

Définition d'un environnement de modélisation et de développement pour la plateforme matérielle NATvision : intégration de fonctions d'apprentissage pour le traitement vidéo

Doctorant : Denis SHEMONAEV

Encadrants : Christophe JEGO, Bertrand LE GAL, Jeremie CRENNÉ (Laboratoire IMS)

Anthony BESSEAU (Société EMG2)

La plateforme NATvision[1] développée par l'entreprise NAT GmbH, est une solution de calcul intensif (HPC) FPGA, basée sur le standard MicroTCA. Conçue autour d'un cœur Zynq Ultrascale+ (combinaison d'un FPGA et d'un SoC ARM), elle permet l'acquisition synchronisée et le traitement en temps réel de plusieurs flux vidéos provenant de caméras distantes. Le standard MicroTCA confère à cette plateforme une grande modularité et évolutivité, permettant de combiner aussi divers autres moyens d'acquisition, de communication ou de calcul à base de FPGA, CPU ou GPU. Les domaines applicatifs sont multiples que ce soit pour les transports, la sécurité, les industries, le médical ou encore la logistique. Pour ces domaines applicatifs, la vision par ordinateur peut répondre aux besoins de contrôle qualité ou d'inspection visuelle, de surveillance, de diagnostic voire également pour la détection de comportements suspects dans les lieux publics. Or, avec l'émergence de l'industrie 4.0 et des usages de l'apprentissage automatique, de nouveaux services doivent être intégrés dans des systèmes embarqués, au plus près des capteurs : on parle d'applications "Edge Computing" en contraste avec des traitements sur le "Cloud".

Il apparaît, depuis le début de la thèse, que des algorithmes moins énergivores au niveau des traitements de données, souvent dérivés de YOLO[2] ont été conçus récemment pour cibler de tels systèmes embarqués[3]. Cependant, ces derniers s'orientent principalement vers des cœurs de processeurs ARM ou GPU. Ces processeurs ne permettent pas d'obtenir un compromis intéressant entre performance, flexibilité et consommation énergétique, à la différence du FPGA qui semble être plus intéressant[4]. La plateforme NATvision s'inscrit donc parfaitement dans ce cadre mais elle souffre malheureusement du manque de solutions permettant d'exploiter à la fois les entrées vidéo et l'IA accélérée par exécution sur circuit FPGA. Les premiers travaux de cette thèse visaient donc à identifier ces lacunes et envisager un environnement haut niveau tirant partie de la pluralité des flux vidéo, de la modularité d'un circuit FPGA et des nouvelles avancées de l'IA. Il a été décidé de s'orienter vers Vitis AI, un outil développé par Xilinx, utilisé en combinaison de la bibliothèque PyTorch, pour développer et exploiter dans un premier temps un réseau personnalisé pour la classification d'images issues de la base de donnée CIFAR-10[5]. En parallèle, un environnement de prototypage se rapprochant des possibilités offertes par la plateforme NATvision a été mis en place. Des cartes KV260 embarquant le nouveau SOM KRIA[4] ont été choisies pour récupérer des flux vidéos de manière synchrone et les rediriger vers la carte ZCU102 (à base de FPGA) ou vers un ordinateur (à base de CPU ou de GPU). Cette plateforme d'expérimentation permet ainsi de conserver un flot flexible et de mieux répartir les tâches. Enfin, un algorithme de calibration a été mis en œuvre pour exploiter au mieux les informations de détection selon le positionnement des caméras. Celui-ci doit être encore révisé pour mieux combiner les détections de caméras multiples dans des configurations diverses.

À ce jour, les premières expérimentations ont mis en avant la difficulté de porter des algorithmes d'IA sur circuit FPGA. Bien que des outils existent, ceux-ci sont souvent encore en cours de développement et ne sont pas exécutables aisément. Aucun outil n'a été réellement développé pour permettre l'exploitation sur circuit FPGA de plusieurs entrées vidéo dans des applications de vision industrielle. Les étapes suivantes dans le cadre de cette thèse visent donc à la mise en œuvre et à l'évaluation d'un flot de conception exploitable pour la plateforme NATvision dans une approche multi-caméra.

[1] « NATvision - Accelerated Machine Vision ». (2020), adresse : <https://www.natvision.com/>.

[2] J. REDMON et A. FARHADI, « YOLOv3 : An Incremental Improvement », *arXiv*, 2018.

[3] Z. JIANG, L. ZHAO, S. LI et Y. JIA, « Real-time object detection method for embedded devices », 2011.

[4] G. MALIPEDDI, Q. HALL et J. KHURANA, « Kria K26 SOM : The Ideal Platform for Vision AI at the Edge White Paper », p. 9, 2021. adresse : https://www.xilinx.com/content/dam/xilinx/support/documentation/white_papers/wp529-som-benchmarks.pdf.

[5] A. KRIZHEVSKY, « Learning Multiple Layers of Features from Tiny Images », 2009.