

# Réseau de neurones convolutifs sur System on Chip

**Louis Delabays**

Travail de Bachelor 2018

Filière informatique – Informatique Industrielle Embarquée

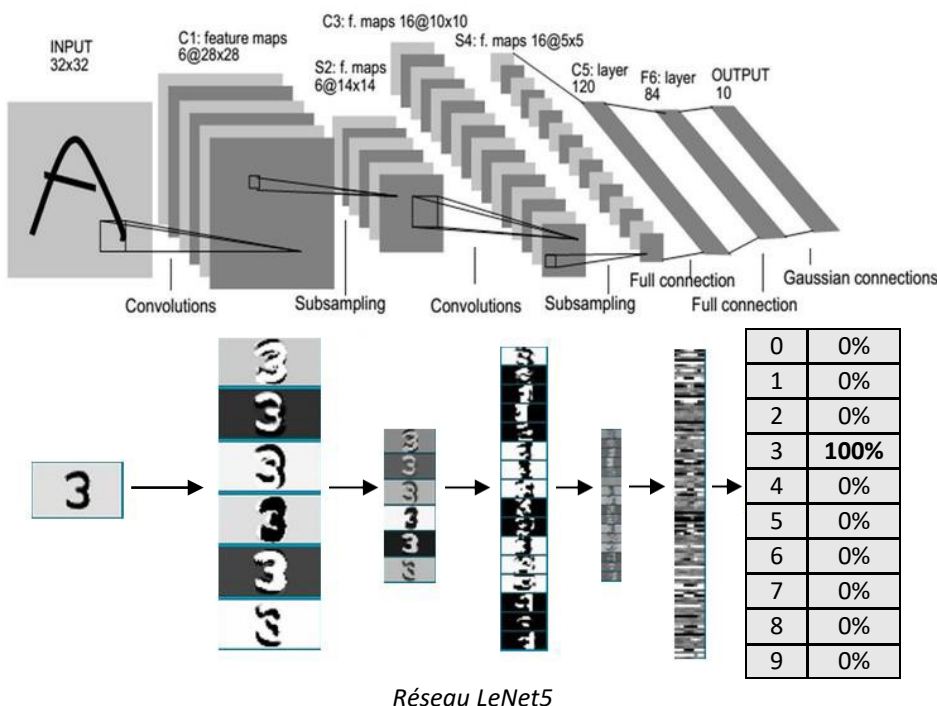
Professeure : Nuria Pazos

Expert-e-s: Maurizio Denna

## Description

L'utilisation des réseaux de neurones convolutifs est en plein essor pour la reconnaissance d'objets dans des images ou des vidéos. Ces réseaux ont besoin d'énormément de ressources de calcul. De ce fait, les temps de traitement sont importants. Pour porter ces réseaux sur des plateformes embarquées, il est nécessaire de créer des implémentations adaptées. La plupart des opérations effectuées par un réseau de neurones sont hautement parallélisables. En particulier la couche de convolution qui prend en général la majorité du temps et des ressources pour l'exécution d'un réseau.

Dans le cadre du projet Bonseyes, ce travail porte sur la conception et l'implémentation d'un accélérateur matériel de convolution. Il doit accélérer les opérations de convolution et de pooling du réseau LeNet5. Ce réseau sert à la classification de chiffres manuscrits. Il est par exemple utilisé pour lire automatiquement les numéros postaux sur les enveloppes. L'accélérateur, ou coprocesseur, est écrit en VHDL et est déployé sur un système sur puce qui contient 2 processeurs et une FPGA.



## Déroulement

Voici les différentes phases du projet:

- Familiarisation avec les réseaux de neurones convolutifs (CNN)
- Conception de l'architecture du coprocesseur de convolution
- Réalisation du coprocesseur en VHDL
- Création de l'interface CPU – FPGA pour le transfert des données
- Validation et test des performances de la solution proposée
- Réalisation d'un démonstrateur capable de classifier des images

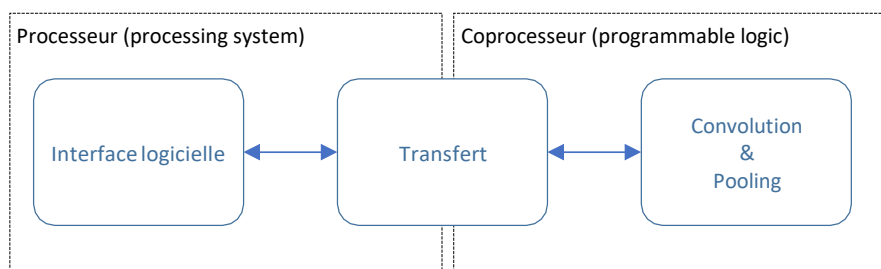
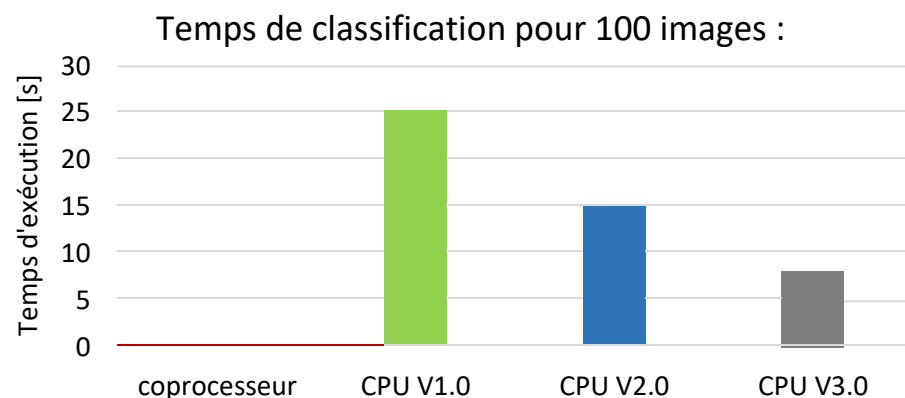


Schéma général du système

## Résultats

Le coprocesseur est intégré dans l'implémentation du réseau de neurones et la classification fonctionne.

Le gain de performance du coprocesseur varie de x4 à x12 par rapport aux implémentations purement CPU testées sur la plateforme embarquée.



## Perspectives

Le coprocesseur développé est spécifique au réseau LeNet5. Beaucoup de réseaux de neurones différents existent pour tous types d'applications de reconnaissance et d'apprentissage. Ce projet propose une première version du coprocesseur de convolution qui a permis d'identifier les points critiques. Cela est utile pour le développement d'une version généraliste du coprocesseur qui supporte les différentes implémentations des couches de convolution de la plupart des réseaux de neurones. Le déploiement sur un système sur puce plus grand est aussi envisageable.