the behavioral responses of an insect to chemical stimuli: towards machine-animal computational technologies

E. Fazzari, F. Carrara, F. Falchi, C. Stefanini, D. Romano. International Journal of Machine Learning and Cybernetics, Springer [34]

Orthoptera are insects with excellent olfactory sense abilities due to their antennae richly equipped with receptors. This makes them interesting model organisms to be used as biosensors for environmental and agricultural monitoring. Herein, we investigated if the house cricket Acheta domesticus can be used to detect different chemical cues by examining the movements of their antennae and attempting to identify specific antennal displays associated to different chemical cues exposed (e.g., sucrose or ammonia powder). A neural network based on state-of-the-art techniques (i.e., SLEAP) for pose estimation was built to identify the proximal and distal ends of the antennae. The network was optimised via grid search, resulting in a mean Average Precision (mAP) of 83.74%. To classify the stimulus type, another network was employed to take in a series of keypoint sequences, and output the stimulus classification. To find the best one-dimensional convolutional and recurrent neural networks, a genetic algorithm-based optimisation method was used. These networks were validated with iterated K-fold validation, obtaining an average accuracy of 45.33% for the former and 44% for the latter. Notably, we published and introduced the first dataset on cricket recordings

that relate this animal's behaviour to chemical stimuli. Overall, this study proposes a novel and simple automated method that can be extended to other animals for the creation of Biohybrid Intelligent Sensing Systems (e.g., automated video-analysis of an organism's behaviour) to be exploited in various ecological scenarios.

3.1.19
Using semantic story maps to describe a territory beyond its map

V. Bartalesi, G. Coro, E. Lenzi, N. Pratelli, P. Pagano, F. Felici, M. Moretti, G. Brunori. Semantic Web, IOS Press. [9]

The paper presents the Story Map Building and Visualizing Tool (SMBVT) that allows users to create story maps within a collaborative environment and a usable Web interface. It is entirely open-source and published as a free-to-use solution. It uses Semantic Web technologies in the back-end system to represent stories through a reference ontology for representing narratives. It builds up a user-shared semantic knowledge base that automatically interconnects all stories and seamlessly enables collaborative story building. Finally, it operates within an Open-Science oriented e-Infrastructure, which enables data and information sharing within communities of narrators, and adds multi-tenancy, multi-user, security, and access-control facilities. SMBVT represents narratives as a network of spatiotemporal events related by semantic relations and standardizes the event descriptions by assigning internationalized resource identifiers (IRIs) to the event components, i.e., the entities that take part in the event (e.g., persons, objects, places, concepts). The tool automatically saves the collected knowledge as a Web Ontology Language (OWL) graph and openly publishes it as Linked Open Data. This feature allows connecting the story events to other knowledge bases. To evaluate and demonstrate our tool, we used it to describe the Apuan Alps territory in Tuscany (Italy). Based on a user-test evaluation, we assessed the tool's effectiveness at building story maps and the ability of the produced story to describe the territory beyond the map.

3.1.20
Using Semantic Web to create and explore an index of toponyms cited in Medieval geographical works

V. Bartalesi, N. Pratelli, E. Lenzi, P. Pontari. Journal on Computing and Cultural Heritage, Association for Computing Machinery [11]

Western thought in European history was mainly affected by the image of the world created during the Middle Ages and Renaissance. The most popular reason to travel during the Middle Ages was taking a pilgrimage. Jerusalem, Rome, and Santiago de Compostela were the most popular destinations. It is not surprising that a lot of works written by travellers as guides for pilgrims exist. By the beginning of the Renaissance, a more precise image of the world was defined thanks to the discovery of ancient geographical models, especially the work of Ptolemy. The three years (2020-2023) Italian National research project IMAGO - Index Medii Aevi Geographiae Operum - aims to provide a systematic overview of the medieval and renaissance Latin geographical literature using the Semantic Web technologies and the LOD paradigm. Indeed, until now, this literature has not been studied using digital methods. In particular, this paper presents how we formally represented the knowledge about the toponyms, or place names, in the IMAGO ontology. To maximise the interoperability, we developed the IMAGO ontology as an extension of two reference vocabularies: the CIDOC CRM and its extension FRBRoo, including its in-progress reformulation, LRMoo. Furthermore, we used Wikidata as reference knowledge base. As case study, we chose to represent the knowledge related to the toponyms cited by the Italian poet Dante Alighieri in his Latin works. We carried out a first experiment for visualising the knowledge about these toponyms on a map and in the form of tables and CSV files.

3.2 Books

In this section, we report books and monographies of which we acted as authors.

3.2.1
Learning to Quantify

A. Esuli., A. Fabris, A. Moreo, F. Sebastiani, Springer [32].

This open-access book provides an introduction and an overview of learning to quantify (a.k.a. "quantification"), i.e. the task of training estimators of class proportions in unlabeled data by means of supervised learning. In data science, learning to quantify is a task of its own related to classification yet different from it, since estimating class proportions by simply classifying all data and counting the labels assigned by the classifier is known to often return inaccurate ("biased") class proportion estimates. The book introduces learning to quantify by looking at the supervised learning methods that can be used to perform it, at the evaluation measures and evaluation protocols that should be used for evaluating the quality of the returned predictions, at the numerous fields of human activity in which the use of quantification techniques may provide improved results with respect to the naive use of classification techniques, and at advanced topics in quantification research. The book is suitable to researchers, data scientists, or PhD students, who want to come up to speed with the state of the art in learning to quantify, but also to researchers wishing to apply data science technologies to fields of human activity (e.g., the social sciences, political science, epidemiology, market research) which focus on aggregate ("macro") data rather than on individual ("micro") data.

3.2.2
Semantic Web - Introduction to Semantic Web languages

C. Meghini, V. Bartalesi. Simonelli Editore [53].

The Web makes a very large amount of information available to users in the form of documents. The Semantic Web is a fundamental extension of the web as it allows, in addition to documents, the sharing of data (including document metadata) in a standard format along with their semantic context expressed in a formal and shared language. Applications in documentary science, biology, cultural heritage and electronic commerce have already demonstrated the validity of this approach. This volume constitutes a gentle introduction to the technologies and languages of the semantic web, clearly illustrating the steps necessary to transform a product published on the web into a set of data that can be processed and reused across applications, users and communities. This is the second monograph of the ebook series "Digital Culture Notebooks" edited by the Laboratory of

Digital Culture of the University of Pisa (http://www.labcd.unipi.it) and published by Simonelli editore. The series houses short monographs on tools and research in the field of Digital Humanities which emerged from the work of teachers and students who collaborate with the Laboratory itself. It aims to support a wider dissemination of digital culture, understood as the field in which the humanities and some sectors of informatics interact and collaborate.

3.3 Proceedings

In this section, we report the paper we published in alphabetic order of the first author. Our works were presented, and published in the proceedings of the following conferences:

- **BUILD-IT** – BUILding a DIgital Twin: requirements, methods, and applications [8]
- **CBMI** – International Conference on Content-based Multimedia Indexing. [4]
- **ECAI** – 27TH European Conference on Artificial Intelligence [64]
- **ECIR** – European Conference on Information Retrieval. [63]
- **EVALITA** – Evaluation of NLP and Speech Tools for Italian. [59]
- **IberLEF@SEPLN** – Iberian Languages Evaluation Forum 2023 [38]
- **ICCSA** – International Conference on Computational Science and Its Applications. [61]
- **ICIAP** – International Conference on Theory and Practice of Digital Libraries 2023 [22, 54]
- **ICMR** – ACM International Conference on Multimedia Retrieval [2]
- **IMPROVE** – 3rd International Conference on Image Processing and Vision Engineering [23]
- **IRCDL** – International Conference on Theory and Practice of Digital Libraries 2023 [31]
- **Ital-IA** – Convegno Nazionale CINI sull'Intelligenza Artificiale [26, 46, 18, 20, 71]
- **MMM** – 29th International Conference on MultiMedia Modeling [3]
- **SEBD** – Proceedings of the 31st Symposium of Advanced Database Systems. [43, 45]
- **SIGIR** – 46th International ACM SIGIR Conference on Research and Development in Information Retrieval [55]
- **SISAP** – 16th International Conference on Similarity Search and Applications [19]
- **Text2Story** – The 6th International Workshop on Narrative Extraction from Texts: Text2Story 2023 [10, 6]
- **VISIGRAPP 2023** – 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications [37, 36]

3.3.1 A Graph Neural Network Approach for Evaluating Correctness of Groups of Duplicates

M. De Bonis, F. Minutella, F. Falchi, P. Manghi. IRCDL 2023 [31]

Unlabeled entity deduplication is a relevant task already studied in the recent literature. Most methods can be traced back to the following workflow: entity blocking phase, in-block pairwise comparisons between entities to draw similarity relations, closure of the resulting meshes to create groups of duplicate entities, and merging group entities to remove disambiguation. Such methods are effective but still not good enough whenever a very low false positive rate is required. In this paper, we present an approach for evaluating the correctness of "groups of duplicates", which can be used to measure the group's accuracy hence its likelihood of false-positiveness. Our novel approach is based on a Graph Neural Network that exploits and combines the concept of Graph Attention and Long Short Term Memory (LSTM). The accuracy of the proposed approach is verified in the context of Author Name Disambiguation applied to a curated dataset obtained as a subset of the OpenAIRE Graph that includes PubMed publications with at least one ORCID identifier.

3.3.2 AIMH at MULTI-Fake-DetectIVE: System Report

G. Puccetti, A. Esuli. EVALITA [59]

This report describes our contribution to the EVALITA 2023 shared task MULTI-Fake-DetectIVE which involves the classification of news including textual and visual components. To experiment on this task we focus on textual data augmentation, extending the Italian text and the Images available in the training set using machine translation models and image captioning ones. To train using different set of input features, we use different transformer encoders for each variant of text (Italian, English) and modality (Image). For Task 1, among the models we test, we find that using the Italian text together with its translation improves the model performance while the captions don´t provide any improvement. We test the same architecture also on Task 2 although in this case we achieve less satisfactory results.

3.3.3 AIMH Lab 2022 Activities for Healthcare

F. Carrara, L. Ciampi, M. Di Benedetto, F. Falchi, C. Gennaro, G. Amato. Ital-IA 2023 [18]

The application of Artificial Intelligence technologies in healthcare can enhance and optimize medical diagnosis, treatment, and patient care. Medical imaging, which involves Computer Vision to interpret and understand visual data, is one area of healthcare that shows great promise for AI, and it can lead to faster and more accurate diagnoses, such as detecting early signs of cancer or identifying abnormalities in the brain. This short paper provides an introduction to some of the activities of the Artificial Intelligence for Media and Humanities Laboratory of the ISTI-CNR that integrate AI and medical image analysis in healthcare. Specifically, the paper presents approaches that utilize 3D medical images to detect the behavior-variant of frontotemporal dementia, a neurodegenerative syndrome that can be diagnosed by analyzing brain scans. Furthermore, it

illustrates some Deep Learning-based techniques for localizing and counting biological structures in microscopy images, such as cells and perineuronal nets. Lastly, the paper presents a practical and cost-effective AI-based tool for multi-species pupillometry (mice and humans), which has been validated in various scenarios.

3.3.4
AIMH Lab 2022 Activities for Vision

L. Ciampi, G. Amato, P. Bolettieri, F. Carrara, M. Di Benedetto, F. Falchi, C. Gennaro, N. Messina, L. Vadicamo, C. Vairo. Ital-IA 2023 [20]

The explosion of smartphones and cameras has led to a vast production of multimedia data. Consequently, Artificial Intelligence-based tools for automatically understanding and exploring these data have recently gained much attention. In this short paper, we report some activities of the Artificial Intelligence for Media and Humanities (AIMH) laboratory of the ISTI-CNR, tackling some challenges in the field of Computer Vision for the automatic understanding of visual data and for novel interactive tools aimed at multimedia data exploration. Specifically, we provide innovative solutions based on Deep Learning techniques carrying out typical vision tasks such as object detection and visual counting, with particular emphasis on scenarios characterized by scarcity of labeled data needed for the supervised training and on environments with limited power resources imposing miniaturization of the models. Furthermore, we describe VISIONE, our large-scale video search system designed to search extensive multimedia databases in an interactive and user-friendly manner.

3.3.5
AIMH Lab approaches for Deepfake Detection

D.A. Coccomini, R. Caldelli, A. Esuli, F. Falchi, C. Gennaro, N. Messina, G. Amato. Ital-IA 2023 [26]

The creation of highly realistic media known as deepfakes has been facilitated by the rapid development of artificial intelligence technologies, including deep learning algorithms, in recent years. Concerns about the increasing ease of creation and credibility of deepfakes have then been growing more and more, prompting researchers around the world to concentrate their efforts on the field of deepfake detection. In this same context, researchers at ISTI-CNR's AIMH Lab have conducted numerous researches, investigations and proposals to make their own contribution to combating this worrying phenomenon. In this paper, we present the main work carried out in the field of deepfake detection and synthetic content detection, conducted by our researchers and in collaboration with external organizations.

3.3.6
AIMH Lab for a Susteinable Bio-Inspired AI

G. Lagani, F. Falchi, C. Gennaro, G. Amato. Ital-IA 2023 [46]

In this short paper, we report the activities of the Artificial Intelligence for Media and Humanities (AIMH) laboratory of the ISTI-CNR related to Sustainable AI. In particular, we discuss the problem of the environmental impact of AI research, and we discuss a research direction aimed at creating effective intelligent systems with a reduced ecological footprint. The proposal is based on bio-inspired learning, which takes inspiration from the biological processes underlying human intelligence in order to produce more energy-efficient AI systems. In fact, biological brains are able to perform complex computations, with a power consumption which is orders of magnitude smaller than that of traditional AI. The ability to control and replicate these biological processes reveals promising results towards the realization of sustainable AI.

3.3.7
An Optimized Pipeline for Image-Based Localization in Museums from Egocentric Images

N. Messina, F. Falchi, A. Furnari, C. Gennaro, G.M. Farinella. ICIAP 2023 [54]

With the increasing interest in augmented and virtual reality, visual localization is acquiring a key role in many downstream applications requiring a real-time estimate of the user location only from visual streams. In this paper, we propose an optimized hierarchical localization pipeline by specifically tackling cultural heritage sites with specific applications in museums. Specifically, we propose to enhance the Structure from Motion (SfM) pipeline for constructing the sparse 3D point cloud by a-priori filtering blurred and near-duplicated images. We also study an improved inference pipeline that merges similarity-based localization with geometric pose estimation to effectively mitigate the effect of strong outliers. We show that the proposed optimized pipeline obtains the lowest localization error on the challenging Bellomo dataset [11]. Our proposed approach keeps both build and inference times bounded, in turn enabling the deployment of this pipeline in real-world scenarios.

3.3.8
A Web Tool to Create and Visualise Semantic Story Maps

V. Bartalesi, E. Lenzi, N. Pratelli. Text2Story 2023 [10]

This paper presents the Story Map Building and Visualizing Tool (SMBVT), a sofware that allows users to create and visualise semantic story maps using a user-friendly web interface. The tool uses Wikidata as external reference knowledge base and exploits Semantic Web technologies in the back-end system to represent stories modelled on the Narrative ontology, a CRM-based vocabulary for representing narratives. SMBVT is entirely open-source and accessible afer free registration.

3.3.9
A Workflow for Developing Biohybrid Intelligent Sensing Systems

E. Fazzari, F. Carrara, F. Falchi, C. Stefanini, D. Romano. Ital-IA 2023 [35]

Animal are sometime exploited as biosensors for assessing the presence of volatile organic compounds (VOCs) in the environment by interpreting their stereotyped behavioral responses. However, current approaches are based on direct human observation to assess the changes in animal behaviors associated to specific environmental stimuli. We propose a general workflow based on artificial intelligence that use pose estimation and sequence classification technique to automate this process. This study also provides an example of its application studying the antennae movement of an insect (eg a cricket) in response to the presence of two chemical stimuli.

3.3.10
Creating and Visualising Semantic Story Maps

V. Bartalesi. Text2Story 2023 [6]

A narrative is a conceptual basis of collective human understanding. Humans use stories to represent characters' intentions, feelings and the attributes of objects, and events. A widely-held thesis in psychology to justify the centrality of narrative in human life is that humans make sense of reality by structuring events into narratives. Therefore, narratives are central to human activity in cultural, scientic, and social areas. Story maps are computer science realizations of narratives based on maps. They are online interactive maps enriched with text, pictures, videos, and other multimedia information, whose aim is to tell a story over a territory. This talk presents a semi-automatic workow that, using a CRM-based ontology and the Semantic Web technologies, produces semantic narratives in the form of story maps (and timelines as an alternative representation) from textual documents. An expert user rst assembles one territory-contextual document containing text and images. Then, automatic processes use natural language processing and Wikidata services to (i) extract entities and geospatial points of interest associated with the territory, (ii) assemble a logically-ordered sequence of events that constitute the narrative, enriched with entities and images, and (iii) openly publish online semantic story maps and an interoperable Linked Open Data-compliant knowledge base for event exploration and inter-story correlation analyses. Once the story maps are published, the users can review them through a user-friendly web tool. Overall, our workow complies with Open Science directives of open publication and multi-discipline support and is appropriate to convey "information going beyond the map" to scientists and the large public. As demonstrations, the talk will show workow-produced story maps to represent (i) 23 European rural areas across 16 countries, their value chains and territories, (ii) a Medieval journey, (iii) the history of the legends, biological investigations, and AI-based modelling for habitat discovery of the giant squid Architeuthis dux.

3.3.11
Crowdsim2: an open synthetic benchmark for object detectors

P. Foszner, A. Szczęsna, L. Ciampi, N. Messina, A. Cygan, B. Bizoń, M. Cogiel, D. Golba, E. Macioszek, M. Staniszewski. VISIGRAPP 2023 [36]

Data scarcity has become one of the main obstacles to developing supervised models based on Artificial Intelligence in Computer Vision. Indeed, Deep Learning-based models systematically struggle when applied in new scenarios never seen during training and may not be adequately tested in non-ordinary yet crucial real-world situations. This paper presents and publicly releases CrowdSim2, a new synthetic collection of images suitable for people and vehicle detection gathered from a simulator based on the Unity graphical engine. It consists of thousands of images gathered from various synthetic scenarios resembling the real world, where we varied some factors of interest, such as the weather conditions and the number of objects in the scenes. The labels are automatically collected and consist of bounding boxes that precisely localize objects belonging to the two object classes, leaving out humans from the annotation pipeline. We exploited this new benchmark as a testing ground for some state-of-the-art detectors, showing that our simulated scenarios can be a valuable tool for measuring their performances in a controlled environment.

3.3.12
Detecting Generated Text and Attributing Language Model Source with Fine-tuned Models and Semantic Understanding

M. Gambini, M. Avvenuti, F. Falchi, M. Tesconi, T. Fagni. IberLEF 2023 [38]

The improvements in natural language generation have led to the development of sophisticated language models capable of generating long and short texts that are incredibly difficult to distinguish from human-written ones. This remarkable generative capability has spread concerns about the potential misuse of such language models, such as the spread of misinformation, plagiarism, and causing disruption in the education system. Therefore, it is important to have automatic systems to distinguish generated texts from human-authored ones (deepfake text detection), as well as recognise the language model which produced a certain text for legal and security issues (generative language model attribution). The aim of the AuTexTification challenge was to address those two tasks on texts generated by state-of-the-art language models like text-davinci-003, being one of the first versions of the powerful ChatGPT. We proposed two detection models for both tasks: fine-tuned BERTweet and TriFuseNet, a three-branched network working on stylistic and contextual features. We achieved an F1 score of 0.616 (0.565) with fine-tuned BERTweet and 0.715 (0.499) with TriFuseNet on the deepfake text detection (generative language model attribution) task. Our results emphasize the significance of leveraging style, semantics, and context to distinguish machine-generated from human-written texts and identify the generative language model source.

3.3.13
Development of a realistic crowd simulation environment for fine-grained validation of people tracking methods

P. Foszner, A. Szczęsna, L. Ciampi, N. Messina, A. Cygan, B. Bizoń, M. Cogiel, D. Golba, E. Macioszek, M. Staniszewski. VISIGRAPP 2023 [37]

Generally, crowd datasets can be collected or generated from real or synthetic sources. Real data is generated by using infrastructure-based sensors (such as static cameras or other sensors). The use of simulation tools can significantly reduce the time required to generate scenario-specific crowd datasets, facilitate data-driven research, and next build functional machine learning models. The main goal of this work was to develop an extension of crowd simulation (named CrowdSim2) and prove its usability in the application of people-tracking algorithms. The simulator is developed using the very popular Unity 3D engine with particular emphasis on the aspects of realism in the environment, weather conditions, traffic, and the movement and models of individual agents. Finally, three methods of tracking were used to validate generated dataset: IOU-Tracker, Deep-Sort, and Deep-TAMA.

3.3.14
Mathematical models and neural networks for the description and the correction of typical distortions of historical manuscripts

P. Savino, A. Tonazzini. ICCSA 2023 [61]

Historical manuscripts are very often degraded by the seeping or transparency of the ink from the page opposite side. Suppressing the interfering text can be of great aid to philologists and paleographers who aim at interpreting the primary text, and nowadays also for the automatic analysis of the text. We formerly proposed a data model, which approximately describes this damage, to generate an artificial training set able to teach a shallow neural network how to classify pixels in clean or corrupted. This NN has proved to be effective in classifying manuscripts where the degradation can be also widely variable. In this paper, we modify the architecture of the NN to better account for ink saturation in text overlay areas, by including a specific class for these pixels. From the experiments, the improvement of the classification and then the restoration is significant.

3.3.15
MC-GTA: A Synthetic Benchmark for Multi-Camera Vehicle Tracking

L. Ciampi, N. Messina, G.E. Valenti, G. Amato, F. Falchi, C. Gennaro. ICIAP 2023 [22]

Multi-camera vehicle tracking (MCVT) aims to trace multiple vehicles among videos gathered from overlapping and non-overlapping city cameras. It is beneficial for city-scale traffic analysis, management, and security. However, developing MCVT systems is tricky, and the lack of data for training and testing computer vision deep learning-based solutions dampen their real-world applicability. Indeed, creating new annotated datasets is cumbersome as it requires great human effort and often faces privacy concerns. To alleviate this problem, we introduce MC-GTA - Multi Camera Grand Tracking Auto, a synthetic collection of images gathered from the virtual world provided by the highly realistic Grand Theft Auto 5 (GTA) video game. Our dataset has been recorded from several cameras recording urban scenes at various crossroads. The annotations, consisting of bounding boxes localizing the vehicles with associated unique IDs consistent across the video sources, have been automatically generated by interacting with the game engine. We provide a sample of our dataset in Figure 1. To assess this simulated scenario, we conduct a performance evaluation using an MCVT SOTA approach, showing that it can be a valuable benchmark that mitigates the need for real-world data. The MC-GTA dataset and the code for creating new ad-hoc custom scenarios are available at https://github.com/GaetanoV10/GT5-Vehicle-BB.

Figure 1. One of the considered scenarios of our MC-GTA dataset including three pairs of overlapping cameras located at three crossroads [22].

3.3.16
Recent Advancements on Bio-Inspired Hebbian Learning for Deep Neural Networks

G. Lagani. SEBD 2022 [43]

Deep learning is becoming more and more popular to extract information from multimedia data for indexing and query processing. In recent contributions, we have explored a biologically inspired strategy for Deep Neural Network (DNN) training, based on the Hebbian principle in neuroscience. We studied hybrid approaches in which unsupervised Hebbian learning was used for a pre-training stage, followed by supervised fine-tuning based on Stochastic Gradient Descent (SGD). The resulting semi-supervised strategy exhibited encouraging results on computer vision datasets, motivating further interest towards applications in the domain of large scale multimedia content based retrieval.

3.3.17
Scaling Bio-Inspired Neural Features to Real-World Image Retrieval Problems

G. Lagani. SEBD 2023 [45]

In the last decade, approaches in feature extraction for content-based multimedia retrieval exploited neural feature representations to describe complex data types such as images. In particular, recent approaches proposed to leverage bio-inspired learning solutions, which have the advantage to offer better generalization from fewer training samples. However, scaling these solutions to real-world datasets is a challenging problem. In my recent research, I proposed a possible approach to achieve such scalability, based on translating bio-inspired learning models into matrix multiplications, which can efficiently be executed on GPU. In this way, for the first time, I was able to validate bio-inspired methodologies on large-scale datasets such as ImageNet.

3.3.18
SegmentCodeList: Unsupervised Representation Learning for Human Skeleton Data Retrieval

J. Sedmidubský, F. Carrara, G. Amato. ECIR 2023 [63]

Recent progress in pose-estimation methods enables the extraction of sufficiently-precise 3D human skeleton data from ordinary videos, which offers great opportunities for a wide range of applications. However, such spatio-temporal data are typically extracted in the form of a continuous skeleton sequence without any information about semantic segmentation or annotation. To make the extracted data reusable for further processing, there is a need to access them based on their content. In this paper, we introduce a universal retrieval approach that compares any two skeleton sequences based on temporal order and similarities of their underlying segments. The similarity of segments is determined by their content-preserving low-dimensional code representation that is learned using the Variational AutoEncoder principle in an unsupervised way. The quality of the proposed representation is validated in retrieval and classification scenarios; our proposal outperforms the state-of-the-art approaches in effectiveness and reaches speed-ups up to 64x on common skeleton sequence datasets.

3.3.19
Social and hUman ceNtered XR

C. Vairo, M. Callieri, F. Carrara, P. Cignoni, M. Di Benedetto, C. Gennaro, D. Giorgi, G. Palma, L. Vadicamo, G. Amato. Ital-IA 2023 [71]

The Social and hUman ceNtered XR (SUN) project is focused on developing eXtended Reality (XR) solutions that integrate the physical and virtual world in a way that is convincing from a human and social perspective. In this paper, we outline the limitations that the SUN project aims to overcome, including the lack of scalable and cost-effective solutions for developing XR applications, limited solutions for mixing the virtual and physical environment, and barriers related to resource limitations of end-user devices. We also propose solutions to these limitations, including using artificial intelligence, computer vision, and sensor analysis to incrementally learn the visual and physical properties of real objects and generate convincing digital twins in the virtual environment. Additionally, the SUN project aims to provide wearable sensors and haptic interfaces to enhance natural interaction with the virtual environment and advanced solutions for user interaction. Finally, we describe three real-life scenarios in which we aim to demonstrate the proposed solutions.

3.3.20
Text-to-Motion Retrieval: Towards Joint Understanding of Human Motion Data and Natural Language

N. Messina, J. Sedmidubsky, F. Falchi, T. Rebok. SIGIR 2023 [55]

Due to recent advances in pose-estimation methods, human motion can be extracted from a common video in the form of 3D skeleton sequences. Despite wonderful application opportunities, effective and efficient content-based access to large volumes of such spatio-temporal skeleton data still remains a challenging problem. In this paper, we propose a novel content-based text-to-motion retrieval task, which aims at retrieving relevant motions based on a specified natural-language textual description. To define baselines for this uncharted task, we employ the BERT and CLIP language representations to encode the text modality and successful spatio-temporal models to encode the motion modality. We additionally introduce our transformer-based approach, called Motion Transformer (MoT), which employs divided space-time attention to effectively aggregate the different skeleton joints in space and time. Inspired by the recent progress in text-to-image/video matching, we experiment with two widely-adopted metric-learning loss functions. Finally, we set up a common evaluation protocol by defining qualitative metrics for assessing the quality of the retrieved motions, targeting the two recently-introduced KIT Motion-Language and HumanML3D datasets. The code for reproducing our results is available here: https://github.com/mesnico/text-to-motion-retrieval.

3.3.21
The Emotions of the Crowd: Learning Image Sentiment from Tweets via Cross-Modal Distillation

A. Serra, F. Carrara, M. Tesconi, F. Falchi. ECAI 2023 [64]

Trends and opinion mining in social media increasingly focus on novel interactions involving visual media, like images and short videos, in addition to text. In this work, we tackle the problem of visual sentiment analysis of social media images – specifically, the prediction of image sentiment polarity. While previous work relied on manually labeled training sets, we propose an automated approach for building sentiment polarity classifiers based on a cross-modal distillation paradigm; starting from scraped multimodal (text + images) data, we train a student model on the visual modality based on the outputs of a textual teacher model that analyses the sentiment of the corresponding textual modality. We applied our method to randomly collected images crawled from Twitter over three months and produced, after automatic cleaning, a weakly-labeled dataset of Ĩ.5 million images. Despite exploiting noisy labeled samples, our training pipeline produces classifiers showing strong generalization capabilities and outperforming the current state of the art on five manually labeled benchmarks for image sentiment polarity prediction.

3.3.22
Towards digital twins of territories through semantic story maps

V. Bartalesi, G. Coro, E. Lenzi, N. Pratelli, P. Pagano. BUILDIT 2023 [8]

Digital maps greatly support storytelling about territories, especially when enriched with data describing cultural, societal, and ecological aspects, conveying emotional messages that describe the territory as a whole. Story maps are interactive online digital narratives that can describe a territory beyond its map by enriching the map with text, pictures, videos, and other multimedia information. This paper outlines how online story maps can fill the gap between a map and a territory in narratives to create a digital twin of different territories as inter-connected semantic stories.

3.3.23
Traffic Scheduling in Non-Stationary Multipath Non-Terrestrial Networks: A Reinforcement Learning Approach

A. Machumilane, A. Gotta, P. Cassará, C. Gennaro, G. Amato. ICC 2023 [52]

In Non-Terrestrial Networks (NTNs), where LEO satellites and User Equipment (UE) move relative to each other, Line-of-Sight (LOS) tracking, and adapting to channel state variations due to endpoint movements are a major challenge. Therefore, continuous LOS estimation and channel impairment compensation are crucial for a UE to access a satellite and maintain connectivity. In this paper, we propose a Actor-Critic (AC)-Reinforcement Learning (RL) framework for traffic scheduling in NTN scenarios where the channel state is non-stationary due to the variability of LOS, which depends on the current satellite elevation. We deploy the framework as an agent in a Multi-Path Routing (MPR) scheme where the UE can access more than one satellite simultaneously to improve link reliability and throughput. We study how the agent schedules traffic on multiple satellite links by adopting the AC version of RL. The agent continuously trains based on variations in satellite elevation angles, handoffs, and relative LOS probabilities. We compare the agent retraining time with the satellite visibility intervals to investigate the effectiveness of the agent's learning rate. We carry out performance analysis considering the dense urban area of Chicago, where high-

rise buildings significantly affect the LOS. The simulation results show how the learning agent selects the scheduling policy when it is connected to a pair of satellites. The results also show that the retraining time of the learning agent is up to 0.1 times the satellite visibility time at certain elevations, which guarantees efficient use of satellite visibility.

3.3.24

Unsupervised Domain Adaptation for Video Violence Detection in the Wild

L. Ciampi, C. Santiago, J.P. Costeira, F. Falchi, C. Gennaro, G. Amato. IMPROVE 2023 [23]

Video violence detection is a subset of human action recognition aiming to detect violent behaviors in trimmed video clips. Current Computer Vision solutions based on Deep Learning approaches provide astonishing results. However, their success relies on large collections of labeled datasets for supervised learning to guarantee that they generalize well to diverse testing scenarios. Although plentiful annotated data may be available for some pre-specified domains, manual annotation is unfeasible for every ad-hoc target domain or task. As a result, in many real-world applications, there is a domain shift between the distributions of the train (source) and test (target) domains, causing a significant drop in performance at inference time. To tackle this problem, we propose an Unsupervised Domain Adaptation scheme for video violence detection based on single image classification that mitigates the domain gap between the two domains. We conduct experiments considering as the source labeled domain some datasets containing violent/non-violent clips in general contexts and, as the target domain, a collection of videos specific for detecting violent actions in public transport, showing that our proposed solution can improve the performance of the considered models. The considered scenario is illustrated in Figure 2.

Figure 2. The unsupervised domain adaptation approach for video violence detection we proposed to mitigate the domain gap that exists between a source domain (depicted on the left) and a target domain (depicted on the right). The source domain comprises three sets of annotated videos illustrating scenes with both violent and non-violent actions in diverse contexts. In contrast, the target domain consists of unlabeled video clips capturing violent and non-violent actions in public transport settings. [23].

3.3.25

Vec2Doc: Transforming Dense Vectors into Sparse Representations for Efficient Information Retrieval

F. Carrara, C. Gennaro, L. Vadicamo, G. Amato. SISAP 2023 [19]

The rapid development of deep learning and artificial intelligence has transformed our approach to solving scientific problems across various domains, including computer vision, natural language processing, and automatic content generation. Information retrieval (IR) has also experienced significant advancements, with natural language understanding and multimodal content analysis enabling accurate information retrieval. However, the widespread adoption of neural networks has also influenced the focus of IR problem-solving, which nowadays predominantly relies on evaluating the similarity of dense vectors derived from the latent spaces of deep neural networks. Nevertheless, the challenges of conducting similarity searches on large-scale databases with billions of vectors persist. Traditional IR approaches use inverted indices and vector space models, which work well with sparse vectors. In this paper, we propose Vec2Doc, a novel method that converts dense vectors into sparse integer vectors, allowing for the use of inverted indices. Preliminary experimental evaluation shows a promising solution for large-scale vector-based IR problems.

3.3.26

VISIONE: A Large-Scale Video Retrieval System with Advanced Search Functionalities

G. Amato, P. Bolettieri, F. Carrara, F. Falchi, C. Gennaro, N. Messina, L. Vadicamo, C. Vairo. ICMR 2023 [2]

VISIONE is a large-scale video retrieval system that integrates multiple search functionalities, including free text search, spatial color and object search, visual and semantic similarity search, and temporal search. The system leverages cutting-edge AI technology for visual analysis and advanced indexing techniques to ensure scalability. As demonstrated by its runner-up position in the 2023 Video Browser Showdown competition, VISIONE effectively integrates these capabilities to provide a comprehensive video retrieval solution. A system demo is available online, showcasing its capabilities on over 2300 hours of diverse video content (V3C1+V3C2 dataset) and 12 hours of highly redundant content (Marine dataset). The demo can be accessed at https://visione.isti.cnr.it/.

3.3.27

VISIONE at Video Browser Showdown 2023

G. Amato, P. Bolettieri, F. Carrara, F. Falchi, C. Gennaro, N. Messina, L. Vadicamo, C. Vairo. MMM 2023 [3]

In this paper, we present the fourth release of VISIONE, a tool for fast and effective video search on a large-scale dataset. It includes several search functionalities like text search, object and color-based search, semantic and visual similarity search, and temporal search. VISIONE uses ad-hoc textual encoding for indexing and searching video content, and it exploits a full-text search engine as search backend. In this new version of the system, we introduced some changes both to the current search techniques and to the user interface.

3.3.28
VISIONE for Newbies: An Easier-to-Use Video Retrieval System

G. Amato, P. Bolettieri, F. Carrara, F. Falchi, C. Gennaro, N. Messina, L. Vadicamo, C. Vairo CBMI 2023 [4]

This paper presents a revised version of the VISIONE video retrieval system, which offers a wide range of search functionalities, including free text search, spatial color and object search, visual and semantic similarity search, and temporal search. The system is designed to ensure scalability using advanced indexing techniques and effectiveness using cutting-edge Artificial Intelligence technology for visual content analysis. VISIONE was the runner-up in the 2023 Video Browser Showdown competition, demonstrating its comprehensive video retrieval capabilities. In this paper, we detail the improvements made to the search and browsing interface to enhance its usability for non-expert users. A demonstration video of our system with the restyled interface, showcasing its capabilities on over 2,300 hours of diverse video content, is available online at `https://youtu.be/srD3TCUkMSg`.

3.4 Editorials

In this section, we report journals, proceedings, and books for which we acted as editors.

3.4.1
Ital-IA 2023

F. Falchi, F. Giannotti, A. Monreale, C. Boldrini, S. Rinzivillo, S. Colantonio (eds): Proceedings of the Italia Intelligenza Artificiale - Thematic Workshops co-located with the 3rd CINI National Lab AIIS Conference on Artificial Intelligence (Ital IA 2023), Pisa, Italy, May 29-30, 2023. CEUR Workshop Proceedings 3486, CEUR-WS.org 2023. [1].

3.5 Preprints

In this section, we report the papers published only in preprint form on publicly accessible archives, in alphabetic order by first author.

3.5.1
Detecting Images Generated by Diffusers

D.A. Coccomini, A. Esuli, F. Falchi, C. Gennaro, G. Amato arXiv:2303.05275. [28]

This paper explores the task of detecting images generated by text-to-image diffusion models. To evaluate this, we consider images generated from captions in the MSCOCO and Wikimedia datasets using two state-of-the-art models: Stable Diffusion and GLIDE. Our experiments show that it is possible to detect the generated images using simple Multi-Layer Perceptrons (MLPs), starting from features extracted by CLIP, or traditional Convolutional Neural Networks (CNNs). We also observe that models trained on images generated by Stable Diffusion can detect images generated by GLIDE relatively well, however, the reverse is not true. Lastly, we find that incorporating the associated textual information with the images rarely leads to significant improvement in detection results but that the type of subject depicted in the image can have a significant impact on performance. This work provides insights into the feasibility of detecting generated images, and has implications for security and privacy concerns in real-world applications. The code to reproduce our results is available at: this https URL[2].

3.5.2
Preprocessing of recto-verso printed documents based on neural networks for text analysis.

P. Savino, A. Tonazzini
ISTI-CNR preprint. [62]

Among the many and varied damages affecting ancient documents, the penetration of ink from one side of the page to the other is one of the most frequent and invasive. In this work, we are interested in binarizing such degraded documents, for the application of OCR or other automatic text analysis tools, which can help philologists and palaeographers in text transcription. We previously proposed a data model that roughly describes this damage for front-to-back documents, and used it to generate an artificial training set that can teach a shallow neural network how to classify pixels on both sides into clean or corrupt. We show that this joint processing of the two sides of the document can significantly improve binarization and therefore OCR and other text analysis tasks, compared to the separate processing of the single sides, using the same information.

3.5.3
Scalable Bio-Inspired Training of Deep Neural Networks with Fasthebb

G. Lagani, F. Falchi, C. Gennaro, H. Fassold, G. Amato
doi:10.2139/ssrn.4566658 [49]

Recent work on sample efficient training of Deep Neural Networks (DNNs) proposed a semi-supervised methodology based on biologically inspired Hebbian learning, combined with traditional backprop-based training. Promising results were achieved on various computer vision benchmarks, in scenarios of scarce labeled data availability. However, current Hebbian learning solutions can hardly address large-scale scenarios due to their demanding computational cost. In order to tackle this limitation, this contribution develops a novel solution by reformulating Hebbian learning rules in terms of matrix multiplications, which can be executed more efficiently on GPU. We experimentally show that the proposed approach, named FastHebb, accelerates training speed up to 70 times, allowing us to gracefully scale Hebbian learning experiments on large datasets and network architectures such as ImageNet and VGG.

3.5.4
Synaptic Plasticity Models and Bio-Inspired Unsupervised Deep Learning: A Survey.

G. Lagani, F. Falchi, C. Gennaro, G. Amato
arXiv:2307.16236. [28]

Recently emerged technologies based on Deep Learning (DL) achieved outstanding results on a variety of tasks in the field of Artificial Intelligence (AI). However, these encounter several challenges related to robustness to adversarial inputs, ecological impact, and the necessity of huge amounts of training data. In response, researchers are focusing more and more interest on biologically grounded mechanisms, which are appealing due to the impressive capabilities exhibited by biological brains. This survey explores

[2] `https://github.com/davide-coccomini/Detecting-Images-Generated-by-Diffusers`

a range of these biologically inspired models of synaptic plasticity, their application in DL scenarios, and the connections with models of plasticity in Spiking Neural Networks (SNNs). Overall, Bio-Inspired Deep Learning (BIDL) represents an exciting research direction, aiming at advancing not only our current technologies but also our understanding of intelligence.

3.5.5
Spiking Neural Networks and Bio-Inspired Supervised Deep Learning: A Survey.

G. Lagani, F. Falchi, C. Gennaro, G. Amato

arXiv:2307.16235. [47]

Recently emerged technologies based on Deep Learning (DL) achieved outstanding results on a variety of tasks in the field of Artificial Intelligence (AI). However, these encounter several challenges related to robustness to adversarial inputs, ecological impact, and the necessity of huge amounts of training data. In response, researchers are focusing more and more interest on biologically grounded mechanisms, which are appealing due to the impressive capabilities exhibited by biological brains. This survey explores a range of these biologically inspired models of synaptic plasticity, their application in DL scenarios, and the connections with models of plasticity in Spiking Neural Networks (SNNs). Overall, Bio-Inspired Deep Learning (BIDL) represents an exciting research direction, aiming at advancing not only our current technologies but also our understanding of intelligence.

3.5.6
The devil is in the fine-grained details: Evaluating open-vocabulary object detectors for fine-grained understanding.

L. Bianchi, F. Carrara, N. Messina, C. Gennaro, F. Falchi

arXiv.2311.17518. [13]

Recent advancements in large vision-language models enabled visual object detection in open-vocabulary scenarios, where object classes are defined in free-text formats during inference. In this paper, we aim to probe the state-of-the-art methods for open-vocabulary object detection to determine to what extent they understand fine-grained properties of objects and their parts. To this end, we introduce an evaluation protocol based on dynamic vocabulary generation to test whether models detect, discern, and assign the correct fine-grained description to objects in the presence of hard-negative classes. We contribute with a benchmark suite of increasing difficulty and probing different properties like color, pattern, and material. We further enhance our investigation by evaluating several state-of-the-art open-vocabulary object detectors using the proposed protocol and find that most existing solutions, which shine in standard open-vocabulary benchmarks, struggle to accurately capture and distinguish finer object details. We conclude the paper by highlighting the limitations of current methodologies and exploring promising research directions to overcome the discovered drawbacks. Data and code are available at this https URL[3].

[3] https://github.com/lorebianchi98/FG-OVD

3.5.7
ViLMA: A Zero-Shot Benchmark for Linguistic and Temporal Grounding in Video-Language Models.

I. Kesen, A. Pedrotti, M. Dogan, M. Cafagna, E. Acikgoz, L. Parcalabescu, I. Calixto, A. Frank, A. Gatt, A. Erdem, E. Erdem

arXiv.2311.07022. [42]

With the ever-increasing popularity of pretrained Video-Language Models (VidLMs), there is a pressing need to develop robust evaluation methodologies that delve deeper into their visio-linguistic capabilities. To address this challenge, we present ViLMA (Video Language Model Assessment), a task-agnostic benchmark that places the assessment of fine-grained capabilities of these models on a firm footing. Task-based evaluations, while valuable, fail to capture the complexities and specific temporal aspects of moving images that VidLMs need to process. Through carefully curated counterfactuals, ViLMA offers a controlled evaluation suite that sheds light on the true potential of these models, as well as their performance gaps compared to human-level understanding. ViLMA also includes proficiency tests, which assess basic capabilities deemed essential to solving the main counterfactual tests. We show that current VidLMs' grounding abilities are no better than those of vision-language models which use static images. This is especially striking once the performance on proficiency tests is factored in. Our benchmark serves as a catalyst for future research on VidLMs, helping to highlight areas that still need to be explored.

4. Dissertations

4.1 PhD Thesis

4.1.1
Bio-Inspired Approaches for Deep Learning: From Spiking Neural Networks to Hebbian Plasticity

Gabriele Lagani, PhD in Computer Science, University of Pisa, 2023 [44].

In the past few years, Deep Neural Network (DNN) architectures have achieved outstanding results in several Artificial Intelligence (AI) domains. Even though DNNs draw inspiration from biology, the training methods based on the backpropagation algorithm (backprop) lack neuroscientific plausibility. The goal of this dissertation is to explore biologically-inspired solutions for the learning task. These are interesting because they can help to reproduce features of the human brain, for example, the ability to learn from a little experience. The investigation is divided into three phases: first, I explore a novel AI solution based on simulating neuronal biological cultures with a high level of detail, using biologically faithful Spiking Neural Network (SNN) models; second, I investigate neuroscientifically grounded Hebbian learning rules, applied to traditional DNNs in combination with backprop, using computer vision as a case study; third, I consider a more applicative perspective, using neural features derived from Hebbian learning for multimedia content retrieval tasks. I validate the proposed methods on different benchmarks, including MNIST, CIFAR, and ImageNet, obtaining promising results, especially in learning scenarios with scarce data. Moreover, to the best of my knowledge, for the first time, I am able to bring bio-inspired Hebbian methods to ImageNet scale, consisting of over 1

million images.

4.2 Master of Science Dissertations

4.2.1

Deep Learning methods for Visual Fish re-identification

Francesco Del Turco, Artificial Intelligence and Data Engineering, 2023. [69]. Advisors: M.G.C.A. Cimino, F. Falchi, C. Fabio, C. Bibbiani, C. Sangiacomo

Fishes are widely used in scientific research due to some particular characteristics, like the strong similarity of their genome with the humans' one or the very short life cycles; moreover, many of them provide simpler systems than other animals like mice and pigs for the study of complex processes. A typical task to be performed when studying different fish individuals during their life is re-identification, in which individuals of the same species under study must be recognized at different growth stages in order to understand the progress of the breeding, matching different weights and lengths obtained at different times. For this task, in some cases, it is enough to use GPS tags or microchips, but they're not a feasible solution for tiny fishes like the Danio rerio (zebrafish), which is one of the most used animals in scientific research, in particular considering their juvenile stages. The purpose of this study is, therefore, to exploit a metric learning approach applied through a Triplet Loss Network to build software able to perform, starting from an input image, the re-identification task in order to ease the whole study process by providing the most similar identities to the researcher and decrease the selection pool for re-identification. With this approach, we were able to obtain a 96% recall@10 over a group of 30 identities and 44.57% over 180 identities taken from our own zebrafish dataset.

4.2.2

Design and Development of Artificial Intelligence Techniques for Detecting People at Sea in aerial Images

Francesco Campilongo, Artificial Intelligence and Data Engineering, 2023. [16]. Advisors: M.G.C.A. Cimino, C. Gennaro, L. Ciampi, L. Vadicamo.

Modern camera-equipped Unmanned Aerial Vehicles (UAVs) can play an essential role in accelerating the localization and rescue of people. To this end, Artificial Intelligence (AI) techniques can be leveraged to automatically understand visual data acquired by drones. This has the potential to reduce search times and ultimately save human lives significantly. This thesis focuses on the field of object detection, a fundamental computer vision task, with a specific emphasis on its application to detecting people in open water environments from drone-view imagery. These kinds of scenarios are remarkably different compared to the generic ones for many reasons: different points of view from which the images are taken, different shapes of the objects seen from above, and changeable backgrounds due to various colors of water and weather conditions. To address these challenges, we conduct an experimental evaluation using various state-of-the-art deep neural networks over two datasets of images taken from UAVs at different altitudes, the Sea Drones See and the MOBDrone datasets. Specifically, we consider three popular generic object detectors (VarifocalNet, TOOD, and the latest version of YOLO (v8)), stressing their generalization capabilities and measuring their performances in various experimental settings. Through our experiments, we achieve notable improvements in detection performance. We started from a mean Average Precision (mAP) of 0.378 for the best performer, VarifocalNet pre-trained on a general-context dataset suitable for object detection, to a mAP of 0.647 using the same neural network fine-tuned over the specific drone-view datasets. Furthermore, we evaluate the performance also in terms of efficiency, showing that the latest version of YOLO demonstrates remarkably faster inference times, approximately nine times faster than VarifocalNet, while also consuming less memory. Our results highlight the significant potential of using AI techniques to improve search and rescue operations in drone-view imagery, paving the way to their applicability even directly onboard the UAVs.

4.2.3

Design and development of Artificial Intelligence algorithms for the analysis of EEG signals in Autism Spectrum Disorder

Giulio Federico, Artificial Intelligence and Data Engineering, 2023. [68]. Advisors: G. Claudio, A. Giuseppe, F. Fabrizio, L. Billeci

Autism Spectrum Disorders (ASD) are brain development alterations with onset in the first three years of life that lead to difficulties mainly in learning, social relationships and language, as well as often repetitive behaviors. The term "spectrum" underlines how autism never presents itself in the same form but varies according to the person and the time, making it difficult to understand the boundaries with the naked eye or with more traditional methods. The main reason of this work is to further investigate these boundaries, first analyzing which characteristics to date are typical in both anatomical and functional terms in both groups, and then proceeding in this same work not only with the power distribution analysis in the bands frequency drivers over the entire scalp but also doing a connectivity analysis on each of the regions since autism is more a connectivity issue than a power issue. Through a statistical analysis and subsequently through artificial intelligence algorithms we will try to understand which brain characteristics could better distinguish one group from another, in particular we will do clustering on the original features deriving from the application of algorithms to extract power and connectivity and subsequently transforming these features into others through neural networks to understand which representations best lend themselves to this discrimination. The greater understanding of these boundaries in the pre-childhood age will help to act in a timely manner to suggest any behavioral and/or pharmacological therapies capable of greatly reducing the effects of the disorder.

4.2.4

Design and Development of a System for Counting-Related Visual Question Answering

Tommaso Amarante, Artificial Intelligence and Data Engineering, 2023. [68]. Advisors: M.G.C.A. Cimino, F. Falchi, N. Messina, L. Ciampi

The challenging task of Visual Question Answering (VQA) requires a thorough comprehension of both visual content and natural language processing. Open- ended counting is a special case of VQA where the goal is to answer specific and possibly complex questions about the number of objects present in images. However,

even if counting is essential in many real-world applications, the de- velopment and assessment of counting algorithms within the VQA domain are limited by the scarcity of particular annotations for counting-related questions in existing VQA datasets. To fill this gap, in this dissertation, we present Object- CountingVQA, a brand-new dataset that focuses on the CountingVQA task. This new benchmark comprises more than 2000 images, with more than 5500 associated question-answer combinations. One feature of ObjectCountingVQA is that, in comparison to other benchmarks, it comprises more complex questions that include adjectives and spatial relationships; this provides a challenging setup for current VQA algorithms. We build those question-answer pairs automatically by using chatGPT, a popular artificial intelligence chatbot developed by OpenAI, starting from the structured data of the scene graphs of Visual Genome, a popular dataset for object detection and visual understanding. Then, we use Ground- ingDINO, a powerful open-ended object detector, to automatically validate the pairs and perform a first selection of good candidates. Finally, to ensure that all the questions and answers were accurate enough to be exploited as a reliable benchmark, we manually checked the generated annotations. To demonstrate the potential of our ObjectCountingVQA dataset, we conduct an experimental evaluation using a state-of-the-art VQA model, MOVIE+MCAN. According to our findings, this newly introduced benchmark presents fresh chal- lenges for the current VQA models, emphasizing the demand for specific counting methods and challenging benchmarks.

4.2.5
Design and development of cross-modal retrieval techniques based on transformer architectures

Lorenzo Bianchi, Artificial Intelligence and Data Engineering, 2023. [12]. Advisors: M.G.C.A. Cimino, C. Gennaro, F. Falchi, N. Messina

Human beings experience the world in a multi-modal manner. We elaborate thoughts combining pieces of information about objects we see, sounds we hear, tactile sensations we feel, odors we smell, and so on. In the last years, the progresses in deep learning techniques made machines more capable of understanding the meaning of texts, images, audio, and videos. By understanding hidden semantics connections between these different types of unstructured data, we can elaborate jointly on this information to approach multimodal problems, to resemble what humans do in everyday life. The work of this thesis will vert on the joint processing of images and natural language sentences. In particular, we will study the technologies behind cross-modal retrieval models between these two types of information. We will exploit new combinations of technologies and techniques to improve the results obtained by ALADIN, a cross-modal image-text retrieval model which reaches performances near the competitors, the large Vision-Language Transformers while being 90 times faster. By introducing some modifications to the visual pipeline in the backbone of the architecture we were able to improve the model's performance. In particular, we improved the results presented in the original paper regarding the recall@k metric for the alignment head, the head of the model which aligns in a fine-grained manner the images and the texts representation. On the MS COCO dataset, we improved the rsum, the sum of the recall@k for the chosen k values (1,5 and 10), by 0.8 points on the 1K test set and by 3.5 points on the 5K test set. The code to reproduce our results is available at https://github.com/lorebianchi98/ALADIN-2.0.

4.2.6
Investigating the problem of distinguishing between native and non-native speakers by their typed texts

Monica De Caro, Digital Humanities, 2023 [17]. Advisors: A. Esuli

This thesis focuses on the differences between native and non-native speakers that emerge by the comparison of their typed texts. The comparison is carried out by mean of machine learning, applying it to datasets composed of texts in English from both native and non-native speakers. In this context, we investigate the impact of phonological features, as they may have a positive impact on this task because they mirror some behaviours that are typical of native speakers. This study is also about some issues concerning the actual need to distinguish between natives and non-natives in contexts that are far from linguistics.

5. Resources

In this section, we report contributions of AIMH having to do with the creation of datasets (Section 5.1), the publication of code (Section 5.2), and the design of shared tasks (Section **??**)

5.1 Datasets

5.1.1
Crowd simulation (CrowdSim2) for tracking and object detection

P. Foszner, A. Szczesna, A. Cygan, B. Bizon, M. Cogiel, D. Golba, L. Ciampi, N. Messina E. Macioszek, M. Staniszewski [66]

CrowdSim2, developed in Unity, serves as a crowd simulation tool specifically designed to generate extensive synthetic data. This generated data from crowd simulation facilitates the evaluation of diverse methods, particularly in terms of tracking multiple individuals and detecting objects, with a focus on pedestrians and cars. The dataset is freely available at `https://zenodo.org/record/7262220`.

5.1.2
FG-OVD: Evaluating open-vocabulary object detectors for fine-grained understanding

L. Bianchi, F. Carrara, N. Messina, C. Gennaro, F. Falchi. [13]

We provide a benchmark suite for evaluating open-vocabulary object detectors for fine-grained understanding of objects, their parts, and their attributes under different difficulty and attribute settings. Project page: `https://lorebianchi98.github.io/FG-OVD/`.

5.1.3
MC-GTA: A Synthetic Benchmark for Multi-Camera Vehicle Tracking

L. Ciampi, N. Messina, G.E. Valenti, G. Amato, F. Falchi, C. Gennaro [22]

MC-GTA - Multi Camera Grand Tracking Auto is a synthetic collection of images gathered from the virtual world provided by the highly realistic Grand Theft Auto 5 (GTA) video game. This

dataset has been recorded from several cameras recording urban scenes at various crossroads and it is suitable for the multi-camera vehicle tracking task. The annotations, consisting of bounding boxes localizing the vehicles with associated unique IDs consistent across the video sources, have been automatically generated by interacting with the game engine.

5.1.4
Pest Sticky Traps: a dataset for Whitefly Pest Population Density Estimation in Chromotropic Sticky Traps

L. Ciampi, V. Zeni, L. Incrocci, A. Canale, G. Benelli, F. Falchi, G. Amato, S. Chessa [25].

The Pest Sticky Traps (PST) dataset is a collection of yellow chromotropic sticky trap pictures specifically designed for training/testing deep learning models to automatically count insects and estimate pest populations. Images were manually annotated by some experts of the Department of Agriculture, Food and Environment of the University of Pisa (Italy) by putting a dot over the centroids of each identified insect. Specifically, we labeled insects as belonging to the category "whitefly" considering two different species, i.e., the sweet potato whitefly (Bemisia tabaci) (Gennadius) and the greenhouse whitefly (Trialeurodes vaporariorum) (Westwood). The dataset is freely available at `https://zenodo.org/records/7801239`.

5.2 Code

5.2.1
A deep learning-based pipeline for whitefly pest abundance estimation on chromotropic sticky traps

L. Ciampi, V. Zeni, L. Incrocci, A. Canale, G. Benelli, F. Falchi, G. Amato, S. Chessa. [24]

`https://ciampluca.github.io/sticky_trap_pest_counting/`

Code for replicating the experiments in [24].

5.2.2
Assess visual sentiment polarity in images.

A. Serra, F. Carrara, M. Tesconi, F. Falchi. [64]

`https://github.com/fabiocarrara/cross-modal-visual-sentiment-analysis`

`https://fabiocarrara.github.io/cross-modal-visual-sentiment-analysis/`*Pre-trained models to perform visual sentiment analysis and code for replicating the experiments in [64]. Project webpage:* `https://fabiocarrara.github.io/cross-modal-visual-sentiment-analysis/`

5.2.3
Detecting Images generated by Diffusers

D.A. Coccomini, A. Esuli, F. Falchi, C. Gennaro, G. Amato arXiv:2303.05275. [28]

`https://github.com/davide-coccomini/Detecting-Images-Generated-by-Diffusers`

Code for replicating the experiments in [28].

5.2.4
MC-GTA: A Synthetic Benchmark for Multi-Camera Vehicle Tracking

L. Ciampi, N. Messina, G.E. Valenti, G. Amato, F. Falchi, C. Gennaro. [22]

`https://github.com/GaetanoV10/GT5-Vehicle-BB`

Code for replicating the experiments in [22].

6. Services

6.1 Services in conferences

In this section, we report the conference in which we were involved in the organization.

6.1.1 Ital-IA 2023

Fabrizio Falchi was General Co-Chair, and Fabio Carrara was Web-Chair of the 3rd of the National Conference on Artificial Intelligence organized by the National Interuniversity Consortium for Informatics (CINI), Ital-IA 2023, Pisa, Italy, May 29-30, 2023.

7. Awards

7.1 International Competitions

7.1.1 Video Browser Showdown

The VISIONE content-based video retrieval system obtained second place at the Video Browser Showdown, The Video Retrieval Competition, with the approach described in [3].

7.2 Best Paper Awards

7.2.1 SIGIR

The paper "Messina, Nicola, et al. "Text-to-Motion Retrieval: Towards Joint Understanding of Human Motion Data and Natural Language." [55] was awarded with the Best Short-paper Award (honorable mention) at the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2023. Taipei, Taiwan, July 23-27, 2023.

7.2.2 IMPROVE

The paper "Unsupervised Domain Adaptation for Video Violence Detection in the Wild", L. Ciampi, C. Santiago, J.P. Costeira, F. Falchi, C. Gennaro, G. Amato [23] won the Best Paper Award at the 3rd International Conference on Image Processing and Vision Engineering, IMPROVE 2023, Prague, Czech Republic, April 21-23, 2023.

7.3 National Awards

7.3.1 ISTI Young Research Awards

The esearchers of the AIMH Lab that won the ISTI Young Research Awards "Matteo Delle Piane" in 2023 are:

- Luca Ciampi, in the Advanced category
- Gabriele Lagani, in the Beginner category

7.3.2 ISTI Grant for Young Mobility

The researchers of the AIMH Lab that won an ISTI Grant for Young Mobility in 2023 are:

- Luca Ciampi, second call

References

[1] Proceedings of the italia intelligenza artificiale - thematic workshops co-located with the 3rd CINI national lab AIIS conference on artificial intelligence (ital IA 2023), pisa, italy, may 29-30, 2023. 3486, 2023.

[2] Giuseppe Amato, Paolo Bolettieri, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro, Nicola Messina, Lucia Vadicamo, and Claudio Vairo. VISIONE: A large-scale video retrieval system with advanced search functionalities. In Ioannis Kompatsiaris, Jiebo Luo, Nicu Sebe, Angela Yao, Vasileios Mazaris, Symeon Papadopoulos, Adrian Popescu, and Zi Helen Huang, editors, *Proceedings of the 2023 ACM International Conference on Multimedia Retrieval, ICMR 2023, Thessaloniki, Greece, June 12-15, 2023*, pages 649–653. ACM, 2023.

[3] Giuseppe Amato, Paolo Bolettieri, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro, Nicola Messina, Lucia Vadicamo, and Claudio Vairo. VISIONE at video browser showdown 2023. In Duc-Tien Dang-Nguyen, Cathal Gurrin, Martha A. Larson, Alan F. Smeaton, Stevan Rudinac, Minh-Son Dao, Christoph Trattner, and Phoebe Chen, editors, *MultiMedia Modeling - 29th International Conference, MMM 2023, Bergen, Norway, January 9-12, 2023, Proceedings, Part I*, volume 13833 of *Lecture Notes in Computer Science*, pages 615–621. Springer, 2023.

[4] Giuseppe Amato, Paolo Bolettieri, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro, Nicola Messina, Lucia Vadicamo, and Claudio Vairo. Visione for newbies: An easier-to-use video retrieval system. In *Proceedings of the 20th International Conference on Content-Based Multimedia Indexing*, CBMI '23, page 158–162, New York, NY, USA, 2023. Association for Computing Machinery.

[5] Francesco Banterle, Alessandro Artusi, Alejandro Moreo, Fabio Carrara, and Paolo Cignoni. Nor-vdpnet++: Real-time no-reference image quality metrics. *IEEE Access*, 11:34544–34553, 2023.

[6] Valentina Bartalesi. Creating and visualising semantic story maps. In Aachen CEUR-WS.org, editor, *Proceedings of Text2Story — Sixth Workshop on Narrative Extraction From Texts (Text2Story 2023) held in conjunction with the 45th European Conference on Information Retrieval (ECIR 2023) Dublin, Ireland, April 2, 2023.*, volume 3370 of *CEUR workshop proceedings*, pages 3–4, 2023.

[7] Valentina Bartalesi, Gianpaolo Coro, Emanuele Lenzi, Pasquale Pagano, and Nicolò Pratelli. From unstructured texts to semantic story maps. *International Journal of Digital Earth*, 16:pp 234–250, 2023.

[8] Valentina Bartalesi, Gianpaolo Coro, Emanuele Lenzi, Nicolò Pratelli, and Pasquale Pagano. Towards digital twins of territories through semantic story maps. In *BUILD-IT 2023 Workshop - BUILding a Digital Twin: requirements, methods, and applications, Rome, Italy*, pages 41–45, 2023.

[9] Valentina Bartalesi, Gianpaolo Coro, Emanuele Lenzi, Nicolò Pratelli, Pagano Pasquale, Francesco Felici, Michele Moretti, and Gianluca Brunori. Using semantic story maps to describe a territory beyond its map. *Semantic web*, 14:pp. 1255–1272, 2023.

[10] Valentina Bartalesi, Emanuele Lenzi, and Nicolò Pratelli. A web tool to create and visualise semantic story maps. In Aachen CEUR-WS.org, editor, *Proceedings of Text2Story — Sixth Workshop on Narrative Extraction From Texts (Text2Story 2023) held in conjunction with the 45th European Conference on Information Retrieval (ECIR 2023) Dublin, Ireland, April 2, 2023.*, volume 3370 of *CEUR workshop proceedings*, pages 163–169, 2023.

[11] Valentina Bartalesi, Nicolò Pratelli, Emanuele Lenzi, and Paolo Pontari. Using semantic story maps to describe a territory beyond its map. *Journal on Computing and Cultural Heritage*, 16:pp 1–18, 2023.

[12] Lorenzo Bianchi. Design and development of cross-modal retrieval techniques based on transformer architectures. Master's thesis, M.Sc. in Artificial Intelligence and Data Engineering, University of Pisa, 2023.

[13] Lorenzo Bianchi, Fabio Carrara, Nicola Messina, Claudio Gennaro, and Fabrizio Falchi. The devil is in the fine-grained details: Evaluating open-vocabulary object detectors for fine-grained understanding, 2023.

[14] Michele De Bonis, Fabrizio Falchi, and Paolo Manghi. Graph-based methods for author name disambiguation: a survey. *PeerJ Comput. Sci.*, 9:e1536, 2023.

[15] Mirko Bunse, Alejandro Moreo, Fabrizio Sebastiani, and Martin Senz. Regularization-based methods for ordinal quantification. *CoRR*, abs/2310.09210, 2023.

[16] Francesco Campilongo. Design and development of artificial intelligence techniques for detecting people at sea in aerial images. Master's thesis, M.Sc. in Artificial Intelligence and Data Engineering, University of Pisa, 2023.

[17] Monica De Caro. Investigating the problem of distinguishing between native and non-native speakers by their typed texts. Master's thesis, M.Sc. in Digital Humanities, University of Pisa, 2023.

[18] Fabio Carrara, Luca Ciampi, Marco Di Benedetto, Fabrizio Falchi, Claudio Gennaro, and Giuseppe Amato. AIMH lab 2022 activities for healthcare. In Fabrizio Falchi, Fosca Giannotti, Anna Monreale, Chiara Boldrini, Salvatore Rinzivillo, and Sara Colantonio, editors, *Proceedings of the Italia Intelligenza Artificiale - Thematic Workshops co-located with the 3rd CINI National Lab AIIS Conference on Artificial Intelligence (Ital IA 2023), Pisa, Italy, May 29-30, 2023*, volume 3486 of *CEUR Workshop Proceedings*, pages 128–133. CEUR-WS.org, 2023.

[19] Fabio Carrara, Claudio Gennaro, Lucia Vadicamo, and Giuseppe Amato. Vec2doc: Transforming dense vectors into sparse representations for efficient information retrieval. In *International Conference on Similarity Search and Applications*, pages 215–222. Springer, 2023.

[20] Luca Ciampi, Giuseppe Amato, Paolo Bolettieri, Fabio Carrara, Marco Di Benedetto, Fabrizio Falchi, Claudio Gennaro, Nicola Messina, Lucia Vadicamo, and Claudio Vairo. AIMH lab 2022 activities for vision. In Fabrizio Falchi, Fosca Giannotti, Anna Monreale, Chiara Boldrini, Salvatore Rinzivillo, and Sara Colantonio, editors, *Proceedings of the Italia Intelligenza Artificiale - Thematic Workshops co-located with the 3rd CINI National Lab AIIS Conference on Artificial Intelligence (Ital IA 2023), Pisa, Italy, May 29-30, 2023*, volume 3486 of *CEUR Workshop Proceedings*, pages 538–543. CEUR-WS.org, 2023.

[21] Luca Ciampi, Fabio Carrara, Valentino Totaro, Raffaele Mazziotti, Leonardo Lupori, Carlos Santiago, Giuseppe Amato, Tommaso Pizzorusso, and Claudio Gennaro. Learning to count biological structures with raters' uncertainty. *Medical Image Analysis*, 80:102500, aug 2022.

[22] Luca Ciampi, Nicola Messina, Gaetano Emanuele Valenti, Giuseppe Amato, Fabrizio Falchi, and Claudio Gennaro. Mc-gta: A synthetic benchmark for multi-camera vehicle tracking. In Gian Luca Foresti, Andrea Fusiello, and Edwin Hancock, editors, *Image Analysis and Processing – ICIAP 2023*, pages 316–327, Cham, 2023. Springer Nature Switzerland.

[23] Luca Ciampi, Carlos Santiago, João Paulo Costeira, Fabrizio Falchi, Claudio Gennaro, and Giuseppe Amato. Unsupervised domain adaptation for video violence detection in the wild. In Francisco H. Imai, Cosimo Distante, and Sebastiano Battiato, editors, *Proceedings of the 3rd International Conference on Image Processing and Vision Engineering, IMPROVE 2023, Prague, Czech Republic, April 21-23, 2023*, pages 37–46. SCITEPRESS, 2023.

[24] Luca Ciampi, Valeria Zeni, Luca Incrocci, Angelo Canale, Giovanni Benelli, Fabrizio Falchi, Giuseppe Amato, and Stefano Chessa. A deep learning-based pipeline for whitefly pest abundance estimation on chromotropic sticky traps. *Ecological Informatics*, 78:102384, 2023.

[25] Luca Ciampi, Valeria Zeni, Luca Incrocci, Angelo Canale, Giovanni Benelli, Fabrizio Falchi, Giuseppe Amato, and Stefano Chessa. Pest Sticky Traps: a dataset for Whitefly Pest Population Density Estimation in Chromotropic Sticky Traps, April 2023.

[26] Davide Alessandro Coccomini, Roberto Caldelli, Andrea Esuli, Fabrizio Falchi, Claudio Gennaro, Nicola Messina, and Giuseppe Amato. AIMH lab approaches for deepfake detection. In Fabrizio Falchi, Fosca Giannotti, Anna Monreale, Chiara Boldrini, Salvatore Rinzivillo, and Sara Colantonio, editors, *Proceedings of the Italia Intelligenza Artificiale - Thematic Workshops co-located with the 3rd CINI National Lab AIIS Conference on Artificial Intelligence (Ital IA 2023), Pisa, Italy, May 29-30, 2023*, volume 3486 of *CEUR Workshop Proceedings*, pages 432–436. CEUR-WS.org, 2023.

[27] Davide Alessandro Coccomini, Roberto Caldelli, Fabrizio Falchi, and Claudio Gennaro. On the generalization of deep learning models in video deepfake detection. *Journal of Imaging*, 9(5), 2023.

[28] Davide Alessandro Coccomini, Andrea Esuli, Fabrizio Falchi, Claudio Gennaro, and Giuseppe Amato. Detecting images generated by diffusers, 2023.

[29] Silvia Corbara, Alejandro Moreo, and Fabrizio Sebastiani. Same or different? diff-vectors for authorship analysis. *CoRR*, abs/2301.09862, 2023.

[30] Silvia Corbara, Alejandro Moreo, and Fabrizio Sebastiani. Syllabic quantity patterns as rhythmic features for latin authorship attribution. *J. Assoc. Inf. Sci. Technol.*, 74(1):128–141, 2023.

[31] Michele De Bonis, Filippo Minutella, Fabrizio Falchi, and Paolo Manghi. A graph neural network approach for evaluating correctness of groups of duplicates. In Omar Alonso, Helena Cousijn, Gianmaria Silvello, Mónica Marrero, Carla Teixeira Lopes, and Stefano Marchesin, editors, *Linking Theory and Practice of Digital Libraries*, pages 207–219, Cham, 2023. Springer Nature Switzerland.

[32] Andrea Esuli, Alessandro Fabris, Alejandro Moreo, and Fabrizio Sebastiani. *Learning to Quantify*, volume 47 of *The Information Retrieval Series*. Springer, 2023.

[33] Alessandro Fabris, Andrea Esuli, Alejandro Moreo, and Fabrizio Sebastiani. Measuring fairness under unawareness of sensitive attributes: A quantification-based approach. *J. Artif. Intell. Res.*, 76:1117–1180, 2023.

[34] Edoardo Fazzari, Fabio Carrara, Fabrizio Falchi, Cesare Stefanini, and Donato Romano. Using ai to decode the behavioral responses of an insect to chemical stimuli: towards machine-animal computational technologies. *International Journal of Machine Learning and Cybernetics*, pages 1–10, 2023.

[35] Edoardo Fazzari, Fabio Carrara, Fabrizio Falchi, Cesare Stefanini, and Donato Romano. A workflow for developing biohybrid intelligent sensing systems. In *Proceedings of the Italia Intelligenza Artificiale - Thematic Workshops co-located with the 3rd CINI National Lab AIIS Conference on Artificial Intelligence (Ital IA 2023), Pisa, Italy, May 29-30, 2023*, volume 3486 of *CEUR Workshop Proceedings*, pages 555–560. CEUR-WS.org, 2023.

[36] Paweł Foszner, Agnieszka Szczęsna, Luca Ciampi, Nicola Messina, Adam Cygan, Bartosz Bizoń, Michał Cogiel, Dominik Golba, Elżbieta Macioszek, and Michał Staniszewski. Crowdsim2: An open synthetic benchmark

for object detectors. In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. SCITEPRESS - Science and Technology Publications, 2023.

[37] Paweł Foszner, Agnieszka Szczęsna, Luca Ciampi, Nicola Messina, Adam Cygan, Bartosz Bizoń, Michał Cogiel, Dominik Golba, Elżbieta Macioszek, and Michał Staniszewski. Development of a realistic crowd simulation environment for fine-grained validation of people tracking methods. In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. SCITEPRESS - Science and Technology Publications, 2023.

[38] Margherita Gambini, Marco Avvenuti, Fabrizio Falchi, Maurizio Tesconi, and Tiziano Fagni. Detecting generated text and attributing language model source with fine-tuned models and semantic understanding. In Manuel Montes-y-Gómez, Francisco Rangel, Salud María Jiménez Zafra, Marco Casavantes, Begoña Altuna, Miguel Ángel Álvarez Carmona, Gemma Bel-Enguix, Luis Chiruzzo, Iker de la Iglesia, Hugo Jair Escalante, Miguel Ángel García Cumbreras, José Antonio García-Díaz, José Ángel González Barba, Roberto Labadie Tamayo, Salvador Lima, Pablo Moral, Flor Miriam Plaza del Arco, and Rafael Valencia-García, editors, *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023) co-located with the Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), Jaén, Spain, September 26, 2023*, volume 3496 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2023.

[39] Jose L Salazar González, Juan A Álvarez-García, Fernando J Rendón-Segador, and Fabio Carrara. Conditioned cooperative training for semi-supervised weapon detection. *Neural networks*, 167:489–501, 2023.

[40] Pablo González, Alejandro Moreo, and Fabrizio Sebastiani. Binary quantification and dataset shift: An experimental investigation. *CoRR*, abs/2310.04565, 2023.

[41] Muhammad Hanif, Anna Tonazzini, Syed Fawad Hussain, Usman Habib, Emanuele Salerno, Pasquale Savino, and Zahid Halim. Blind bleed-through removal in color ancient manuscripts. *Multimedia Tools and Applications*, 82(8), 2023.

[42] Ilker Kesen, Andrea Pedrotti, Mustafa Dogan, Michele Cafagna, Emre Can Acikgoz, Letitia Parcalabescu, Iacer Calixto, Anette Frank, Albert Gatt, Aykut Erdem, and Erkut Erdem. Vilma: A zero-shot benchmark for linguistic and temporal grounding in video-language models, 2023.

[43] Gabriele Lagani. Recent advancements on bio-inspired hebbian learning for deep neural networks. In Giuseppe Amato, Valentina Bartalesi, Davis Bianchini, Claudio Gennaro, and Riccardo Torlone, editors, *Proceedings of the 30th Italian Symposium on Advanced Database Systems, Tirrenia (PI), Italy, June 19-22, 2022*, volume 3194, pages 610–615. CEUR-WS, 2022.

[44] Gabriele Lagani. *Bio-Inspired Approaches for Deep Learning: From Spiking Neural Networks to Hebbian Plasticity*. PhD thesis, Dottorato in Informatica, University of Pisa, Italy, 2023.

[45] Gabriele Lagani. Scaling bio-inspired neural features to real-world image retrieval problems. In Diego Calvanese, Claudia Diamantini, Guglielmo Faggioli, Nicola Ferro, Stefano Marchesin, Silvello Gianmaria, and Letizia Tanca, editors, *Proceedings of the 31st Symposium of Advanced Database Systems, Galzingano Terme, Italy, July 2nd to 5th, 2023*, volume 3478, pages 711–717. CEUR-WS, 2023.

[46] Gabriele Lagani, Fabrizio Falchi, Claudio Gennaro, and Giuseppe Amato. AIMH lab for a susteinable bio-inspired AI. In Fabrizio Falchi, Fosca Giannotti, Anna Monreale, Chiara Boldrini, Salvatore Rinzivillo, and Sara Colantonio, editors, *Proceedings of the Italia Intelligenza Artificiale - Thematic Workshops co-located with the 3rd CINI National Lab AIIS Conference on Artificial Intelligence (Ital IA 2023), Pisa, Italy, May 29-30, 2023*, volume 3486 of *CEUR Workshop Proceedings*, pages 575–584. CEUR-WS.org, 2023.

[47] Gabriele Lagani, Fabrizio Falchi, Claudio Gennaro, and Giuseppe Amato. Spiking neural networks and bio-inspired supervised deep learning: A survey, 2023.

[48] Gabriele Lagani, Fabrizio Falchi, Claudio Gennaro, and Giuseppe Amato. Synaptic plasticity models and bio-inspired unsupervised deep learning: A survey, 2023.

[49] Gabriele Lagani, Fabrizio Falchi, Claudio Gennaro, Hannes Fassold, and Giuseppe Amato. Scalable bio-inspired training of deep neural networks with fasthebb, 2023.

[50] Jakub Lokoč, Stelios Andreadis, Werner Bailer, Aaron Duane, Cathal Gurrin, Zhixin Ma, Nicola Messina, Thao-Nhu Nguyen, Ladislav Peška, Luca Rossetto, et al. Interactive video retrieval in the age of effective joint embedding deep models: lessons from the 11th vbs. *Multimedia Systems*, 29(6):3481–3504, 2023.

[51] Leonardo Lupori, Valentino Totaro, Sara Cornuti, Luca Ciampi, Fabio Carrara, Edda Grilli, Aurelia Viglione, Francesca Tozzi, Elena Putignano, Raffaele Mazziotti, Giuseppe Amato, Claudio Gennaro, Paola Tognini, and Tommaso Pizzorusso. A comprehensive atlas of perineuronal net distribution and colocalization with parvalbumin in the adult mouse brain. *Cell Reports*, 42(7), 2023. All Open Access, Gold Open Access, Green Open Access.

[52] Achilles Machumilane, Alberto Gotta, Pietro Cassará, Claudio Gennaro, and Giuseppe Amato. Traffic scheduling in non-stationary multipath non-terrestrial networks:

A reinforcement learning approach. In *ICC 2023 - IEEE International Conference on Communications*, pages 4094–4099, 2023.

[53] Carlo Meghini and Valentina Bartalesi. *Semantic Web - Introduction to Semantic Web languages*. Simonelli Editore, Milano (Italia), 2023.

[54] Nicola Messina, Fabrizio Falchi, Antonino Furnari, Claudio Gennaro, and Giovanni Maria Farinella. An optimized pipeline for image-based localization in museums from egocentric images. In Gian Luca Foresti, Andrea Fusiello, and Edwin Hancock, editors, *Image Analysis and Processing – ICIAP 2023*, pages 512–524, Cham, 2023. Springer Nature Switzerland.

[55] Nicola Messina, Jan Sedmidubský, Fabrizio Falchi, and Tomás Rebok. Text-to-motion retrieval: Towards joint understanding of human motion data and natural language. In Hsin-Hsi Chen, Wei-Jou (Edward) Duh, Hen-Hsen Huang, Makoto P. Kato, Josiane Mothe, and Barbara Poblete, editors, *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023*, pages 2420–2425. ACM, 2023.

[56] Alessio Molinari, Andrea Esuli, and Fabrizio Sebastiani. Improved risk minimization algorithms for technology-assisted review. *Intell. Syst. Appl.*, 18:200209, 2023.

[57] Mbasa Joaquim Molo, Emanuele Carlini, Luca Ciampi, Claudio Gennaro, and Lucia Vadicamo. Teacher-student models for ai vision at the edge: A car parking case study. In *19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024)*, 2024. accepted.

[58] Alejandro Moreo, Andrea Pedrotti, and Fabrizio Sebastiani. Generalized funnelling: Ensemble learning and heterogeneous document embeddings for cross-lingual text classification. *ACM Trans. Inf. Syst.*, 41(2):36:1–36:37, 2023.

[59] Giovanni Puccetti and Andrea Esuli. AIMH at multi-fake-detective: System report (short paper). In Mirko Lai, Stefano Menini, Marco Polignano, Valentina Russo, Rachele Sprugnoli, and Giulia Venturi, editors, *Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2023), Parma, Italy, September 7th-8th, 2023*, volume 3473 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2023.

[60] Pasquale Savino and Anna Tonazzini. Training a shallow nn to erase ink seepage in historical manuscripts based on a degradation model. *Neural computing & applications*. Accepted for publication.

[61] Pasquale Savino and Anna Tonazzini. In O. Gervasi, B. Murgante, A.M.A.C. Rocha, C. Garau, F. Scorza, Y. Karaca, and C.M. Torre, editors, *Computational Science and Its Applications - ICCSA 2023 Workshops, Athens, Greece, 3-6 July 2023*, volume 14108 of *Lecture notes in computer science*, Cham, 2023. Springer Nature Switzerland.

[62] Pasquale Savino and Anna Tonazzini. Preprocessing of recto-verso printed documents based on neural networks for text analysis, 2023. 3rd Conference on Digital Preservation and processing technology of Written Heritage, in conjunction with the 7th IEEE International Congress on Information Science and Technology (IEEE CiSt'23), Agadir - Essaouira, Morocco, 16-22/12/2023.

[63] Jan Sedmidubský, Fabio Carrara, and Giuseppe Amato. Segmentcodelist: Unsupervised representation learning for human skeleton data retrieval. In Jaap Kamps, Lorraine Goeuriot, Fabio Crestani, Maria Maistro, Hideo Joho, Brian Davis, Cathal Gurrin, Udo Kruschwitz, and Annalina Caputo, editors, *Advances in Information Retrieval - 45th European Conference on Information Retrieval, ECIR 2023, Dublin, Ireland, April 2-6, 2023, Proceedings, Part II*, volume 13981 of *Lecture Notes in Computer Science*, pages 110–124. Springer, 2023.

[64] Alessio Serra, Fabio Carrara, Maurizio Tesconi, and Fabrizio Falchi. The emotions of the crowd: Learning image sentiment from tweets via cross-modal distillation. In Kobi Gal, Ann Nowé, Grzegorz J. Nalepa, Roy Fairstein, and Roxana Radulescu, editors, *ECAI 2023 - 26th European Conference on Artificial Intelligence, September 30 - October 4, 2023, Kraków, Poland - Including 12th Conference on Prestigious Applications of Intelligent Systems (PAIS 2023)*, volume 372 of *Frontiers in Artificial Intelligence and Applications*, pages 2089–2096. IOS Press, 2023.

[65] Mattia Setzu, Silvia Corbara, Anna Monreale, Alejandro Moreo, and Fabrizio Sebastiani. Explainable authorship identification in cultural heritage applications: Analysis of a new perspective. *CoRR*, abs/2311.02237, 2023.

[66] Agnieszka Szczęsna, Paweł Foszner, Adam Cygan, Bartosz Bizoń, Michał Cogiel, Dominik Golba, Luca Ciampi, Nicola Messina, Elżbieta Macioszek, and Michał Staniszewski. Crowd simulation (CrowdSim2) for tracking and object detection, February 2023.

[67] Costantino Thanos, Carlo Meghini, Valentina Bartalesi, and Gianpaolo Coro. An exploratory approach to data driven knowledge creation. *Journal of Big Data*, 19, 2023.

[68] Amarante Tommaso. Design and development of a system for counting-related visual question answering. Master's thesis, M.Sc. in Artificial Intelligence and Data Engineering, University of Pisa, 2023.

[69] Francesco Del Turco. Deep learning methods for visual fish re-identification. Master's thesis, M.Sc. in Artificial Intelligence and Data Engineering, University of Pisa, 2023.

[70] Lucia Vadicamo, Giuseppe Amato, and Claudio Gennaro. Induced permutations for approximate metric search. *Information Systems*, 119:102286, 2023.

[71] Claudio Vairo, Marco Callieri, Fabio Carrara, Paolo Cignoni, Marco Di Benedetto, Claudio Gennaro, Daniela Giorgi, Gianpaolo Palma, Lucia Vadicamo, and Giuseppe Amato. Social and human centered XR. In Fabrizio Falchi, Fosca Giannotti, Anna Monreale, Chiara Boldrini, Salvatore Rinzivillo, and Sara Colantonio, editors, *Proceedings of the Italia Intelligenza Artificiale - Thematic Workshops co-located with the 3rd CINI National Lab AIIS Conference on Artificial Intelligence (Ital IA 2023), Pisa, Italy, May 29-30, 2023*, volume 3486 of *CEUR Workshop Proceedings*, pages 48–53. CEUR-WS.org, 2023.

[72] Xenophon Zabulis, Nikolaos Partarakis, Ioanna Demeridou, Paraskevi Doulgeraki, Emmanouil Zidianakis, Antonis Argyros, Maria Theodoridou, Yannis Marketakis, Carlo Meghini, Valentina Bartalesi, Nicolò Pratelli, Christian Holz, Paul Streli, Manuel Meier, Matias Katajavaara Seidler, Laura Werup, Peiman Fallahian Sichani, Sotiris Manitsaris, Gavriela Senteri, Arnaud Dubois, Chistodoulos Ringas, Aikaterini Ziova, Eleana Tasiopoulou, Danai Kaplanidi, David Arnaud, Patricia Hee, Gregorio Canavate, Marie-Adelaide Benvenuti, and Jelena Krivokapic. A roadmap for craft understanding, education, training, and preservation. *Heritage*, 6(7):5305–5328, 2023.