PhD Program in Sciences & Technologies for Electronics & Telecommunication

YEAR 2021

Curriculum in

Computer Vision, Pattern Recognition and Machine Learning

Visione Computazionale, Riconoscimento e Apprendimento Automatico

ATTENTION

The PhD application also implies to submit 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

Research themes

Theme A
3D scene understanding with geometrical and deep learning reasoning
Tutor: Alessio Del Bue, Yiming Wang

Classical multi-view geometry problems make use of geometric reasoning to infer the scene 3D structure and its dynamic. These approaches often neglects the semantic composition of the scene that instead provides important cues about objects motion and their current spatial configuration. Instead, such context and semantic information can be given by current deep learning architectures but very few works attempted to merge geometrical reasoning with such semantic information. This research theme will have the aim to bridge this gap and to provide solutions that can be applied on intelligent system for robotics and autonomous driving. This research will study different methods and tools involving scene graphs, object detection with either multi-view images or 3D data, deep learning methods for scene representation and large-scale 3D reconstruction in dynamic environments. The research activities will build over and expand the current expertise on these problems developed under the Spatial AI topics at PAVIS research line.

Theme B

Artificial Intelligence for Human Behavior Analysis

Tutor: Alessio Del Bue, Cigdem Beyan

Development of techniques and frameworks for the analysis of human behaviours including actions, expressions, social signals and emotions. To recognize and interpret human behavior, it is fundamental to identify the subjects involved to an action, which can be realized by a single person or a group of people. In this context, methods for detecting, tracking persons and groups using video sequences acquired from cameras, which might be distributed in the environment in several sparse locations, and from other types of sensors (e.g., microphones) will be considered. Further, human body modeling is a mandatory task to extract from images the different components of the human body, like head, torso, arms, legs, etc., so as to estimate posture, gesture and other social cues widely known as useful hints to classify behavior of the persons and groups and recognize the situations. Computer vision and machine learning constitute the focus of this research, and particular attention will be specifically devoted to deep learning models. Applications in the contest of surveillance and security as well as in the biomedical field (e.g., monitoring of elderly people) are envisioned.

Theme C

People and Object Re-identification in the wild

Tutor: Alessio Del Bue, Vittorio Murino

Study and development of deep learning biometric techniques for scene analysis and understanding. The research will mainly focus on person or object characterization in a sensor network using attributes coming from multi-modal data in challenging conditions (e.g., person in crowd or object in cluttered environment). The so-called re-identification problem aims to recover the identity of persons or objects as viewed in different times and places. Not only optical cameras will be used, but other information derived from different sensors may also be utilized (e.g., range, thermal and others). The robustness to environmental (real) conditions and the non-cooperation of the subjects are the main features to which the developed techniques will have to cope with to deploy this technology in real scenarios. Computer vision and machine learning constitute the focus of this research, and attention will also be devoted to deep learning models.

Theme D

Deep Learning for Multi-modal scene understanding

Tutor: Alessio Del Bue, Pietro Morerio, Vittorio Murino

This topic is related to the research and implementation of algorithms that leverage multimodal data, namely data coming from different sensors, for general purposes such as classification, recognition and, more in general, scene understanding. In particular, we primarily aim at exploiting optical (RGB) sensors, range or depth sensors (e.g. kinect), thermal sensors (thermal and near-infrared cameras), acoustic sensors (monoaural microphones, and our in-house opto-acoustical sensor, DualCam) to detect persons and objects, tracking and classifying objects, events and behaviors in general. The main focus will be on learning and developing new deep learning methods, as particularly suitable to merge heterogeneous information coming from different sensor modalities.

Theme E

Weakly-Supervised and Unsupervised Deep Learning

Tutor: Alessio Del Bue, Jacopo Cavazza

One of the key factors behind the recent popularity of deep learning algorithms is the possibility of leveraging a large corpus of labelled data. Despite gathering massive amount of data is nowadays not problematic, differently, data annotation is surely a major bottleneck. In fact, not only it is time-consuming and economically expensive, but it is also prone to errors since requiring human intervention. This research topic focuses on how to relax the level of supervision required to develop (deep) machine learning algorithms. The ultimate goal is to devise new computational techniques which exploit and discover geometrical inner properties of the data (e.g., semi-supervised learning, clustering), while also considering the transfer of knowledge from existing labelled datasets in order to recognize categories and classes for which a little or even no labelled data is provided (few-shot or zero-shot learning).

Theme F

Visual Reasoning with Knowledge and Graph Neural Networks for scene understanding

Tutor: Alessio Del Bue, Stuart James, Sebastiano Vascon

Machine ability to detect objects within images has surpassed human ability, however, when posed with relatively simple more complex tasks machines quickly struggle. This theme focuses on developing AI systems that are able to access knowledge stored in Knowledge Graphs to understand the world around the camera view. Few works have successfully integrated knowledge in Computer Vision systems and knowledge graphs provide one avenue. This research will study methods to integrate knowledge for user interaction via retrieval or visual question and answering within real world environments. In particular, shallow and deep graph-based methodologies are promising computational framework to include external knowledge and also maintaining a high degree of interpretability, a necessary feature for modern AI systems.

Theme G

Distributed AI in sensor networks and robotic platforms

Tutor: Alessio Del Bue, Lorenzo Natale, Agnieszka Wykowska

The integration of smart building technology and robotics has great potentials. Smart buildings equipped with robots have a much high level of autonomy, because they can physically interact with the environment and humans, in addition they can actively inspect the scene to get additional information if needed. Robots, on the other hand, can have access to a larger set of sensors and computing power, than what is available on-board. In this setting robots can monitor the environment from a different perspective, and adapt their behaviour depending on the situation. This research theme we will develop AI approaches for egocentric and allocentric scene understanding by leveraging information form robotic platforms and camera networks deployed in indoor environments. Topics of research are related to behaviour understanding, dynamic scene analysis, domain adaptation for object detection from different views, and attention mechanisms using egocentric and allocentric data. This research will be implemented on real hardware platforms already deployed at IIT with the target to deploy AI assistive systems that can interact with and support humans in several high-level tasks.