

6.2 - Machine learning avec CUDA

Le machine learning Immich fournit la recherche intelligente et la reconnaissance faciale. CUDA est pris en charge sur les GPU NVIDIA avec une capacité de calcul 5.2 ou supérieure.

Conditions actuelles

- GPU NVIDIA compatible ;
- pilote officiel NVIDIA 545 ou plus récent ;
- accès au GPU depuis le conteneur ;
- image machine learning CUDA ;
- mémoire GPU suffisante pour le modèle.

La Quadro P2000 répond à la condition de capacité de calcul et ton pilote 570.172.08 dépasse le minimum documenté.

Docker Compose

Le service doit utiliser une image se terminant par `-cuda` et réserver un GPU. Exemple conceptuel :

```
immich-machine-learning:
  image: ghcr.io/immich-app/immich-machine-learning:${IMMICH_VERSION:-release}-cuda
  deploy:
    resources:
      reservations:
        devices:
          - driver: nvidia
            count: 1
            capabilities:
              - gpu
```

Utiliser de préférence les fichiers officiels `hwaccel.ml.yml` correspondant à la version Immich.

TrueNAS

Dans les paramètres de l'application :

1. activer le GPU pour le machine learning ;
2. sélectionner la P2000 ;
3. redéployer ;
4. vérifier le conteneur ;
5. lancer une tâche de visages ou recherche intelligente.

Vérifier

```
sudo docker exec NOM_CONTENEUR_ML nvidia-smi  
sudo docker logs --tail 200 NOM_CONTENEUR_ML
```

Chercher :

```
CUDAExecutionProvider  
Available ORT providers
```

Surveiller :

```
watch -n 1 nvidia-smi
```

Gestion de la VRAM

La P2000 possède 5 Go de VRAM. Éviter de lancer simultanément un import massif, plusieurs transcodages et de nombreux traitements ML sans surveillance.

Réduire si nécessaire :

- le nombre de workers ;
- la concurrence des tâches ;
- la taille ou complexité du modèle ;
- les transcodages parallèles.

Après activation

Il n'est pas nécessaire de relancer tous les anciens jobs uniquement pour profiter de l'accélération. Les prochaines tâches utiliseront le backend disponible.

Référence : <https://docs.immich.app/features/ml-hardware-acceleration/>

Revision #1

Created 2026-08-08 02:04:18 CEST by Admin

Updated 2026-08-08 02:04:18 CEST by Admin