PhD Program in Science & Technologies for Electronics & Telecommunication Engineering

YEAR 2024 - 2025

Curriculum in

Computer 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

Generative Multi-modal models for spatial perception and robotic behavior generation

Tutor: Alessio Del Bue, Pietro Morerio

This topic focuses on leveraging multiple sources of sensory input to enable robots to learn and execute complex behaviors in general 3D environments. By integrating data from cameras, LiDAR, and inertial measurement units (IMUs), robots can perceive and understand the 3D structure and semantic of the environment more comprehensively. Generative AI models can be employed to process and fuse these diverse modalities effectively. This fusion allows robots to make informed decisions and adapt their actions in real-time based on the rich information received from their surroundings. Ultimately, multi-modal models empower robots to exhibit more versatile and adaptive behaviors, making them more capable and efficient in a wide range of tasks and environments.

Theme B

Topological Deep Learning and Graph Neural Networks

Tutor: Alessio Del Bue, Pietro Morerio

Topological deep learning integrates techniques from deep learning with algebraic topology to analyze complex data structures, supported on topological domains, e.g. graphs and hypergraphs. Unlike traditional methods, it focuses on capturing intrinsic geometric properties, regardless of dimensionality or shape. By representing data as topological spaces, topological deep learning uncovers hidden patterns in irregular or noisy datasets. Graph Convolutional Networks (GCNs) are the state-of-the-art tools used to solve tasks on data that exhibit a graph-like structure. They are specifically designed to capture and exploit the graph's topology that models the interaction between different entities in a complex system via a graph-based convolution operator. In spectral-based GCNs, a category of GCNS, one of the crucial components is constructing operators (satisfying certain mathematical properties) for convolutions on graphs. We are interested in investigating the design of novel operators to address several limitations observed in the state of the art, including handling multigraphs, temporal graphs, and hypergraphs.

Theme C

Leveraging generative AI to learn in imperfect data regime

Tutor: Vittorio Murino

This topic focuses on non-ideal scenarios for deep learning approaches, in which the data can be unlabeled, noisy labelled, class-imbalanced, biased, or available just as a bunch of samples. We will address both unimodal and multimodal settings, where multisensory (multimodal) data can be acquired or extracted with different spatial resolution and loosely synchronized in time. The core technical idea to tackle these issues is based on the development of data augmentation and generation methods. The intuition is that the problems derived from data with missing, partially missing or wrong annotations, as well as under-represented or mis-represented classes, can be faced by finding adequate data transformation of the available data, or even by generating suitable synthetic data able to correct or rebalance the flaws of the available samples. Besides, along the same strategic line, this theme also aims at developing self-supervised training methods leveraging the different multimodal data and the distillation paradigm in imperfect multimodal settings, especially when labels are fully unavailable. Language models and multimodal (vision+language) foundation models are also to be considered. Possible actual application test cases range from the areas of industrial process control, behavior analysis, biomedical and health, and many others.

Theme D

Learning with imperfect data: Transfer Learning, Unsupervised Domain Adaptation and Domain Generalization

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 issue, two main strategies are unsupervised 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 generalizing to domains that are utterly unseen during training. The design of new training procedures to solve these tasks, which involve unsupervised learning and generative methods, even in multimodal settings, and the identification of novel application

scenarios represent the main directions of this research.