

24-25 Mai 2024

Certificat d'université en

Intelligence Artificielle

HackIA25

Atelier d'Intelligence Artificielle (I-ISIA-202)

Edge AI System for smart homes



Stéphane DUPONT

<https://hackia.eu/>

Contact: stephane.dupont@umons.ac.be

- I. Introduction & Programme du Workshop**
 - II. Objectif du workshop**
 - III. Tâche 0 : collecte et préparation des données**
 - IV. Tâche 1 : choix, développement et entraînement des modèles**
 - V. Tâche 2 : intégration du dispositif multi-caméra robotisé, et ressource Edge : Jetson Xavier**
 - VI. Tâche 3 : optimisation/compression des modèles**
 - VII. Conclusions**
-

Programme du Workshop

24 Mai 2025

08h00 à 09h00	Accueil des participants
09h00 à 10h00	Présentation du challenge Pr. Stéphane Dupont et pause-café
10h00 à 13h00	Session de travail N° 01
13h00 à 14h00	Lunch
14h00 à 16h45	Session de travail N° 02
16h45 à 17h00	Pause-café
17h00 à 20h00	Session de travail N° 03
20h00 à 21h30	Dîner
21h30 à 23h00	Activité jeu

Programme du Workshop

25 Mai 2025



07h00 à 08h30	Petit déjeuner
08h30 à 10h45	Session de travail N° 04
10h45 à 11h00	Pause-café
11h00 à 13h00	Session de travail N° 05
13h00 à 14h00	Lunch Sandwich
14h00 à 17h00	Préparation des démos et pitches
17h00 à 17h30	Pause-café
17h30 à 18h30	Présentations devant jury
18h30 à 19h30	Remise des prix, cérémonie de clôture et cocktail

Coachs Techniques |



Titulaire

Pr. Stéphane Dupont

UMONS



Ir. Aurélie Cools

UMONS



Mr. Maksym Iakovenko

UMONS



Mr. Etienne Julien

UMONS



Membres de Jury

Membres du Jury |



Pr. Thierry Dutoit

UMONS



Pr. Pierre Manneback

UMONS



Pr. Xavier Siebert

UMONS



Pr. Sidi Ahmed Mahmoudi

UMONS

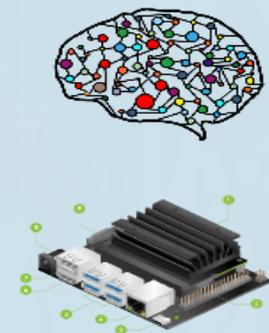


Introduction

- Mise en œuvre et exploitation des connaissances acquises lors des défis IA
- Déploiement de modèles Deep Learning dans des applications concrètes

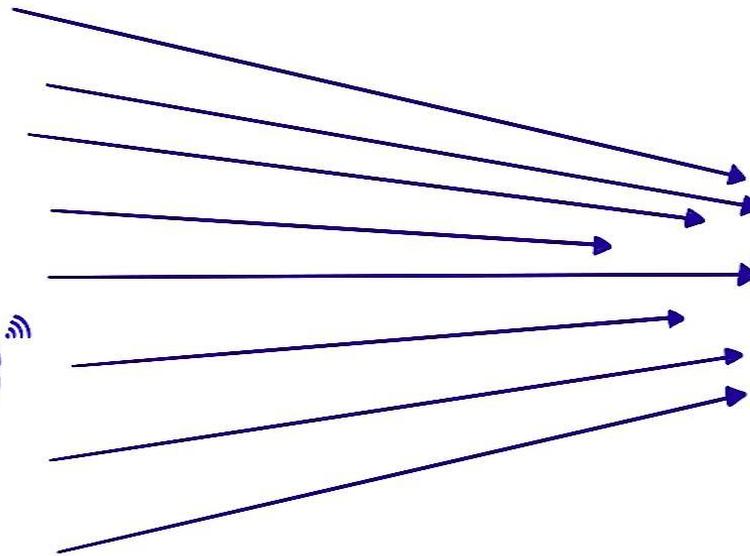
« **modules pour maisons intelligentes** »

- Application des modèles sur des flux vidéo e temps réel
- Déploiement de solutions Deep Learning sur ressources Edge AI

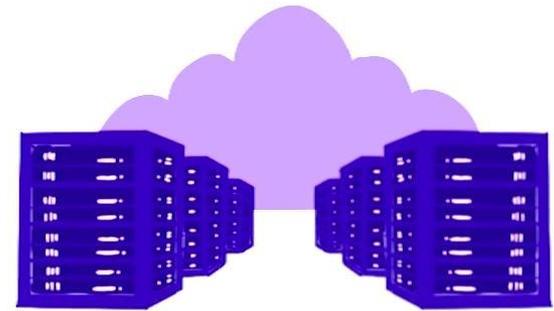


Que signifie Edge Computing ?

OBJETS CONNECTÉS



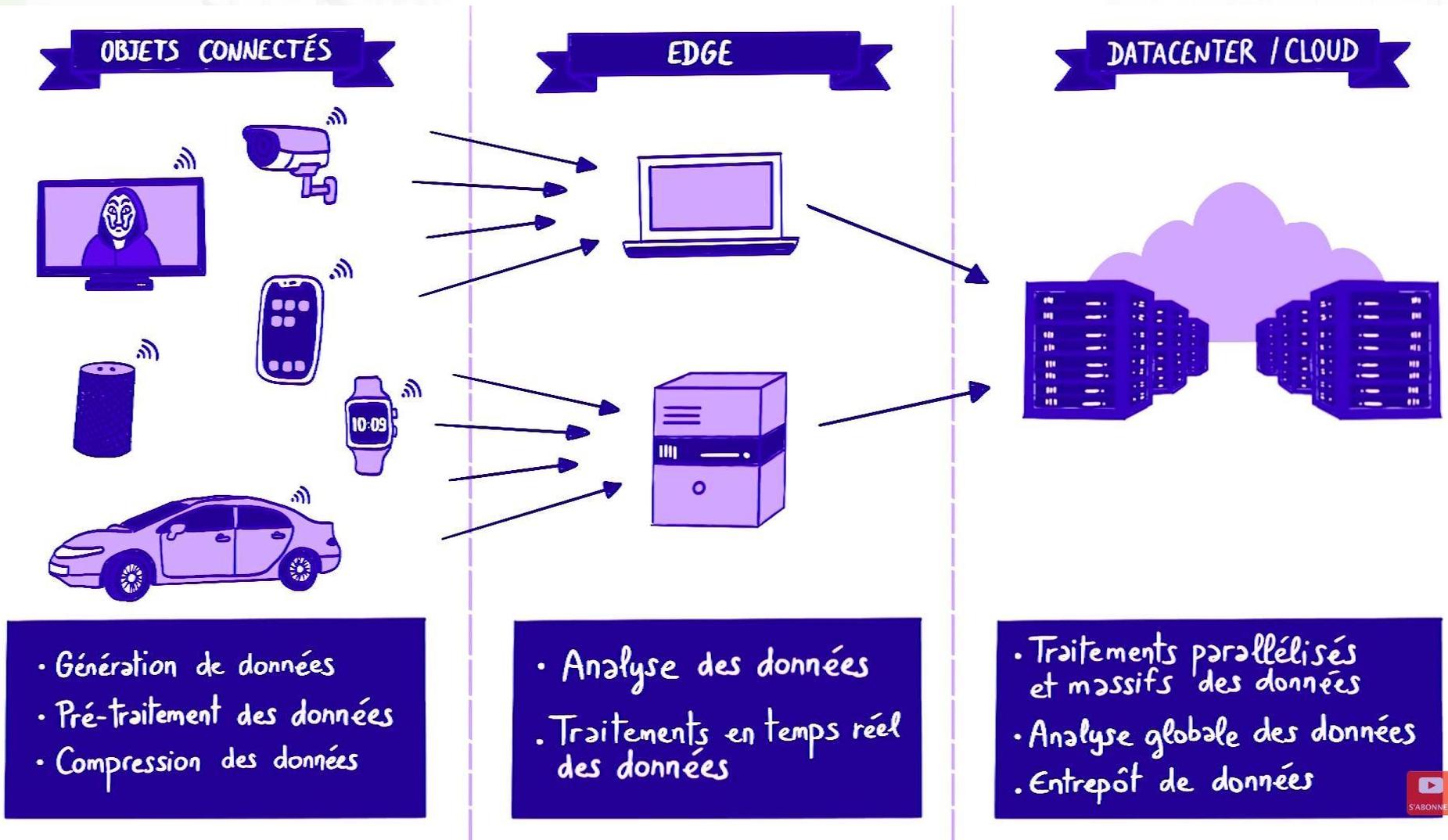
DATACENTER / CLOUD



- Saturation de la bande passante
- Augmentation de la latence de traitement des données

Vie privée, confidentialité, protection des données

Que signifie Edge Computing ?



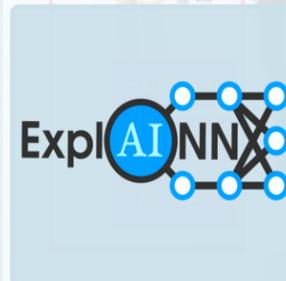
- I. Introduction & Programme du Workshop
 - II. Objectif du workshop
 - III. Tâche 0 : collecte et préparation des données
 - IV. Tâche 1 : choix, développement et entraînement des modèles
 - V. Tâche 2 : intégration du dispositif multi-caméra robotisé, et ressource Edge : Jetson Xavier
 - VI. Tâche 3 : optimisation/compression des modèles
 - VII. Conclusions
-

Objectif du Workshop

- **Développer** de réseaux de neurones profonds pour localiser des objets et des personnes sur des images, reconnaître visages, etc.

Dispositif EdgeAI avec 3 caméras à votre disposition

- **Intégrer** les modèles dans un système Edge IA pour flux vidéo en temps réel, et en particulier:
 - portage de la solution sur la ressource Edge IA « **Jetson Xavier** »
 - pilotage d'une camera robotisée.
- **Optimiser** et explorer le compromis entre **précision, temps de calcul, espace mémoire**



FEMALE
ADULT
MOVE



Objectif du Workshop

- **Projet impliquant quatre tâches:**

0 - Collecte et préparation/annotation des données ;

1 - Développement et entraînement des modèles Deep Learning ;

2 - Portage sur ressource Edge : Jetson Xavier ;

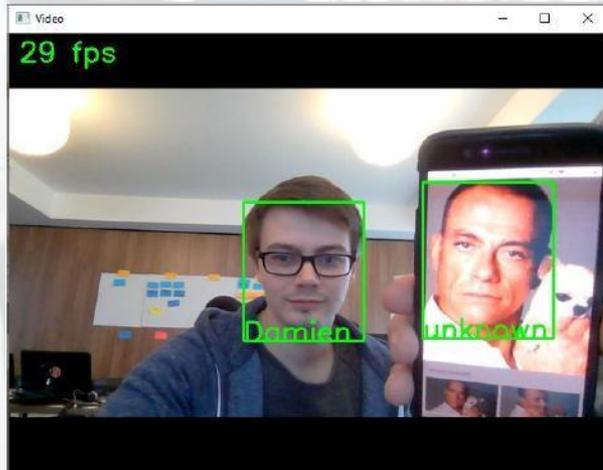
3 - Optimisation/compression des modèles ;

- I. Introduction & Programme du Workshop
 - II. Objectif du workshop
 - III. Tâche 0 : collecte et préparation des données
 - IV. Tâche 1 : choix, développement et entraînement des modèles
 - V. Tâche 2 : intégration du dispositif multi-caméra robotisé, et ressource Edge : Jetson Xavier
 - VI. Tâche 3 : optimisation/compression des modèles
 - VII. Conclusions
-

- Programme imposé



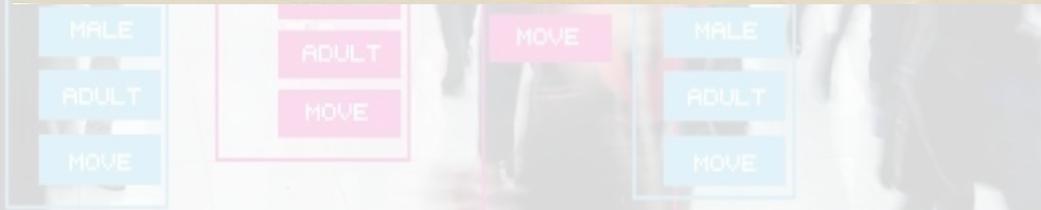
- 3ème au choix



Collecte et préparation des données

Détection de chute – relativement **rare** dans les données d'apprentissage

Détection d'objets "perdus" – de **taille relativement petite** et rares



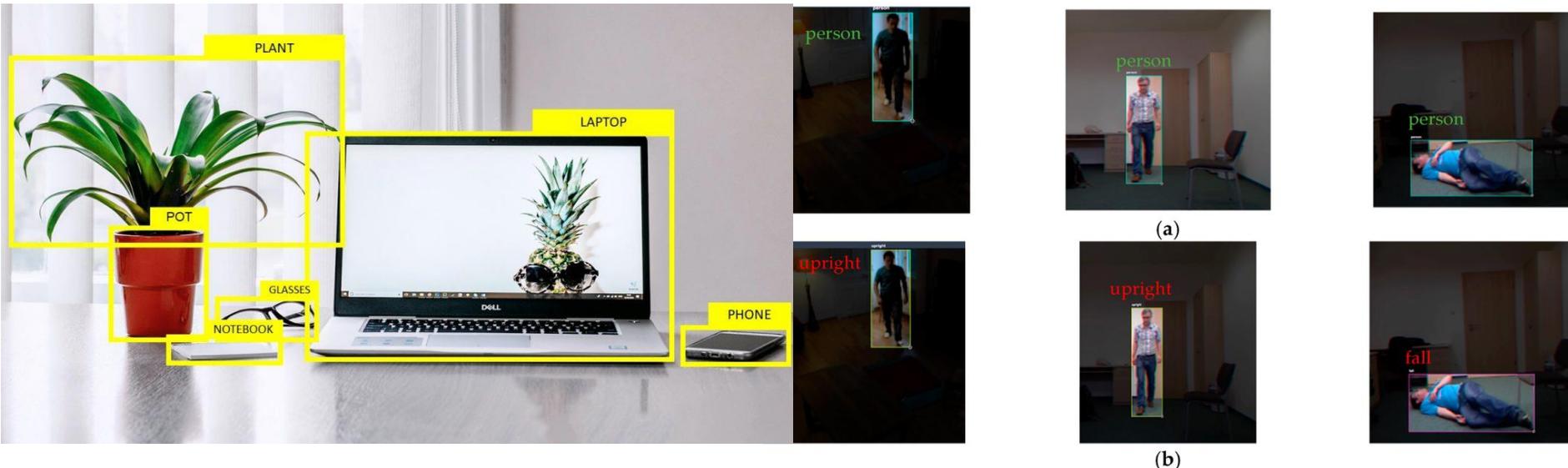
Collecte et préparation des données

1. Photographier et filmer un maximum

- Chutes (on se contentera de simuler) => donner de sa personne
- Objets plus ou moins isolés. Quatre categories
 - trousseau clé, porte-feuille, lunettes, smartphone

2. Annoter ces images

- Rectangles englobants les objets + etiquette (type de posture de la personne, type d'objet)
- Points spécifiques sur le corps de la personne (nez, épaule gauche, etc.)



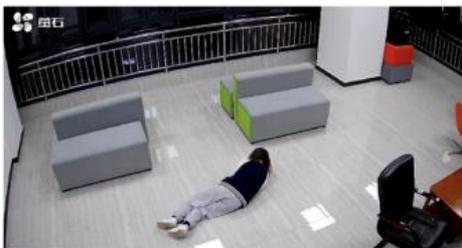
Collecte et préparation des données

1. Photographier et filmer un maximum

- Chutes (on se contentera de simuler) => donner de sa personne
- Objets plus ou moins isolés. Quatre categories
 - trousseau clé, porte-feuille, lunettes, smartphone

2. Annoter ces images

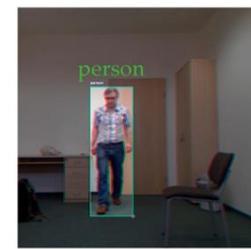
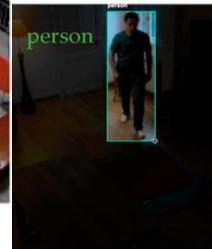
- Rectangles englobants les objets + etiquette (type de posture de la personne, type d'objet)
- Points spécifiques sur le corps de la personne (nez, épaule gauche, etc.)



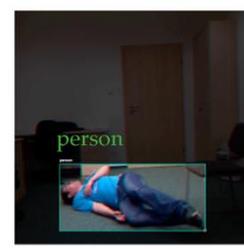
(a) example of our own dataset



(b) fall detection result



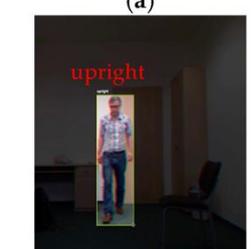
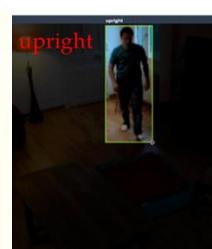
(a)



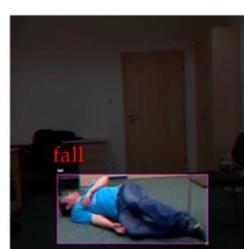
(a) human pose estimation



(a) optical flow extraction



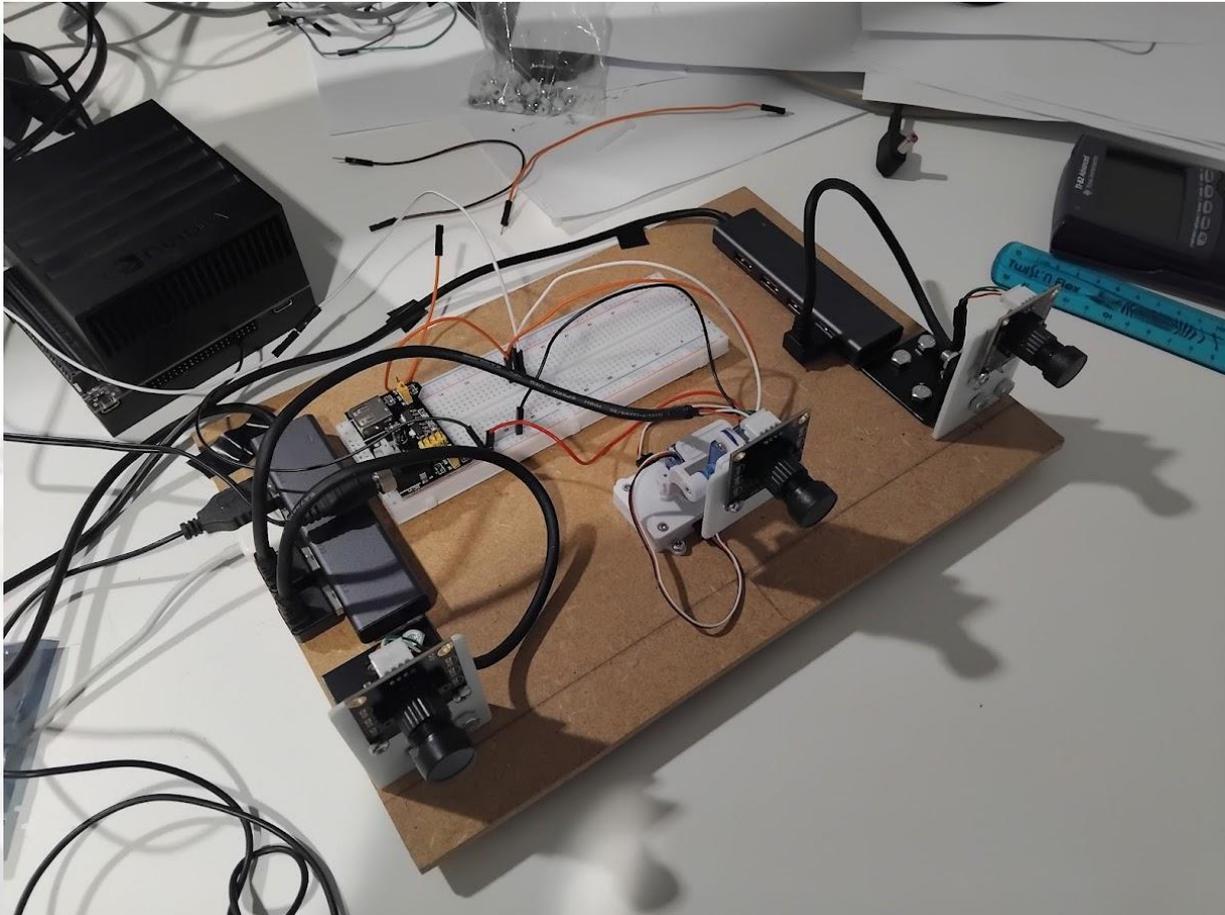
(b)



- I. Introduction & Programme du Workshop
 - II. Objectif du workshop
 - III. Tâche 0 : collecte et préparation des données
 - IV. Tâche 1 : choix, développement et entraînement des modèles
 - V. Tâche 2 : intégration du dispositif multi-caméra robotisé, et ressource Edge : Jetson Xavier
 - VI. Tâche 3 : optimisation/compression des modèles
 - VII. Conclusions
-

- I. Introduction & Programme du Workshop
 - II. Objectif du workshop
 - III. Tâche 0 : collecte et préparation des données
 - IV. Tâche 1 : choix, développement et entraînement des modèles
 - V. Tâche 2 : intégration du dispositif multi-caméra robotisé, et ressource Edge : Jetson Xavier
 - VI. Tâche 3 : optimisation/compression des modèles
 - VII. Conclusions
-

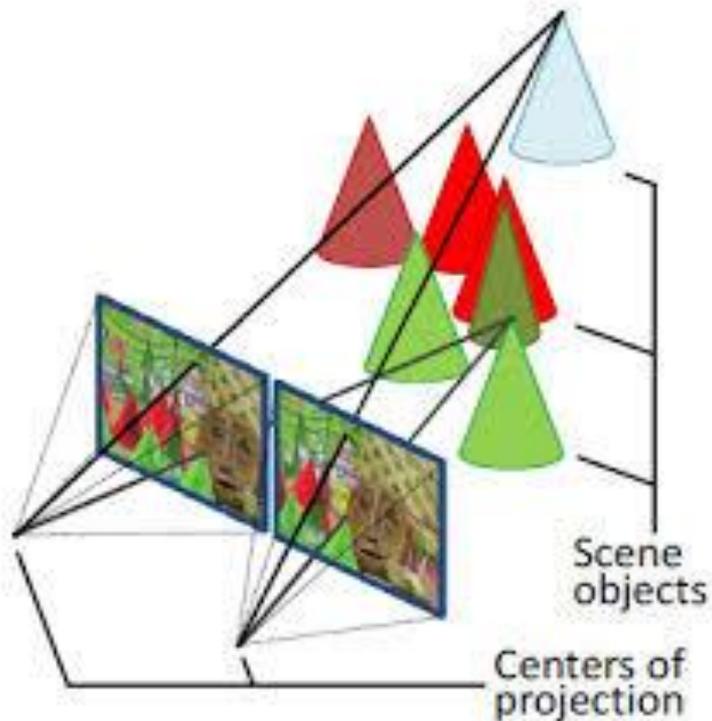
Edge AI for Smart Cities



Opportunités

- Multiple image et méthodes d'ensemble

Code de démarrage sera fourni

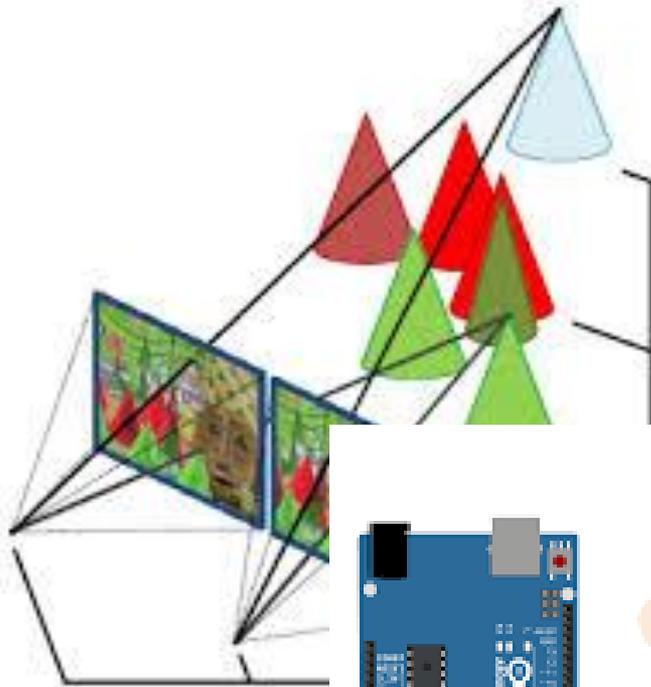


Opportunités

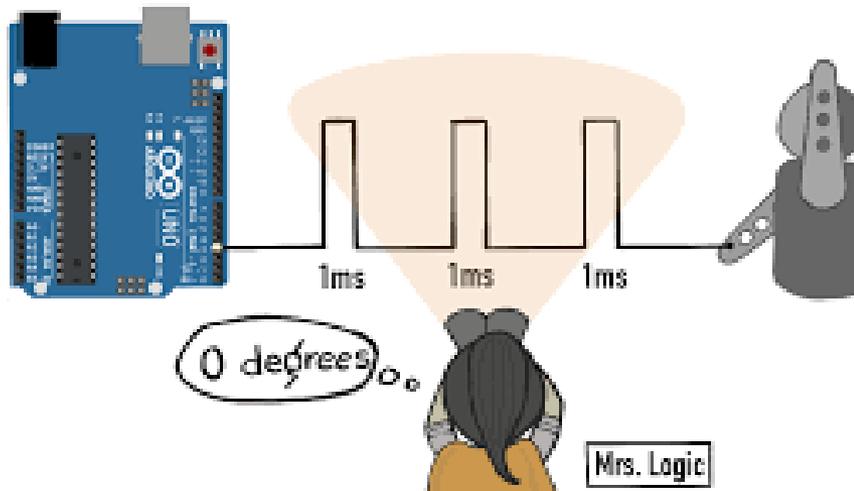
- Multiple image et méthodes d'ensemble
- Localisation 3D par stéréoscopie

Code de démarrage sera fourni

Intégration senseurs, robot, jetson



SERVO MOTOR CONTROL



Opportunités

- Multiple image et méthodes d'ensemble
- Localisation 3D par stéréoscopie
- Diriger la caméra (et potentielle zoomer) sur une zone d'intérêt grâce au robot pan-tilt

Code de démarrage sera fourni

- I. Introduction & Programme du Workshop
 - II. Objectif du workshop
 - III. Tâche 0 : collecte et préparation des données
 - IV. Tâche 1 : choix, développement et entraînement des modèles
 - V. Tâche 2 : intégration du dispositif multi-caméra robotisé, et ressource Edge : Jetson Xavier
 - VI. Tâche 3 : optimisation/compression des modèles
 - VII. Conclusions
-

Edge AI

- **Optimiser et accélérer** l'algorithme : temps d'inférence des modèles
- Privilégier l'utilisation des réseaux de neurones de **plus petites tailles**
- **Compresser** les modèles développés: pruning, quantification, distillation de connaissances, etc.



Vous pouvez combiner les méthodes de compression

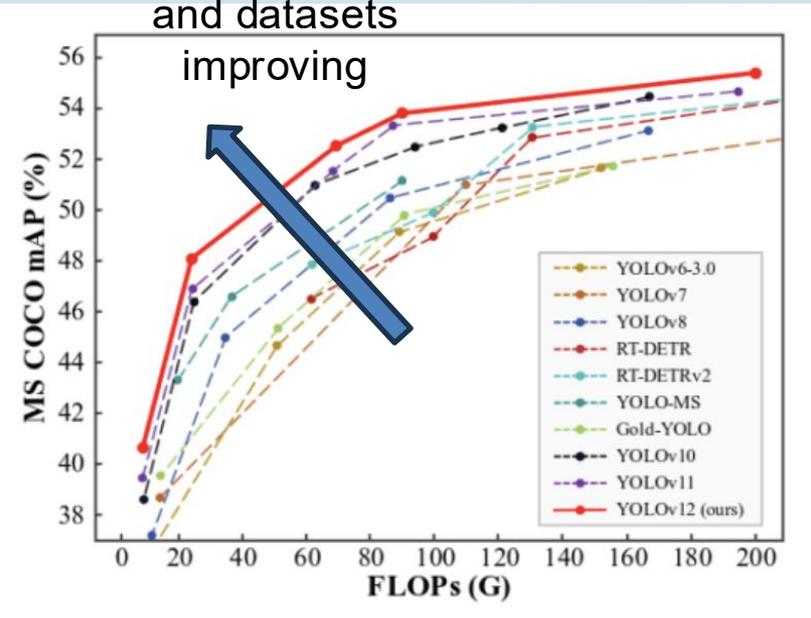
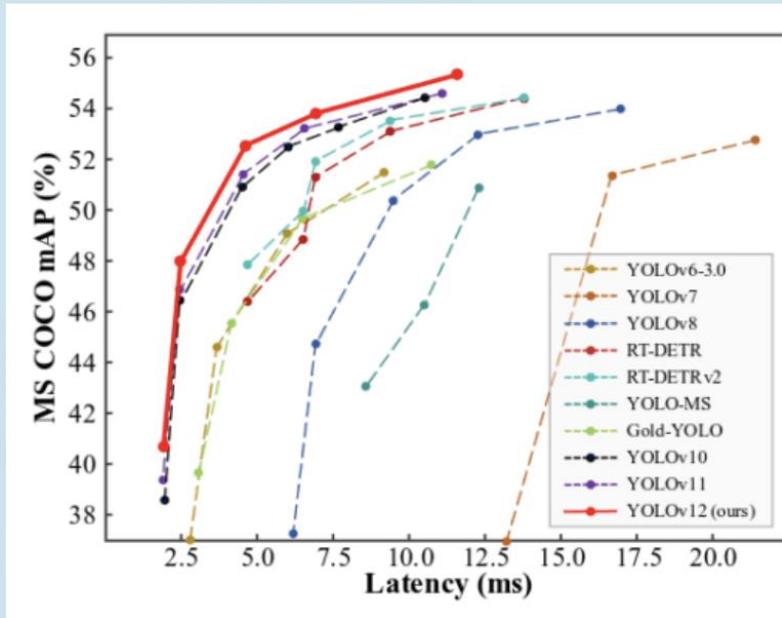
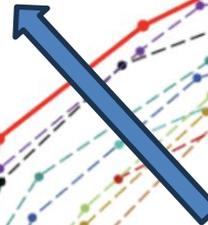
- **Objectif** : meilleur compromis entre précision, temps de calcul espace mémoire requis.

Edge AI

ML

methodologies
and datasets

improving



- **Objectif** : meilleur compromis entre précision, temps de calcul espace mémoire requis.

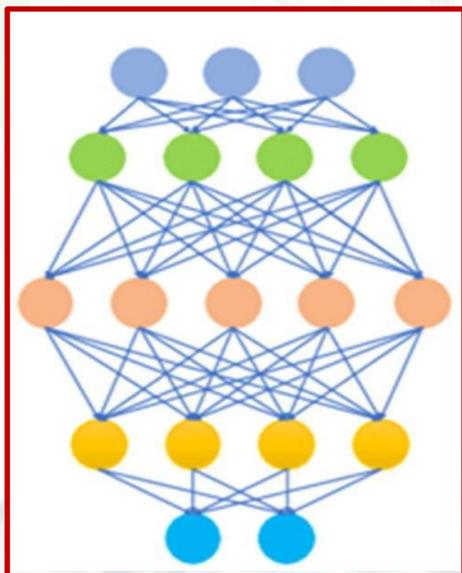
Main approaches of Edge AI in Deep Learning

- a. Pruning
- b. Quantization
- c. Knowledge Distillation



Related work : Pruning

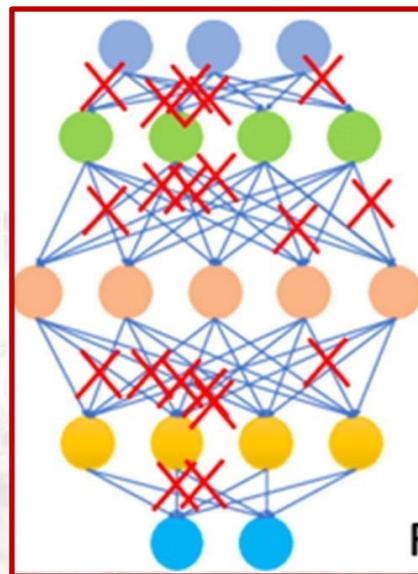
Input



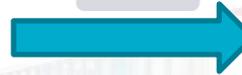
Dense Network



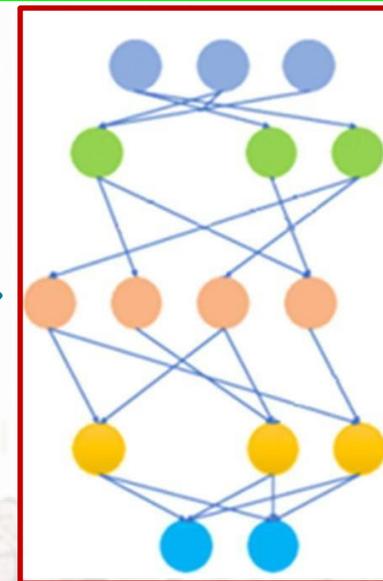
Weight pruning



Sparse Network



Node pruning



Dense Network

Several elements to analyze before pruning :

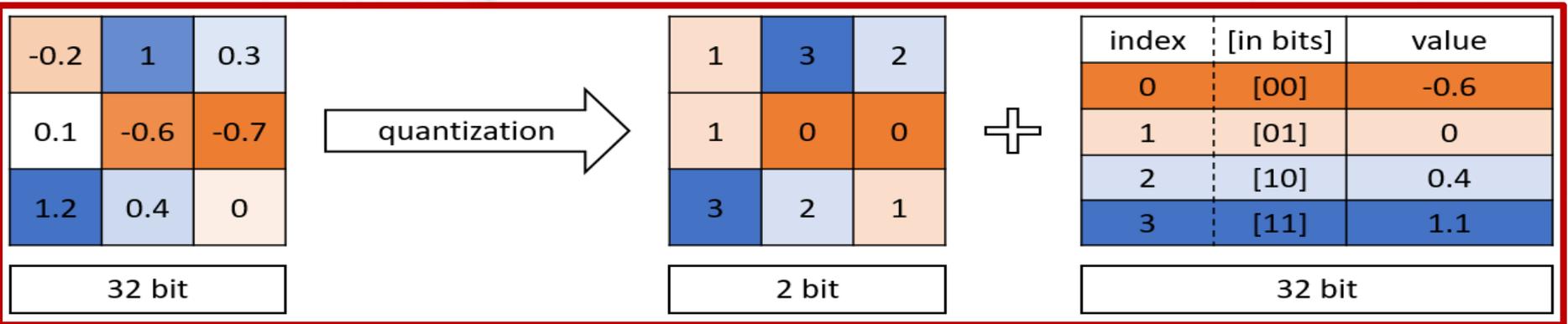
- Pruning methods
- Pruning-zone application
- Neron pruning or/and weight pruning
- Pruning scheduling
- Pruning impact on memory consumption
- Pruning impact on model size
- Pruning impact on computation time

Main approaches of Edge AI in Deep Learning

- a. Pruning
- b. Quantization
- c. Knowledge Distillation

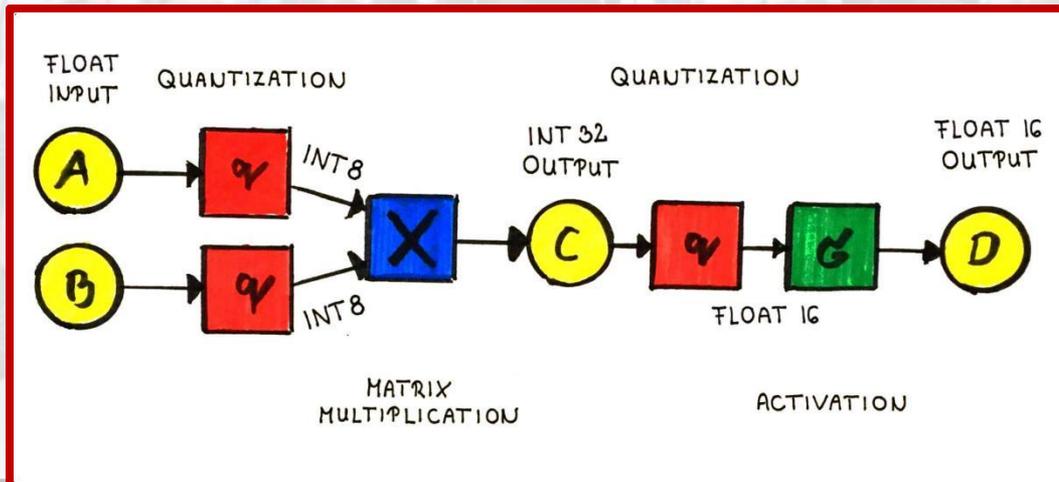


Related work : Quantization

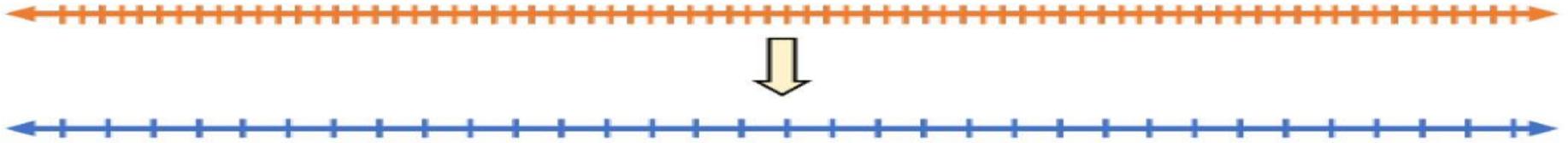


Low precision arithmetic:

<https://www.cs.cornell.edu/courses/cs4787/2020sp/lectures/Lecture21-Slides.pdf>



Related work : Quantization



Benefits:

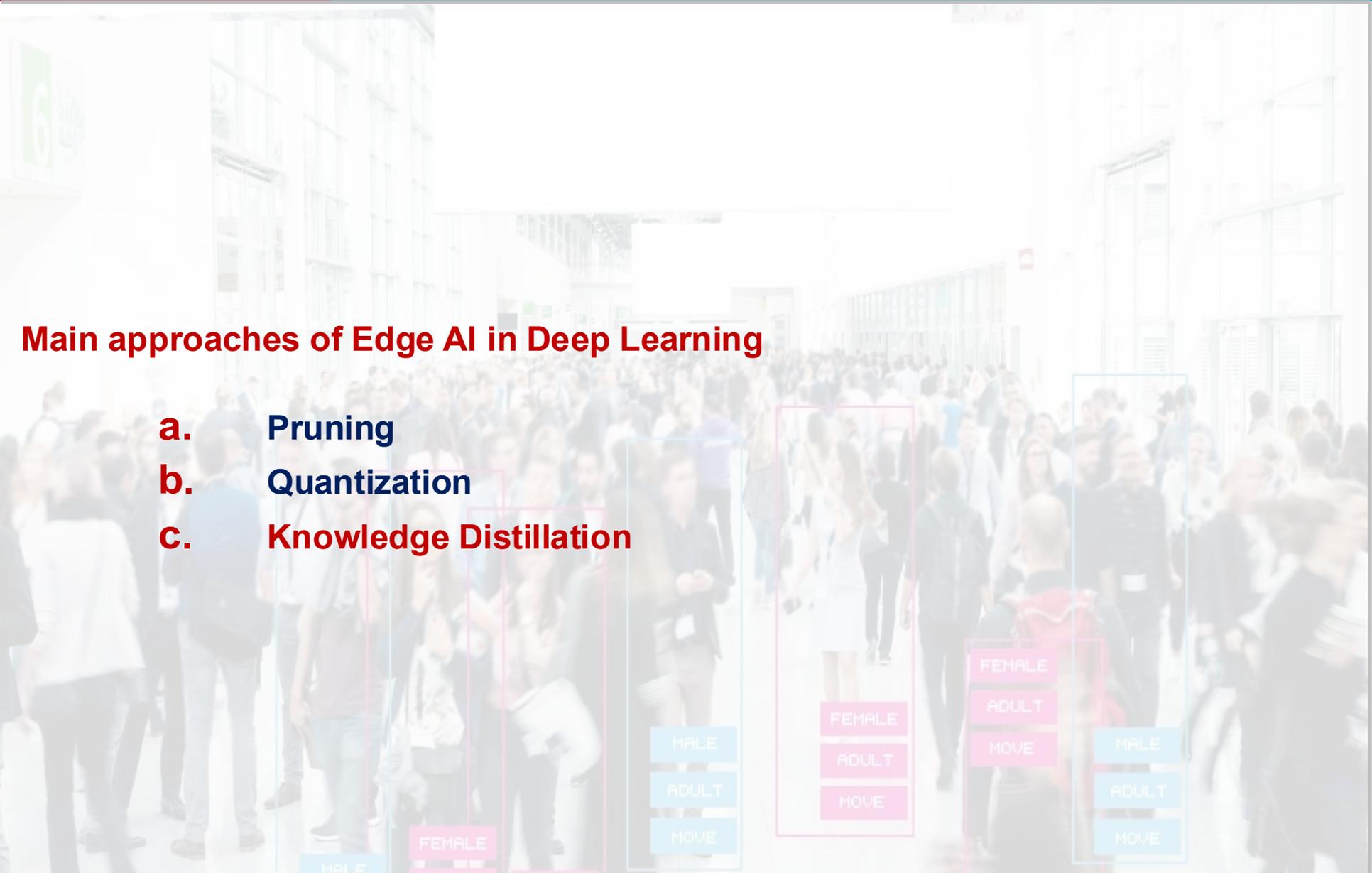
- Faster arithmetic operations
- Reduction in model size
- Compatibility with more (and less) devices

When to apply ?

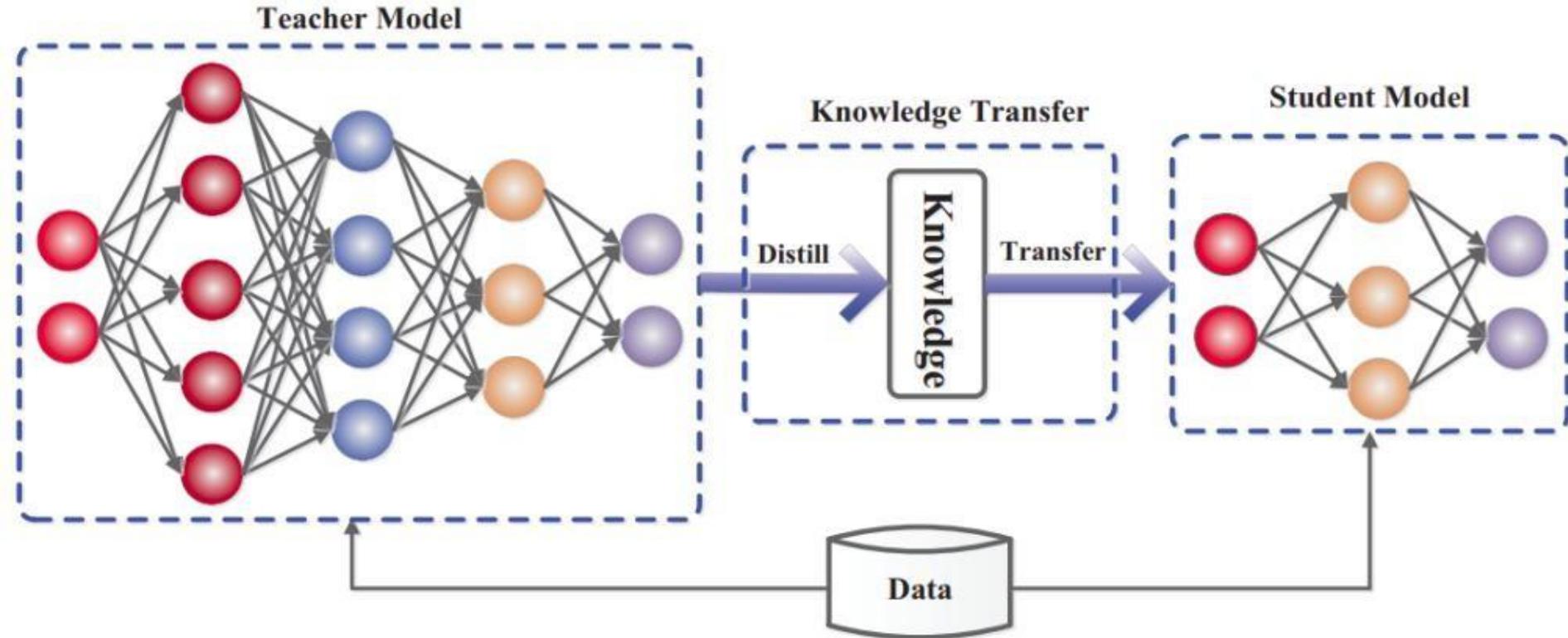
- **Static Post training quantization - PTQ** : quantization of weights/activations (8 bit)
- **Quantization Aware Training – QAT**
- **Dynamic Quantization** : quantization of weights only (both fp16 and int8)

Main approaches of Edge AI in Deep Learning

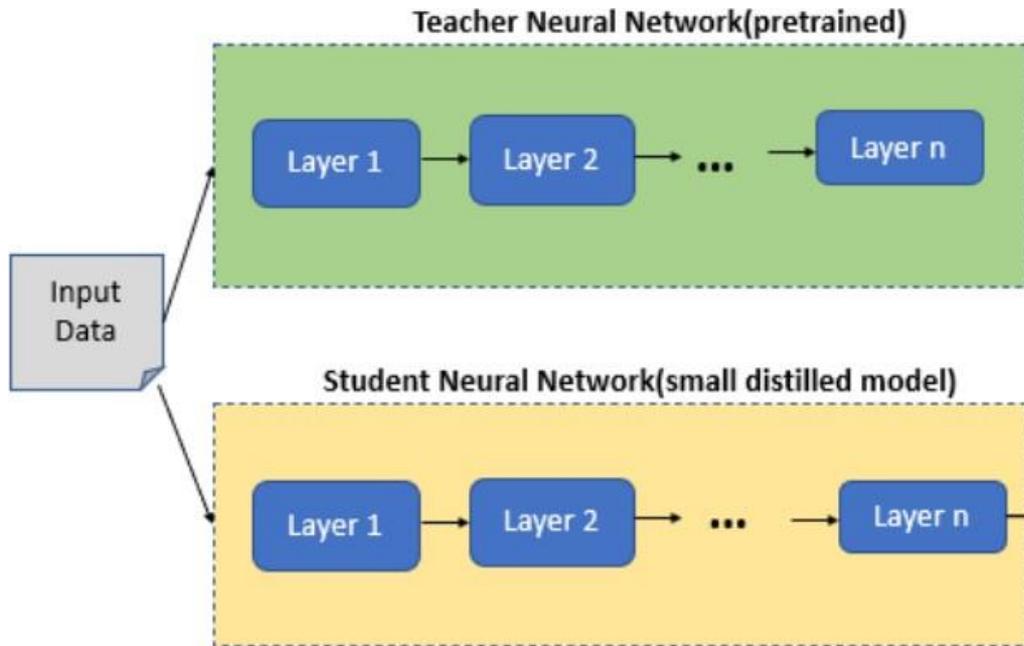
- a. Pruning
- b. Quantization
- c. Knowledge Distillation



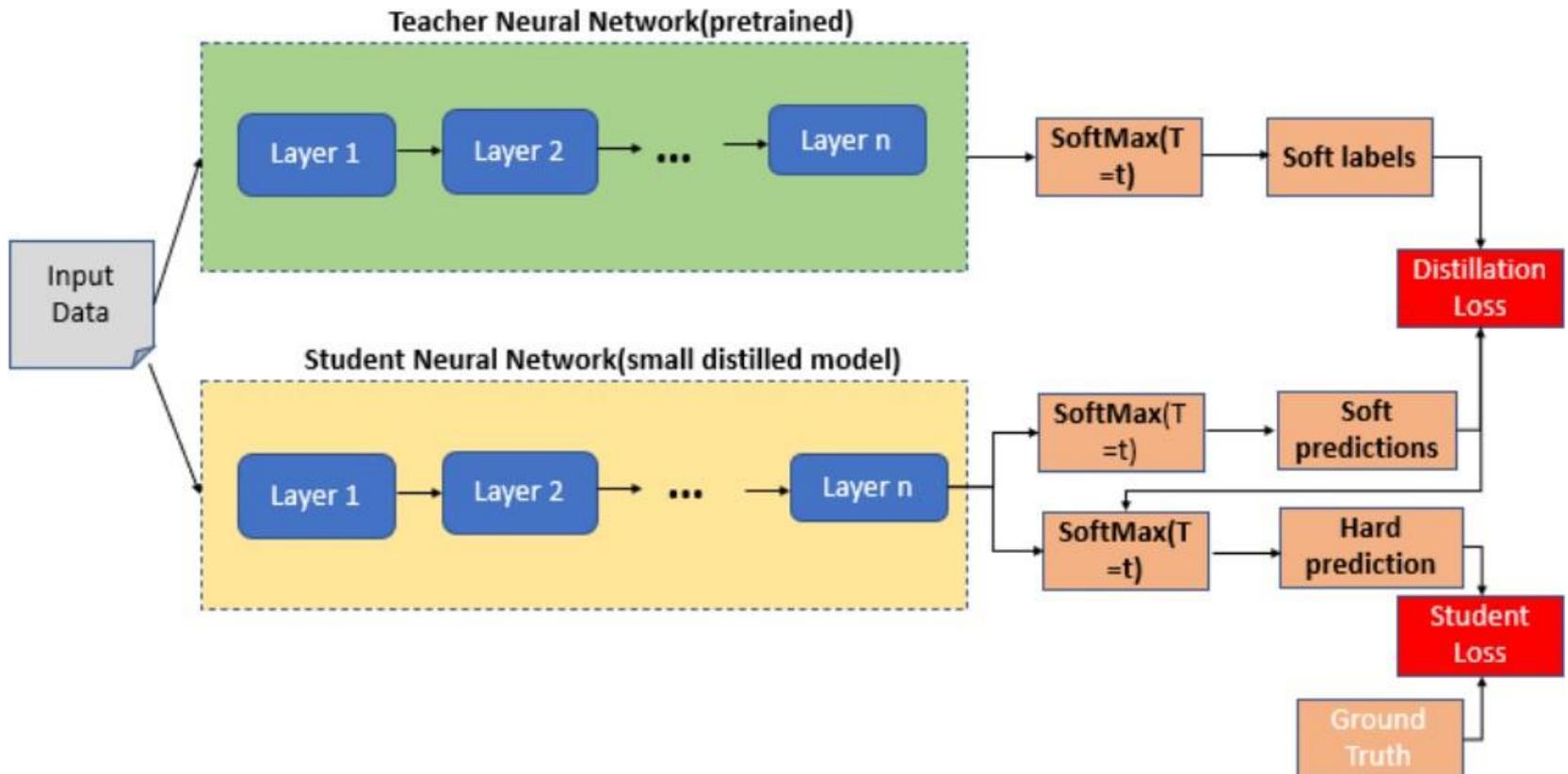
Knowledge Distillation



Knowledge Distillation : Process



Knowledge Distillation : Process



Matériel

Cloud : Google Colab, Google Colab pro, vast ai, etc.

Edge: Environnement « Deeplearning » préconfiguré

Matériel embarqué

[Jetson Xavier](#)

[Jetson Nano](#)

[Movidius](#) (avec Raspberry PI)

[Google Coral](#)

Développement

Git : outils de gestion de versions, de partage de code, etc.



Collaboration

Rejoindre le groupe [Teams](#)

Partage de bases de données annotées et autres



Espace de travail

**Espace
d'intégration et test**

Espace de travail

**Espace
d'intégration et test**

**Vos chambres – mais
avec modération**

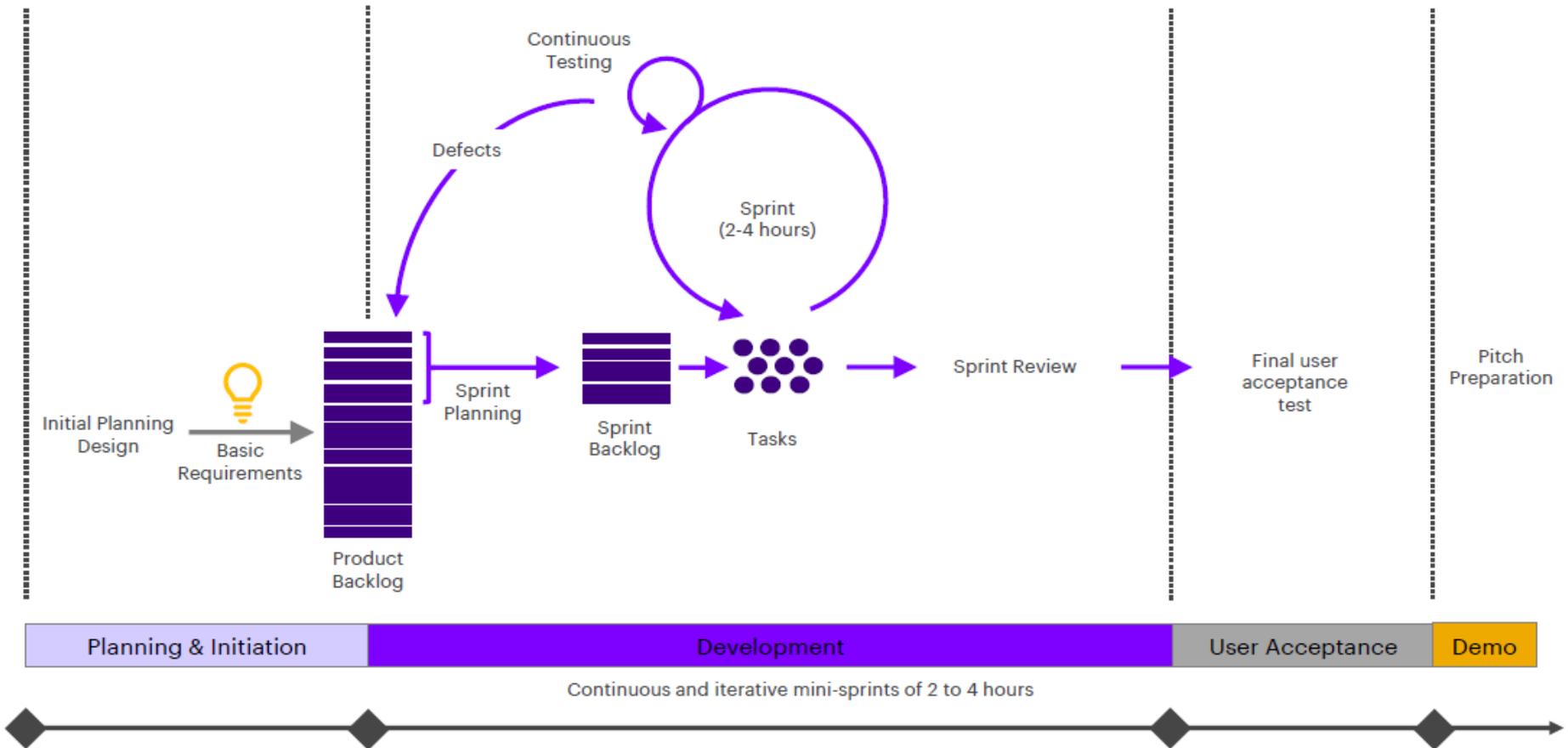
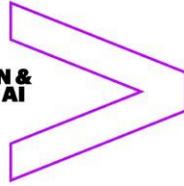
Jetson Xavier



GPU	512-core Volta GPU with Tensor Cores
CPU	8-core ARM v8.2 64-bit CPU, 8MB L2 + 4MB L3
Memory	32GB 256-Bit LPDDR4x 137GB/s
Storage	32GB eMMC 5.1
DL Accelerator	(2x) NVDLA Engines
Vision Accelerator	7-way VLIW Vision Processor
Encoder/Decoder	(2x) 4Kp60 HEVC/(2x) 4Kp60 12-Bit Support
Size	105 mm x 105 mm x 65 mm
Deployment	Module (Jetson AGX Xavier)

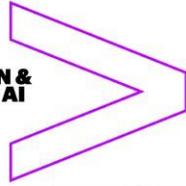
Organisation du travail en équipe

HACKATHON & WORKSHOP AI



Organisation du travail en équipe

HACKATHON &
WORKSHOP AI



Stuff to do (backlog)

Data quality checks

Prepare pitch

Integrate library xx

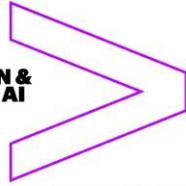
To do

Ongoing

Done

Organisation du travail en équipe

HACKATHON &
WORKSHOP AI



Stuff to do (backlog)

Data quality checks

Prepare pitch

Integrate library xx

To do

Data quality checks

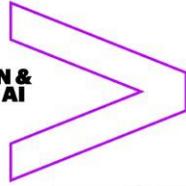
Ongoing

Data quality checks
Thibault

Done

Organisation du travail en équipe

HACKATHON &
WORKSHOP AI



Stuff to do (backlog)

Integrate
library xx

To do

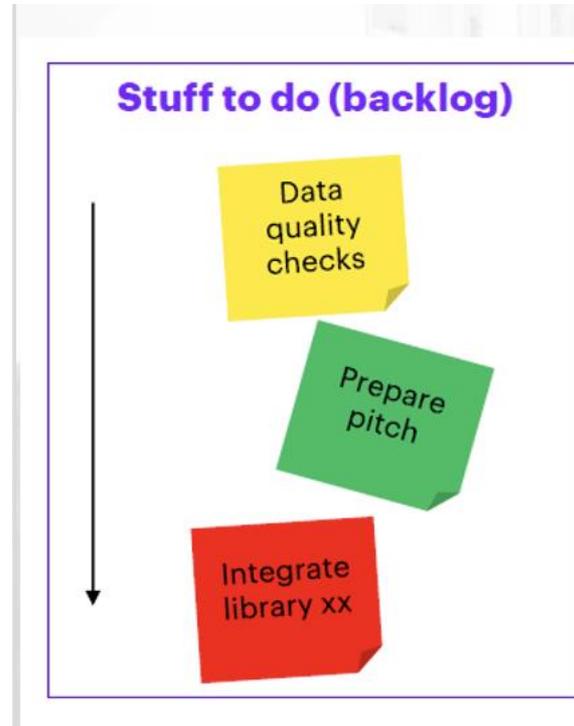
Ongoing

Data
quality
checks
Thibault

Done

Prepare
pitch
Marc

Liste des choses à faire: data collection, etc...



Objectif du Workshop

- **Projet impliquant quatre tâches:**

0 - Collecte et préparation/annotation des données ;

1 - Développement et entraînement des modèles Deep Learning ;

2- Portage sur ressource Edge : Jetson Xavier ;

3 - Optimisation/compression des modèles ;



Programme du Workshop

24 Mai 2025

08h00 à 09h00	Accueil des participants
09h00 à 10h00	Présentation du challenge Pr. Stéphane Dupont et pause-café
10h00 à 13h00	Session de travail N° 01
13h00 à 14h00	Lunch
14h00 à 16h45	Session de travail N° 02
16h45 à 17h00	Pause-café
17h00 à 20h00	Session de travail N° 03
20h00 à 21h30	Dîner
21h30 à 23h00	Activité jeu

Programme du Workshop

25 Mai 2025



07h00 à 08h30	Petit déjeuner
08h30 à 10h45	Session de travail N° 04
10h45 à 11h00	Pause-café
11h00 à 13h00	Session de travail N° 05
13h00 à 14h00	Lunch Sandwich
14h00 à 17h00	Préparation des démos et pitches
17h00 à 17h30	Pause-café
17h30 à 18h30	Présentations devant jury
18h30 à 19h30	Remise des prix, cérémonie de clôture et cocktail

Edge AI for Smart Home

- **Module Edge AI pour maisons intelligentes**
- **Intégration de modèles offrant une haute précision**
- **Traitement temps réel à partir de flux vidéo multiples**
- **Portage sur matériel Edge (optimisation, réduction de mémoire, etc.)**
- **D'autres pistes :**
 - **personnalisation de l'interface graphique**
 - **Envoi de notification et alertes (sms, emails, etc.)**
 - **Innovations personnelles**

Edge AI for Smart Homes

- Remise du travail et résultats sur github (pour le **25/05/2024**)
 - **Présentation des résultats devant le Jury le Dimanche à partir de 17h30**
 - **pitch selon vos modalités (avec support visuel)**
 - **démo live**
- Durée de chaque présentation (groupe) : **10 min. + 10 min. démo**
- Remise des prix à partir de **18h30**



Sponsors

HackIA
UMONS

| inforTech

numediart

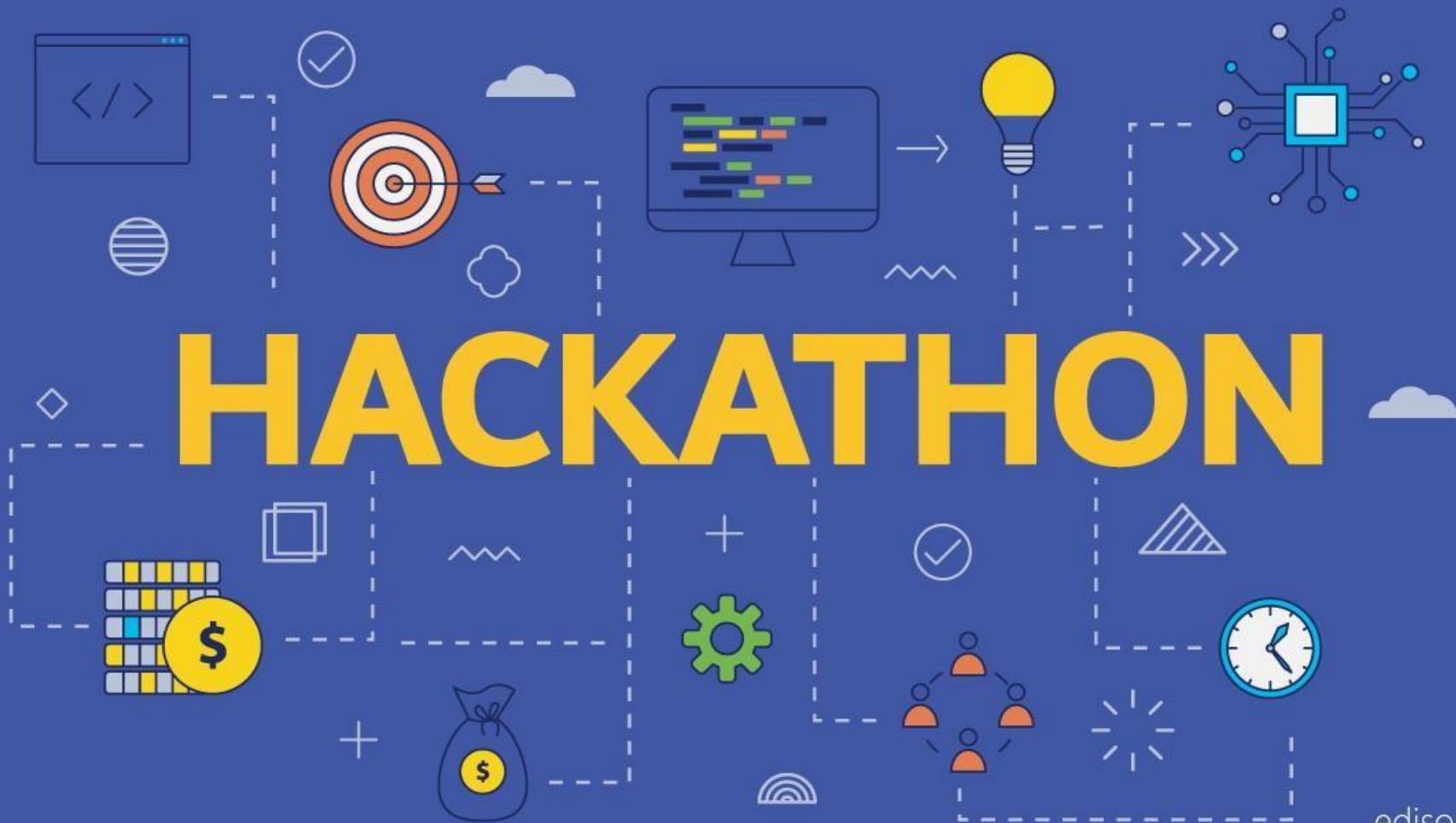
Critères

- Performance et complétude de vos systèmes
- Innovation
- Présentation persuasive

One more thing



HACKATHON



edison365
Making Ideas pay