Master Internship 2018:
Real-time Multi-person 3D pose estimation and tracking using Convolutional Neural Networks

**Key words:** Deep Learning, Convolutional Neural Networks, Pose Estimation, Tracking.

1 Context

Pose estimation (*i.e.*, retrieving the coordinates of body part joints from images) is a major computer vision problem that has recently seen tremendous advances thanks to deep convolutional neural networks (*e.g.*, [CYO+17, NYD16, YLO+17, LTP17]), and recent results have become sufficient to allow for interesting applications such as ergotherapy. Results of 2D pose estimation can be seen on Figure 1.

![Figure 1: Examples of 2D pose estimation on still images using convolutional neural networks.](image)

2 Goals and Challenges

The goal of this internship is to develop solutions for pose estimations in videos such as those shown in Figure 2. To achieve this aim, we want to extend the approach we developed in [LTP17] into two directions.

The first is performing 3D pose estimation, which consists in retrieving real world 3D coordinates for the body part joints, instead of image coordinates. To that extend, volumetric maps of the probability density of each joint are estimated using a deep convolutional network [PZDD17].

Second, we aim at tackling the case where multiple persons are depicted in the video. In that case, not only do we want the pose estimation for each individual but also the tracking information, *i.e.*, which skeleton belongs to who from one frame to another.
Finally, we want to be able to perform such detection in a real-time fashion. As such the computational cost of the designed architecture has to be taken into account and the temporal information in the video has to be exploited.

3 Profile

This opportunity is dedicated to master (or engineering school) students with major in computer science or applied mathematics. Background in machine learning, deep learning and optimization are required. A good understanding of modern ConvNet architectures (VGG, ResNet, DenseNet) is desirable.

Programming experience in python (Numpy, Scipy, matplotlib) is important, with skills to scale algorithms for videos. Prior experience with standard Deep Learning library (e.g. Tensorflow, Keras) would be a plus.

4 Modalities

Internship duration: 6 months, starting February–March 2018.
Location: Cnam Paris, 2 rue conté 75003.
Salary: 1100 € brut / month
Supervision: Nicolas Thome (Cnam), David Picard (ENSEA/CNRS).
Contact: nicolas.thome@cnam.fr, picard@ensea.fr.

This internship is funded by Ergonova with an option for following this work in the context of a PhD.

References


http://www.ergonova.fr/