Workshop 2023
ePicture This
an initiative by PENTA label projects
MANTIS and IMAGINATION with AENEAS support

ABSTRACT PRESENTATION

Title: **Machine Learning for Computer Vision, a Case Study in Hu**
Interaction

Presenter name: Boris Lenseigne, Julia Cohen
Company name / Institute: LR Technologies
Project name: Libellul'ergo
Funding group: ~~Penta / Xecs / Euripides / ECSEL / KDT~~
Abstract can be published on website: ☐ **YES** ☐ ~~**NO**~~

Provide abstract of 500 words maximum. Use font ARIAL, size 11.
If figures are used, the text plus figures must stay within this one page.

Machine learning has been a game changer in the recent years for computer vision applications, especially with deep learning methods. Providing that learning data is available, artificial neural networks outperform handcrafted approaches in object recognition and segmentation, while more dedicated architectures are a solution for the interpretation of image sequences or for higher level interpretation problems. A wide range of pre-existing networks is nowadays available and building an image analysis system with these technologies is mostly a matter of choosing a network topology, feeding it with data, and having it learn to perform the desired task.

The caveat resides in the data to be used for learning. Since artificial neural networks function as black boxes, to achieve a good performance, they must learn from a large amount of carefully labelled data that covers all the scenarios to be encountered and avoids any unwanted bias.

This communication will address these questions in the practical case of human-machine interaction. A consumer grade camera is used to perform eye and face tracking with the intend of using this information to drive the computer's mouse in the most intuitive way possible. Mouse control involves not only moving the mouse on the screen, but also interacting with the elements on the interface with both the left and right buttons, as well as scrolling up and down. Depending on the activity of the user, the face movements can be directly related to the mouse movements (clicking on an icon) or not (on-screen reading). A real-life application could be to enable people with disabilities to use computers, and thus become more independent in their daily life.

In this scenario, we will address the main issues of building AI-based vision systems: the choice of network topology, acquisition of the learning dataset, pre-processing and labelling of the data, learning and evaluation of the model. While doing this we will comment on the traps that paved our way and the strategies we used to solve them. Finally, we will compare this approach to the more traditional one, with the goal of providing insights on the pros and the cons of the ubiquitous usage of machine learning and deep learning when building a computer vision system.