



Stage M2 2018-2019 :

$\sqrt{(CNN)}$

(Réseaux Convolutionnels embarqués pour les Capsules Vidéo Endoscopiques)

30 octobre 2018

1 Description

Depuis 2012, les réseaux convolutionnels, re-popularisés à travers les travaux de l'équipe de G. Hinton, ont démontré leur capacité de reconnaissance surpassant largement la plupart des autres modèles neuronaux ou statistiques [1]. Si ces bons résultats rendent éligibles ces réseaux dès lors que puissance de calcul et énergie sont présentes, il n'en va pas de même lorsque qu'il est envisagé de les intégrer dans un système embarqué temps réel ayant des contraintes énergétiques fortes.

A titre d'exemple, le réseau GoogleNet [2] possède quelques 4 millions de paramètres sur 32 ou 64 bits soit 128 à 256 millions de bits uniquement pour la représentation des données, à ce chiffre il est nécessaire d'y ajouter les opérateurs de calcul. Pour un traitement temps réel, la duplication de ces opérateurs est indispensable, générant un nombre d'opérateurs proportionnels au nombre de données. Cela dépasse largement la capacité d'intégration des circuits électroniques, il y a au moins un ordre de grandeur entre la taille des réseaux de neurones convolutionnels existants et les capacités d'intégration des circuits électroniques embarquables.

Il est nécessaire si nous voulons embarquer un réseau convolutionnel d'étudier comment réduire sa taille afin d'optimiser la quantité de mémoire nécessaire, le nombre de calculs réalisés avec comme impact direct une forte réduction de l'énergie consommée pour un calcul en temps réel.

Pour mesurer la performance d'un réseau de neurones convolutionnel, deux métriques sont utilisées : la sensibilité et la spécificité.

Dans notre cas, la détection des polypes dans une image, la sensibilité est défini comme le taux d'images avec polype qui sont correctement prédites, voir l'équation 1.

$$\text{Sensibilité} = \frac{VP}{VP + FP} \quad (1)$$

La spécificité représente le taux d'images sans polype qui sont correctement prédites.

$$\text{Spécificité} = \frac{VN}{VN + FN} \quad (2)$$

Dans les équations 1 et 2, VP représente le taux de vrais positifs qui indique le nombre d'images avec polypes qui sont correctement classées par le réseau de neurones convolutionnel. De la même façon, VN représente vrai négatif qui indique le nombre de images sans polype qui sont correctement classées par le réseau. FP représente les faux positifs qui indique le nombre de images avec polypes qui sont incorrectement classées par le réseau et FN représente les faux négatifs qui indique le nombre d'images sans polype qui sont incorrectement classées par le réseau.

Lors de la construction d'un réseau convolutionnel, nous cherchons à ce que ces deux métriques soient proches de 1, cette optimisation combinée contribue à la complexité du réseau.

Dans ce stage nous désirons étudier le découpage en deux étapes du traitement, un premier réseau sera conçu pour avoir une sensibilité maximale sans se soucier de la spécificité. Ce premier réseau permettra la sélection d'images qui seront passées à un second réseau dont l'objectif sera d'avoir une spécificité maximale sans se soucier de la sensibilité. Avec cette approche la taille de chaque réseau sera réduite avec comme objectif d'obtenir une réduction de l'ordre de $\sqrt{\text{TailleCNN}}$ où TailleCNN est le nombre de paramètre d'un réseau convolutionnel de type GoogleNet.

Ce stage a pour ambition de réaliser le modèle de ce nouveau réseau convolutionnel. Pour cela le stagiaire aura à sa disposition une banque de 1500 images provenant d'examen endoscopiques.

Une fois le réseau convolutionnel créé, l'étude portera sur son embarquabilité en prenant en compte les contraintes liées à l'intégration de l'algorithme dans un système embarqué en se basant sur des critères de puissance de calcul, de débit d'entrée/sortie, de consommation et de facteur de formes. Une description algorithmique permettant l'utilisation de synthèse de haut niveau sera réalisée.

Des compétences en traitement des image et/ou en conception de systèmes numériques seront appréciés dans le cadre de ce stage.

2 Lieu et Moyens :

Le stage se déroulera au LIP6 à la faculté des Sciences de Sorbonne Université. Les outils utilisés seront des PC sous Linux avec les outils de développement SDSoc et Vivado de Xilinx, Matlab et gcc.

3 Contacts :

Bertrand Granado, Andrea Pinna, Orlando Chuquimia, Hichem Sahbi
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

4 Références bibliographiques

- [1] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, 2012.
- [2] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.