

Efficiency and Compression of Deep Neural Networks

a PostDoc call

Overview

Deep Neural Networks (DNNs) can solve extremely challenging tasks thanks to complex stacks of (convolutional) layers with thousands of neurons, especially to solve computer vision-related tasks like image classification, object detection or image segmentation. In the last few years a lot of focus has been devoted to some different aspects related to deep learning and computer vision: how can we improve the deep models learning strategy? Do we really need all the complexity we are currently using to solve tasks with deep learning? Is there going to be any difference at training and at inference time? How can we improve the computation on embedded devices?

In this postdoc call, these and many other aspects will be explored, in between theory and application. Specifically, the development of algorithms to quantize and compress DNNs, and deployment on a physical embedded device, will be the core project.

The position offered will be in presence at Telecom Paris - Institut Polytechnique de Paris, in the MultiMedia equipe of the IDS department, lasting 18 months. Proper equipment to fulfill the postdoc will be provided. Standard salary assessed on the Ecole basis. The post doc will start at the earliest date possible, to be agreed with the candidate.

Candidate profile

- PhD in Computer Science or related fields, emphasis on Machine learning/Deep learning.
- Deep knowledge of Machine learning and Deep learning.
- Coding in Python (C/CUDA is a plus), experience with Tensorflow/PyTorch or similar.
- Great passion and commitment towards research in AI.
- Proficiency in the English language, both spoken and written.
- Publication in Deep learning and embedded devices.

How to apply?

Send CV and two reference letters to enzo.tartaglione@telecom-paris.fr