Instantiation of deep neural networks on multiview-based embedded systems.

Deep Convolutional Neural Networks (CNNs) have been very successful over the past decade, becoming a standard in computer vision. This success has come with a large computing cost, making the deployment of CNNs a difficult task, especially with real-time constraints. Indeed, deep learning techniques are very computationally intensive, involving up to 30 billion operations to classify a single image. These high computational loads prevent the implementation of deep learning in systems that combine real-time performance and low power consumption.
In order to make this deployment possible, previous works exploit the important parallelism of these algorithms, which requires the use of dedicated hardware platforms. With power management constraints, FPGA-based processing cores are known as the solution of choice for accelerating computer vision applications. CNNs are good candidates where processing naturally takes place in data-flow, making FPGA-based hardware architectures even more relevant.
A first thesis addressed these issues related to the implementation of CNNs on FPGAs. This work aimed to improve the efficiency of implementations through two main optimization strategies: the first explores the model and parameters of CNNs, while the second focuses on hardware architectures adapted to the FPGA.

In this thesis, we propose to address the problem in a distributed form in terms of calculation and perception. The motivation is to demonstrate that we keep a good quality of inference if we degrade a DNN (Deep Neural Network) but that we increase the number of points of view. In other words, increasing visual information could simplify processing architectures.
The idea is therefore to use several cameras, each incorporating a "light" computer that cannot carry a full state-of-the-art DNN. This confrontation problem "information vs degradation" will be a first part of this thesis where it could be interesting to integrate “geometry” between the different cameras as additional information.
However, despite these degradations, it is unlikely that the embedded DNNs are sufficient to obtain a good inference. Thus, we envisage to merge the different flows into a SOC/FPGA-based central computer allowing the partial dynamic reconfiguration of the FPGA area. This study on dynamically reconfigurable architectures for CNNs will be the second part of this thesis.

Supervisor: *François BERRY – PR and Christophe BLANC - MCF*
Team: *Axe ISPR, DREAM – dream-lab.fr*
Laboratory*: Institut Pascal UMR 6602 CNRS*