PhD Program in Sciences & Technologies for Electronics & Telecommunication

YEAR 2019

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_2018.pdf

Research themes

Theme A
Computer vision and deep learning for AI spatial reasoning
Tutor: Alessio Del Bue

The research theme will study and development of techniques and systems for the analysis of the semantic 3D structure of a scene using deep learning and computer vision approaches with specific applications on autonomous vehicles and agents. The aim is to develop next generation AI systems that can understand the spatial arrangement of a generic area and its dynamic in order to allow further physical interaction with or without human presence. In particular, the scene context will be modelled using novel deep architectures based on graphs (e.g. Graph Nets) that can scale from indoor to large outdoor scene. A strong emphasis will be also put on the interpretability and customisation of the learned models in generic scenarios. The developed models will enable next generation system that can effectively understand and interact in the real world. Skills will be developed mainly in Computer Vision and Machine Learning topics and using images, depth sensor modalities from cameras.

Theme B
Zero- and Few-Shot Learning
Tutor: Alessio Del Bue, Vittorio Murino

This topic focuses on the deployment of techniques allowing a recognition system to discriminate classes and categories which differ from the ones considered for training. We encompass cases where a few annotated instances of the novel classes are available (few-shot learning) or, when the novel classes and categories remain totally unseen, auxiliary textual descriptions/attributes are exploited to compensate the absence of visual data (zero-shot learning). Machine learning and computer vision constitute the focus of this research. Multimodal approaches to combine images/videos with attributes/text embeddings are of crucial interest and utility for this research theme. In this respect, a particular attention will be devoted to deep learning frameworks, capable of solving zero- and few-shot learning through an end-to-end pipeline.

Theme C
Online 3D scene understanding with geometrical and deep learning reasoning
Tutor: Alessio Del Bue

Classical multi-view geometry problems make use of geometric reasoning to infer the scene 3D structure and its dynamic, even in real-time. 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 in real-time on intelligent system for robotics and autonomous driving. This research will study different methods and tools involving object detection in multi-view images, deep learning methods for scene representation and large-scale 3D reconstruction in dynamic environments.

Theme D
Re-identification from multi-modal data
Tutor: Alessio Del Bue, Vittorio Murino

Study and development of biometric techniques for scene analysis and understanding using multi-modal sensors. The research will mainly focus on person characterization, with possible focus on the usage of soft biometrics cues (3D, attributes) and in challenging conditions (e.g., crowd). The idea is to recover the identity of persons as viewed in different times and places, also considering face/body attributes, the so-called re-identification problem. Not only optical cameras will be used, but other information derived from different sensors may also be utilized (e.g., range, thermal, event cameras). 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 particular attention will also be specifically devoted to deep learning models.