



# Stage 2021/2022 : Estimation de l'impact des techniques de réduction des réseaux de neurones sur Micro Contrôleur

Janvier 2022

## 1 Description

Le stage concerne l'estimation de l'impact des techniques de réduction de réseaux de neurones convolutifs (CNN) lors de l'inférence sur des cibles embarquées de type Micro Contrôleur (MCU). Il consiste en l'utilisation de ces techniques de réduction avec l'outil TensorFlow, qui utilise le langage Python, sur des réseaux de neurones convolutifs dont MobileNet. Les techniques étudiées seront l'élagage, ou pruning en anglais, la quantification et la distillation.

Les modèles en Python obtenus à l'aide de TensorFlow, seront ensuite traduits en langage C pour être embarqués dans des cibles de type STM32 basées sur des cœurs de processeur ARM Cortex M4 et ARM Cortex M7. Le stagiaire utilisera la méthode EST développée au sein de l'équipe Syel du LIP6 pour mesurer et évaluer l'impact de l'application des techniques de réduction sur les performances du MCU.

## 2 Lieu et Moyens :

Le stage se déroulera au LIP6 à la faculté des Sciences de Sorbonne Université. Le stagiaire disposera de serveur de calcul de type PC avec Linux comme systèmes d'exploitation et de l'ensemble des outils nécessaires à la réalisation de son stage : compilateur C ARM, interpréteur Python, carte embarquée STM32.

## 3 Contacts :

Thomas Garbay, Bertrand Granado, Khalil Hachicha  
Sorbonne Université - Laboratoire d'Informatique de Paris 6  
4, Place Jussieu  
75252 Paris Cedex 05  
Téléphone : 01 44 27 96 33  
Mél: Bertrand.Granado@sorbonne-universite.fr