

Machine learning embarqué pour microcontrôleurs

23 janvier 2024, 13:56

Frank Van den Broek

Cas pratique du machine learning embarqué : vision par ordinateur sur microcontrôleur Raspberry Pi Pico

Récemment, le domaine du machine learning « à la périphérie » (edge ML) était principalement utilisé pour l'inférence soit sur de vrais ordinateurs avec processeur graphique ou sur des téléphones mobiles. Mais avec l'émergence du machine learning « miniature » (TinyML), les réseaux neuronaux et outils d'analyse de données peuvent désormais aussi être déployés sur des microcontrôleurs. Pour montrer les possibilités du TinyML, Sirris a utilisé une application de vision par ordinateur sur un microcontrôleur bon marché : le Raspberry Pi Pico.

Le TinyML est une technologie multidisciplinaire en plein essor issue d'innovations dans les domaines de l'apprentissage automatique, des logiciels et des matériels intégrés ouvrant la voie à une nouvelle classe d'applications intelligentes via l'inférence « sur l'appareil » (près du capteur) sur des systèmes intégrés de très faible puissance. TinyML rend possible de nouveaux types d'applications et de services à la périphérie, qui ne dépendent pas d'un traitement à distance sur le cloud, mais utilisent l'inférence exécutée localement». Puisque les données sont traitées sur l'appareil, et ne sont donc pas transmises via un réseau, TinyML offre de grands avantages en termes de vie privée, latence, efficacité énergétique et fiabilité. La détection audio de mots d'activation sur des produits de géants de la technologie (« Okay Google », « Hey Siri ») était un

cas bien connu de première utilisation de TinyML sur des appareils alimentés par batterie allumés en permanence. Grâce aux avancées récentes dans les domaines des matériels enbarqués, modèles d'apprentissage automatique et plateformes de développement TinyML, le machine learning embarqué est désormais aussi à la portée d'autres industries pour toutes sortes d'applications utilisant des capteurs tels que caméras, accéléromètres ou capteurs de son.

Vision par ordinateur sur un microcontrôleur low-cost

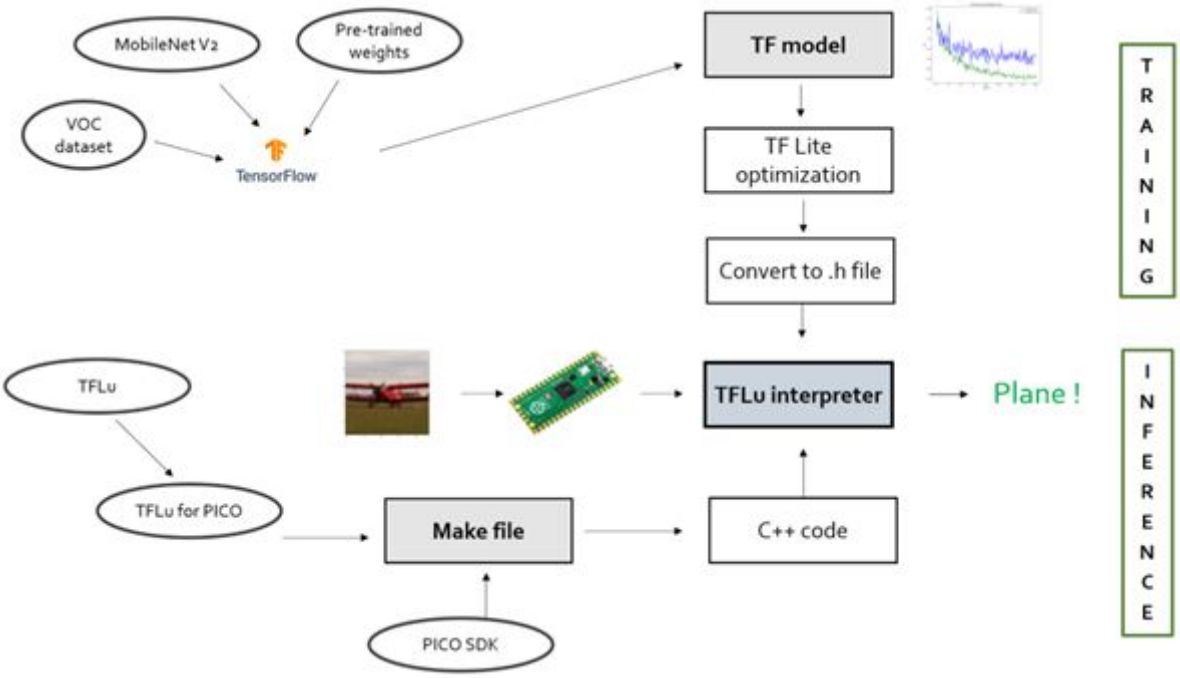
Pour montrer les possibilités de TinyML, Sirris a déployé une application de vision par ordinateur sur un microcontrôleur bon marché : le Raspberry Pi Pico (prix d'achat de moins de 5 EUR). L'objectif était d'évaluer les performances possibles, comme la précision et le temps d'inférence pour une tâche de classification d'image, avec des outils et modèles de vision par ordinateur open source.

La vision par ordinateur est une avancée majeure de la dernière décennie, rendue possible par les algorithmes d'apprentissage profond – des réseaux IA qui nécessitaient beaucoup de mémoire et puissance de calcul et ne pouvaient donc pas être utilisés sur des microcontrôleurs comme le Raspberry Pi Pico, qui possède seulement 264 Ko de mémoire SRAM et 2 Mo de mémoire flash, avec un processeur 133 MHz Arm Cortex-M0+. Mais récemment, de nouvelles architectures de réseau d'apprentissage profond ont été conçues spécialement pour les appareils mobiles et intégrés, qui donnent de bons résultats pour les tâches de vision par ordinateur, même avec des ressources limitées. Pour notre étude, nous avons utilisé MobileNetV2, un réseau neuronal convolutif pré-entraîné et libre d'accès, que nous avons davantage entraîné pour notre application par apprentissage par transfert avec l'ensemble des données provenant de VOC-2012 et de COCO.

Outils open source

Pour entraîner un modèle TinyML, on utilise un procédé semblable à ceux utilisés pour les gros modèles d'apprentissage profond. Dans notre cas, nous avons utilisé Tensorflow, un framework de ML open source de Google produisant un modèle entraîné prêt à passer en production sur un ordinateur ou un serveur.

Le déploiement de ce modèle sur un microcontrôleur requiert cependant de réduire les besoins en mémoire et puissance de calcul au moyen de différentes mesures d'optimisation, comme la quantisation, qui réduit la précision des entrées et paramètres représentés, et pruningl, qui supprime les neurones du réseau qui ont peu d'impact sur le résultat final. Pour ces optimisations, nous avons utilisé TensorFlow Lite, un autre outil open source de Google, que nous avons ensuite converti en fichier d'en-tête C/C++ et avons pu flasher sur le Raspberry Pi Pico. Pour l'interprétation du modèle sur le microcontrôleur, nous avons utilisé la bibliothèque C/C++ TensorFlow Lite for Microcontrollers (TFLu) de Google.



©Sirris

Résultats

Nous avons déployé sans problème une application de vision par ordinateur sur un microcontrôleur Raspberry Pi Pico bon marché avec des outils open source. Le modèle traite des images RVB avec une résolution de 48 x 48 pixels et les classifie dans 3 catégories avec un temps d'inférence de 250 ms.

Au cours de prochains travaux, nous explorerons les possibilités offertes par des microcontrôleurs plus puissants – disposant de plus de mémoire et de puissance de calcul que le Raspberry Pi Pico – en termes notamment de vision par ordinateur et de détection d'anomalies.

Webinaire

Le 31 janvier 2024, nous approfondirons la tâche de classification d'images décrite ci-dessus sur microcontrôleur Raspberry Pi Pico dans un webinaire gratuit de 60 minutes. Au programme de ce webinaire :

- Une discussion du matériel, du microcontrôleur et de l'appareil photo
- Une discussion de l'architecture MobileNet utilisée et une comparaison avec des réseaux alternatifs
- Une présentation des outils open source utilisés pour l'entraînement, l'optimisation et l'inférence : TensorFlow, TensorFlow Lite, TensorFlow Lite for Microcontrollers
- Une discussion des résultats.

Envie de participer ? Inscrivez-vous maintenant à notre webinaire !

Cet article est écrit dans le cadre de notre projet COOCK CORNET EmbedML, financé par le VLAIO.

Auteurs



Frank Van den Broek