Name of the course: Deep Neural Networks for Constrained Devices

Teachers: Edoardo Ragusa; e-mail: edoardo.ragusa@unige.it

Duration of the course: 12 hours

Language: Italian; if requested by foreign students, the course will be held in English.

Aims of the course: the course provides basic knowledge about the key aspects that should be considered when deploying a deep neural network on a constrained device. The course provides a brief introduction to the topic and presents a survey of the most important tools and techniques. Practical examples of quantization, pruning, and design of efficient architectures are provided using Tensorflow lite. In addition, two examples of deployment on Jetson Nano and Arduino Nano are presented.

Teaching program:

1. Introduction and Review of preliminary concepts;
2. Quantization;
3. Pruning
4. Design of efficient architectures;
5. Use case: Jetson Nano
6. Use case: Arduino Nano

Exam modality: Presentation of a project where the student exploits the techniques and the concepts learned during the course

Bibliography: Notes by the teacher (in English), Tensorflow Lite documentation https://www.tensorflow.org/lite

Timetable:

- 05/11/2024 (10-12)
- 07/11/2024 (10-12)
- 12/11/2024 (10-12)
- 14/11/2024 (10-12)
- 26/11/2024 (10-12)
- 28/11/2024 (10-12)

Room The location will be announced before the course begins. The course can also be attended through Microsoft Teams. Credential will be communicated after registration.

Registration: Please, contact the teacher to register for the course.