PhD Program in Science & Technology for Electronic & Telecommunication Engineering

YEAR 2026

Curriculum in

Computational Vision, Automatic Recognition and Learning

Visione Computazionale, Riconoscimento e Apprendimento Automatico

ATTENTION

The PhD application also implies submitting a research proposal under one or more themes chosen among those below indicated.
To write a proper research proposal, please follow the instructions indicated in the following file:
https://pavisdata.iit.it/data/phd/ResearchProjectTemplate.pdf

Research themes

Theme A
Multimodal scene understanding
Tutor: Vittorio Murino

This topic focuses on the research and implementation of algorithms that leverage big data, often multimodal, coming from 24h/7d sensor streams, to monitor a certain environment and understand the scenario. In particular, we aim at exploiting optical (RGB), 3D (e.g. Depth) and acoustic data to detect, track and classify people activities and objects, events and behaviors in general. Due to the inherent ability of deep learning architectures to cope with multimodal data, we first target such class of methods. Other machine learning approaches and fusion methods will in any case be investigated. For instance, additional information contained in different modalities, such as language, can be exploited in the *privileged information learning* framework or within the *Teacher-Student* paradigm.

Theme B

Continual generalized category discovery, incremental learning

Tutors: Vittorio Murino, Cigdem Beyan

This topic focuses on the deployment of techniques allowing a recognition system to discriminate classes and categories at test time, differently from the typical training & test process based on batch. We encompass cases where a few annotated instances of the novel classes are available (few-shot learning) or are provided in an incremental way, with the goal of distinguishing whether such objects belong to the currently trained items or are brand new. This also implies the ability to maintain high classification performance in old classes when learning new ones. Differently from the standard continual learning and incremental scenario in which each new task contains new classes never seen before in the previous tasks, and every task is fully supervised, in the generalized case, only the first task is fully supervised (and is typically large enough to allow learning a strong backbone representation), but each subsequent task contains unlabeled samples from both new and old classes. So, Generalized Category Discovery must simultaneously solve many problems: (i) discover which of the new samples belong to new classes, and which to old ones; (ii) how to represent these new classes and assign new samples to them; and (iii) how to update the backbone and maintain alignment of previous classifiers to drifting representations. This topic has affinities with few-shot and zero-shot learning, and is nowadays also tackled leveraging vision-language models (VLMs) for which we

Theme C

Multimodal Learning and Vision Language Model Adaptation for Temporal Medical Data Analysis

Tutors: Vittorio Murino, Cigdem Beyan, Barbara Toniella Corradini

This PhD topic focuses on multimodal learning for the analysis of temporal medical data, where information evolves over time such as in longitudinal studies, medical videos, and continuous monitoring signals. A key challenge in this domain is the limited availability of annotated data and the difficulty of integrating heterogeneous sources including imaging, clinical text, and sensor data. The project explores how Foundation Models, particularly Vision Language Models, can be adapted to effectively model and reason over temporal data in biomedical settings. Different adaptation strategies will be considered, allowing flexibility across tasks and modalities without restricting the work to a single approach.

The goal is to develop robust and generalizable methods for understanding evolving medical data, enabling applications such as disease progression analysis, video-based diagnostics, and patient monitoring over time.

Theme D

Reinforcement Learning with Curriculum Strategies in Multimodal Healthcare Data

Tutors: Cigdem Beyan, Vittorio Murino

This research explores the integration of reinforcement learning with curriculum learning strategies in the context of multimodal healthcare data. The goal is to investigate how structured learning processes can improve the efficiency, robustness, and generalization of models operating across heterogeneous data sources, such as medical imaging, clinical records, and physiological signals. Particular attention will be devoted to how learning curricula can be designed to progressively guide models from simpler to more complex tasks, potentially

addressing challenges such as data sparsity, modality imbalance, and domain variability. The work will also examine how reinforcement learning frameworks can support adaptive decision-making in dynamic clinical settings. The project aims to contribute methodological insights and practical approaches for leveraging multimodal data in healthcare, with a focus on scalability, interpretability, and real-world applicability.

Theme E
Artificial Intelligence for Physical, Natural and Life Sciences
Tutors: Vittorio Murino, Daniele Berardini, Barbara Corradini

This research theme aims to integrate physically grounded simulation models, traditionally based on differential equations and numerical solvers, with modern machine learning approaches capable of learning from large-scale multimodal data. Particular attention will be devoted to state-of-the-art methodologies such as physics-informed neural networks, graph neural networks, foundation models based on transformers and diffusion, with a particular focus on self-supervised and multimodal learning frameworks. These recent architectures advancements demonstrate how AI can support scientific simulation, forecasting, molecular and protein modelling, and representation learning across heterogeneous data modalities. The research will investigate how to combine scientific constraints, structured latent representations, and multimodal data to build robust, interpretable, and adaptive models applied to complex biochemical data such as sensor signals, molecular graphs, biomedical images, and genomic sequences.

Theme F
Discovery and mitigate bias in deep neural networks
Tutors: Vittorio Murino, Vito Paolo Pastore

Deep neural networks have been shown to achieve outstanding performances in a wide range of tasks, including image classification and segmentation. However, these models struggle when trained on bias datasets, impacting their generalization on unseen data. In such cases, bias may be defined as the existence of strong spurious correlations between data and target, causing a deep model to rely on shortcuts related to bias when making predictions, rather than on semantic attributes of interest. Consequently, they may perform poorly on unseen data not showing bias-related correlations (i.e., unbiased data). Recently, several methods have been proposed to mitigate bias dependency in deep neural networks, either assuming that bias information is available (supervised debiasing) or not (unsupervised debiasing). The latter describes a more realistic scenario, when not only bias attributes but even the presence of bias itself is unknown. This PhD proposal aims to design unsupervised methods for bias mitigation, exploring different computer vision tasks. Starting from state-of-the-art solutions, the research will include identification and bias-discovery in a trained model, possibly using Vision-Language models to provide both visual and textual representation of the identified bias. Furthermore, the research will include exploration of how bias propagates into a neural network, at a layer level, exploring benchmark datasets, with a particular focus on multimodal ones.

Theme G

Federated Learning by multimodal generative modeling

Tutors: Vittorio Murino, Daniele Berardini, Vito Paolo Pastore

Generative modeling has advanced significantly with diffusion-based architectures capable of learning structured data distributions across multiple modalities. However, current approaches assume centralized data access and joint optimization across all modalities and sources, which is often impractical in privacy-sensitive, heterogeneous environments. This project proposes a novel research direction: enabling multimodal generative modeling within a federated learning framework, where distributed clients contribute modality-specific data without sharing raw samples. The goal is to design a generative, diffusion-grounded architecture that supports distributed, unimodal training at the client side while synthesizing coherent multimodal outputs at the global level. Key challenges include learning a shared latent diffusion space from non-IID and modality-partial data, aligning client-specific generation objectives with cross-modal consistency, and addressing modality imbalance across clients. The research will explore encoder-decoder diffusion structures that aggregate cross-client information into a shared generative backbone, paired with modality-specific decoding heads and aggregation-aware optimization protocols. It will also investigate how generative modeling itself can mitigate statistical heterogeneity and coordinate learning across disjoint data distributions and modalities. This work aims to advance multimodal generation in decentralized contexts, with implications for tasks such as privacy-preserving synthesis in healthcare, sensor-level integration in robotics, and other real-world scenarios involving distributed and heterogeneous data sources.

Theme H

Biomedical imaging & Neuroimaging

Tutors: Vittorio Murino, Daniele Berardini, Vito Paolo Pastore

The wide adoption of biomedical sensors (e.g., MRI, TAC, SPECT, MEG, EEG, Fluorescence Microscopy, etc.) in various medical and biological investigations is fostering an increasing interest in advanced tools supporting the expert in the analysis and interpretation of the produced 2D/3D/4D images, both in clinical and scientific applications.

In this perspective, this theme will address research related to the development of computer aided diagnosis (CAD) systems, addressing tasks ranging from image analysis, image segmentation and detection, up to automatic determination of disease biomarkers and more advanced data analysis, with applications to connectomics and radiomics. Particular attention will be devoted to structural data, functional data and the investigation of relationships between function and the underlying structure. To this aim, multimodal data analysis and fusion will play a strategic role. The development of such CAD tools will also require the design of novel computer vision, pattern recognition and machine learning techniques for biomedical data. In this line, particular attention will be devoted to deep learning models. Research will address a range of possible applications, from biomedical image analysis (e.g., image segmentation of particular body organs, and classification of their conditions) to cell imaging (e.g., cells detections and segmentation, relationship between cell morphology and function), with particular attention to problems related to brain imaging such as, e.g., characterization of mental or neuro-degenerative diseases, investigation of cognitive functions, functional and structural connectomics, etc.

Theme I
Computer vision into the wild
Tutor: Vittorio Murino

This research theme focuses on the development of machine learning models for computer vision that can be deployed into the wild. More specifically, one drawback of modern learning systems is that they strongly rely on the characteristics of the data they are trained with. This results in models that poorly generalize to context unexplored during training (for example, consider a home robot that is deployed in a new house). To overcome this liability, two main strategies are domain adaptation and domain generalization. In the former case, we can leverage non-annotated samples from a desired scenario during training, and design models that better *adapt* to that domain. In the latter, the goal is to generalize to domains that are utterly unseen during training. The design of new training procedures to solve these tasks and the identification of novel application settings represent the main directions of this research.

Theme L
Few-Shot and Zero-Shot Deepfake Detection in the Era of Foundation Models
Tutors: Barbara Toniella Corradini, Vittorio Murino

The continuous emergence of highly advanced generative Foundation Models (FMs), capable of synthesizing hyper-realistic and multimodal content, has created an urgent need for robust deepfake detection systems. Because traditional detectors—which rely on specific artifacts from known training data—quickly become obsolete against novel generation architectures, there is a critical demand for verification frameworks capable of identifying synthetic media under strictly limited supervision. To address this challenge, this project tackles deepfake detection in few-shot and zero-shot scenarios, focusing on achieving strong cross-generator generalization to identify manipulated media even from completely unseen generators. We accomplish this by leveraging Multimodal Foundation Models, including Vision-Language Models (VLMs) and Audio-Language Models (ALMs). By exploiting the semantically rich latent spaces of these FMs, we aim to design adaptable semantic and forensic representations, moving away from rigid detectors to flexible frameworks that evaluate visual, acoustic, and cross-modal cues simultaneously. Crucially, our methodology should aim to rely on FMs representations to grasp the intrinsic anomalies of deepfakes, regardless of the underlying generation algorithm. The proposed models should be designed to recognize subtle, multi-domain inconsistencies—such as anomalous temporal dynamics, unnatural facial behaviors, biometric identity mismatches, and misaligned image-sound patterns. By establishing reliable, generator-agnostic detection systems, this research advances critical applications in media forensics, automated misinformation detection, and trustworthy digital communication.

Theme M
Controlling Image Generation in Multimodal Models via Internal Representation
Tutors: Barbara Toniella Corradini, Vittorio Murino

This research addresses a critical limitation in current multimodal generative models: the lack of precise control when relying solely on external text prompts. While modern text-to-image systems generate exceptionally high-quality visuals, their generation processes remain largely opaque and difficult to steer. The core concept of this project is to investigate how semantic, spatial, stylistic, and compositional data is encoded within the latent spaces and intermediate activations of these models. By leveraging these internal representations, we can bypass the ambiguity of text prompts and directly guide the generation process. This enables granular control over object attributes, scene layouts, visual styling, and the interactions between specific concepts, whilst preserving core identities. Our objective is to develop robust methods for interpreting, editing, and controlling the internal features of foundation generative models, aimed at the design of more understandable and explainable AI algorithms. This will facilitate image synthesis that is more reliable, interpretable, and responsive to user intent. Potential applications are vast, encompassing creative content generation, visual editing, medical image synthesis, data augmentation, and safety-aware generation.