MACHINE LEARNING FOR IOT

Développement d'une application basée sur de l'apprentissage automatique pour un système embarqué

VAP SEM
Systèmes Embarqués, Mobilité et objets communicants
2020 - 2021

Auteurs

Nimesh TAHALOOA

Encadrants

Cedric ADJIH (Inria)

Anis LAOUITI (TSP)

Bibliothèques

CONTEXTE

Forte croissance des objets connectés et des applications de machine learning

Traitement des données de ces objets principalement sur le cloud

Souhait de rapprocher les centres de traitement de données aux sources (Edge Computing)

Réduction de latence, économie de bande passante, plus de confidentialité

Transition contrainte par les limitations des systèmes embarqués

Cas d'application : la détection de feu

Feux de forêts plus violents et fréquents

Système de surveillance pour la maison

NVIDIA JETSON NANO

TensorRT

Mini-ordinateur avec un GPU, préchargé avec des bibliothèques d'accélération d'algorithmes de ML

Compression des réseaux de neurones (Clustering, Pruning, Quantification) avec TensorRT

Evaluation de la précision d'inférence du modèle optimisé

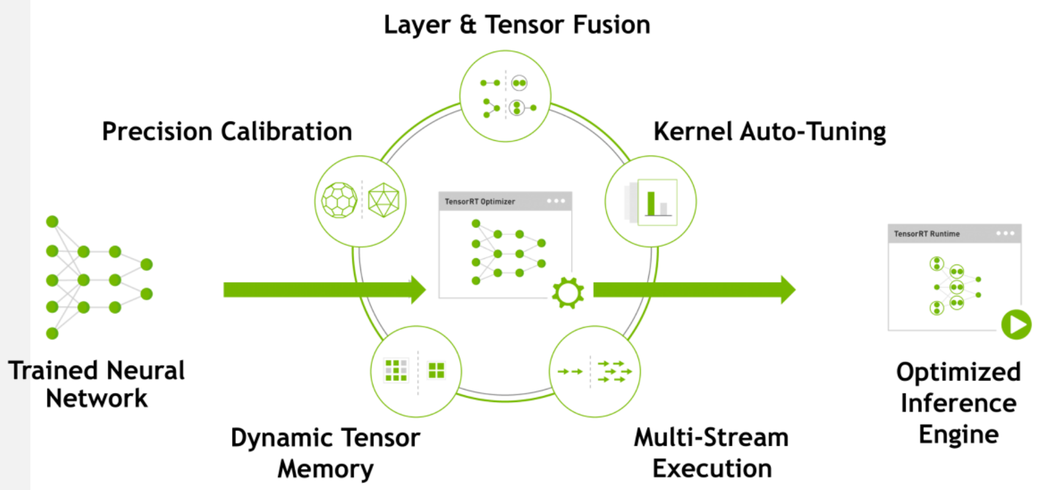


Schéma de fonctionnement de TensorRT (*nvidia.com*)

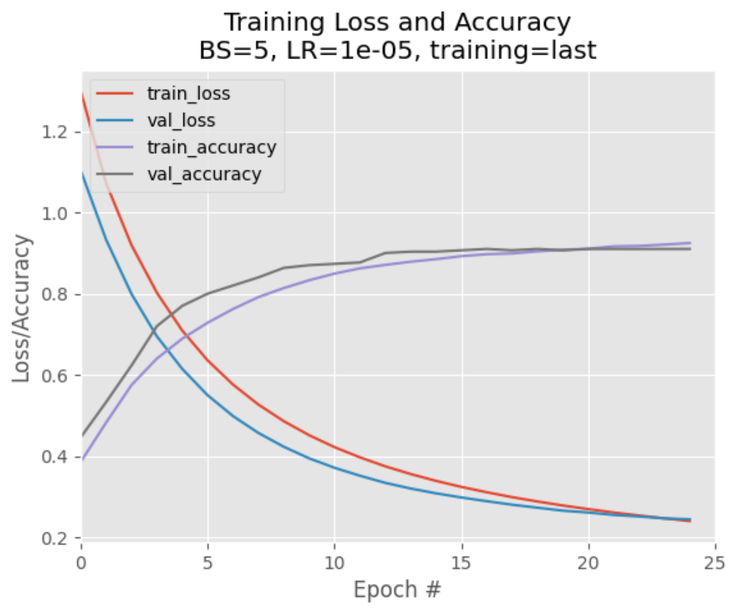
RÉSULTATS

Convergence rapide d'un modèle MobileNet avec 91% de précision d'inférence, entraîné sur la Jetson Nano

Optimisation du modèle avec TensorRT : 10 minutes

Entraînement limitée principalement par la RAM, partagée entre le CPU et le GPU. C'est plus difficile de charger et faire converger des modèles profonds.

Exemple d'inférence (Fumée:74%, Feu: 23%)



Courbes d'entraînement et matrice de confusion du modèle MobileNet, en transfer learning

Contact nimesh.tahalooa@telecom-sudparis.eu

https://github.com/nimesh-t/pfe-ml-for-iot