

2^{ème} **FORUM** Régional IA

19 OCTOBRE 2022

Au LAB'O Village by CA d'Orléans



Atelier 4 - IA Embarquée

Le CRT CRESITT est soutenu par :



L'action de diffusion technologique est cofinancée par l'Union européenne.
L'Europe s'engage en région Centre-Val de Loire avec le Fonds européen de développement régional.



réseau des Centre-Val de Loire développeurs économiques

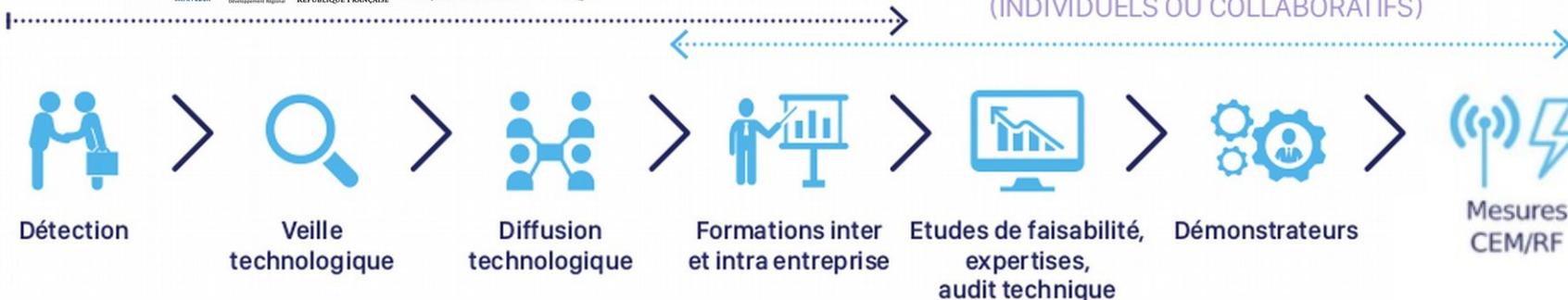


ACTIONS AIDÉES



FACTURATION

PROJETS D'ENTREPRISES (INDIVIDUELS OU COLLABORATIFS)



DE LA DÉTECTION DE BESOIN À L'ÉTUDE TECHNOLOGIQUE

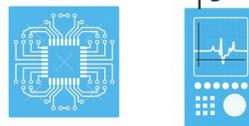




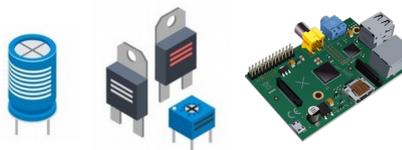
Appui à la mise en œuvre de technologies sans-fil et sans contact



Appui au développement de systèmes embarqués et sécurisés (électronique analogique et numérique, logiciels et noyaux embarqués temps réel)



Appui à l'intégration de capteurs et traitement des signaux (FPGA, ...)



Appui à la conception, validation et optimisation des alimentations et convertisseurs

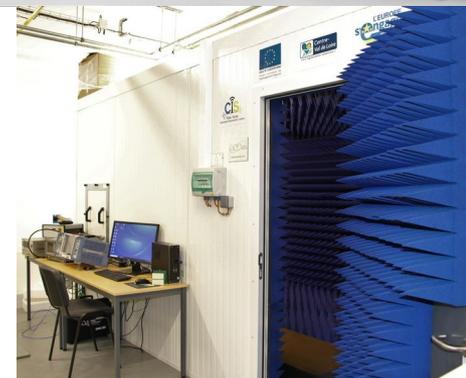
- **Compatibilité Electro-Magnétique**

- Rayonné et Conduit
- Immunité et Emissivité



- **RadioFréquences**

- Choix, adaptation et mesures d'antennes
- Mesures selon directive RED (EN300220 et EN300328)
- Simulations : Électromagnétisme et conception d'antennes avec FEKO d'Altair



- **Systemes autonomes**

- Mesures de consommations électriques
- Utilisation de différentes techniques de récupération d'énergie
- Gestion des différentes sources d'alimentation



Système Autonome à Pile à Combustible et photovoltaïque

- Évaluer les performances d'un algorithme d'intelligence artificielle (IA) sur différentes plateformes matérielles
- Faire un décompte du nombre de personnes dans une scène
- Évaluer les consommations énergétiques associées

- 1) Rechercher des plateformes matérielles
- 2) Rechercher l'algorithme d'intelligence artificielle
- 3) Effectuer le comparatif des performances
 - d'exécution
 - de consommation

CRITÈRES :

- Faible encombrement (max 10cmx10cm)
- Capable de traiter des images vidéos
- Connectivité vidéo disponible (USB, HDMI, etc...)
- Même support logiciel (OS, framework ...)
- Disponible à l'achat (pas si évident!)
- Faible coût (<500€)



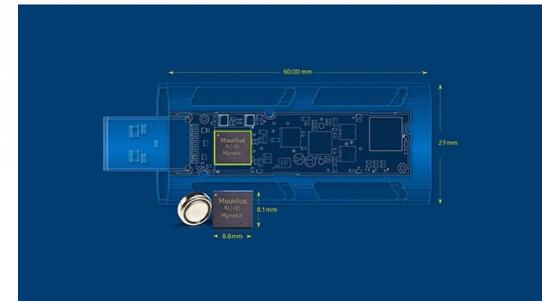
Raspberry Pi 3 B :

- CPU ARM Cortex A53 quad core - 1.2 Ghz x 4
- Multi-processeurs
- 4 ports USB 2.0
- Port HDMI
- Port de caméra CSI
- Port Ethernet
- Ram 1Go
- OS GNU Linux
- 31€ (2019)



Intel Neural Compute Stick 2 :

- Vision Processor Unit Intel Movidius Myriad X avec :
 - 16 processeurs vectoriels
 - Deep Neural Network hardware accelerator
- Traitement d'images et de vidéos
- Stick ~ USB
- IoT
- Toolkit OpenVINO
- 72€ (2019)



NVIDIA Jetson NANO :

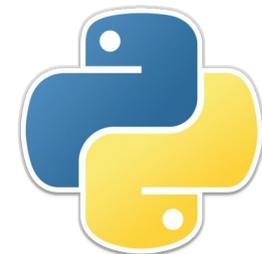
- CPU ARM Cortex A57 quad core - 1.43 Ghz
- GPU Maxwell - 128 cœurs (CUDA)
- 4 port USB 3.0
- Port HDMI
- Port de caméra CSI
- Port Ethernet
- 115€ (2019)



Algorithmes choisis : traitement d'image

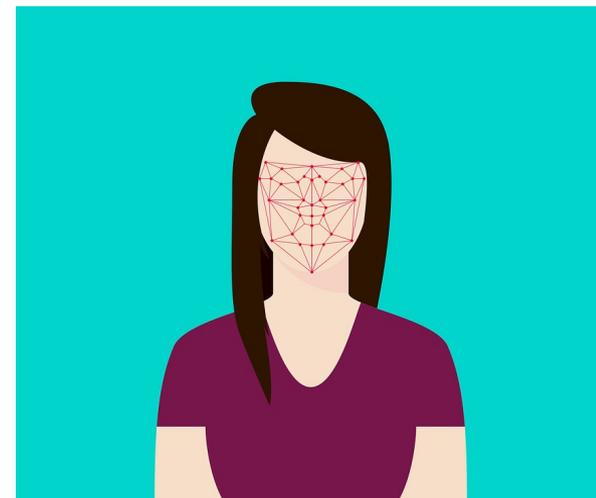
=> Détection de visages par Single Shot Multibox Detecteur (SSD) :

- Rapide pour des applications en temps réel
- Moins de calculs pour des systèmes embarqués
- Détection de haute précision



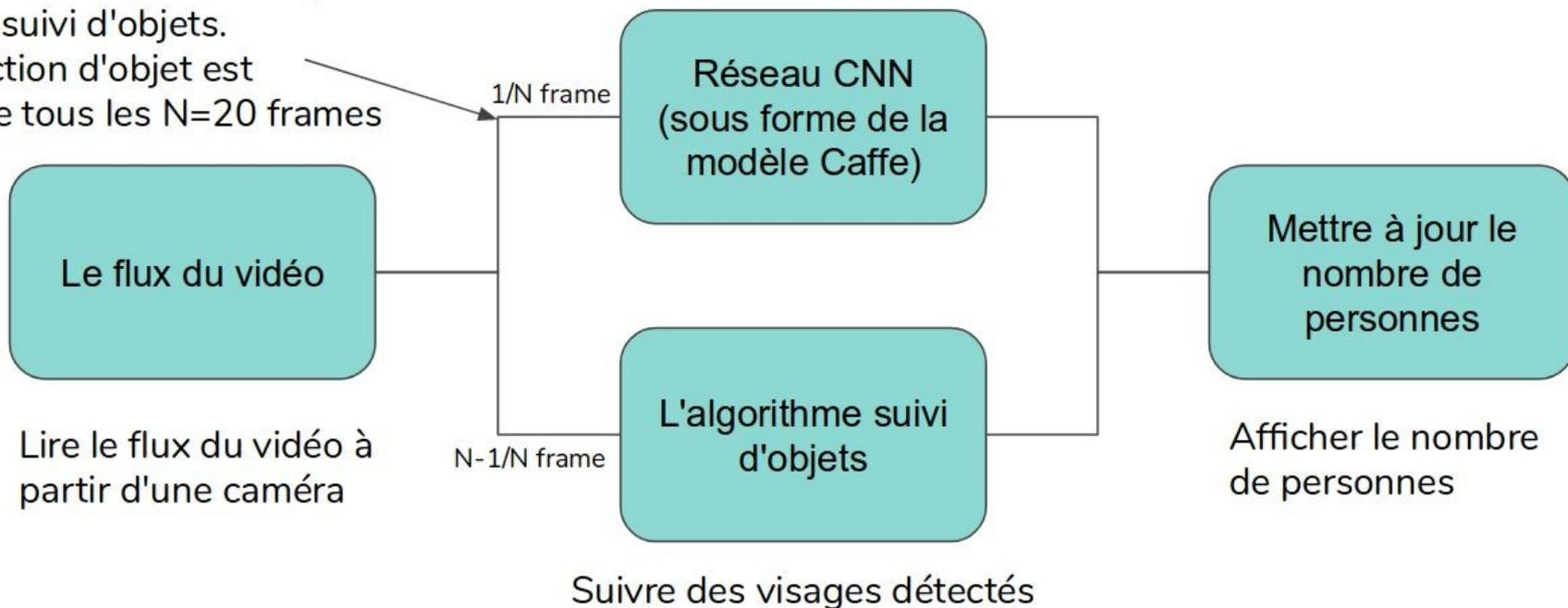
=> Suivi des visages dans une scène : suivi de corrélation

- Langage : Python + bibliothèques
- pour la portabilité entre plateformes



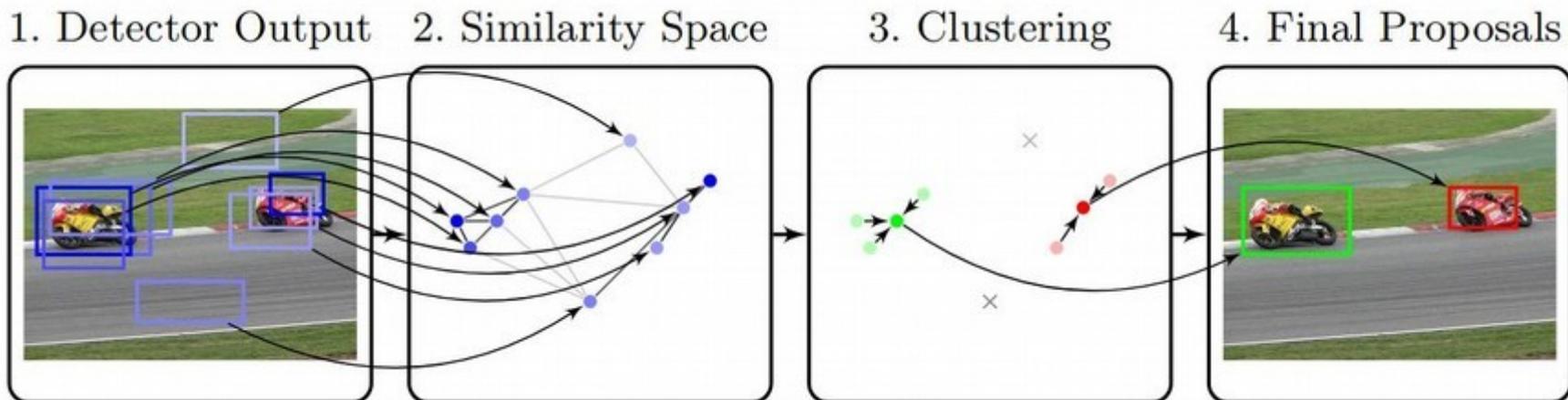
Comme le coût du calcul sur le réseau CNN est plus cher que celui du suivi d'objets. La détection d'objet est exécutée tous les $N=20$ frames

Passer chaque frame d'image au réseau pour détecter les visages

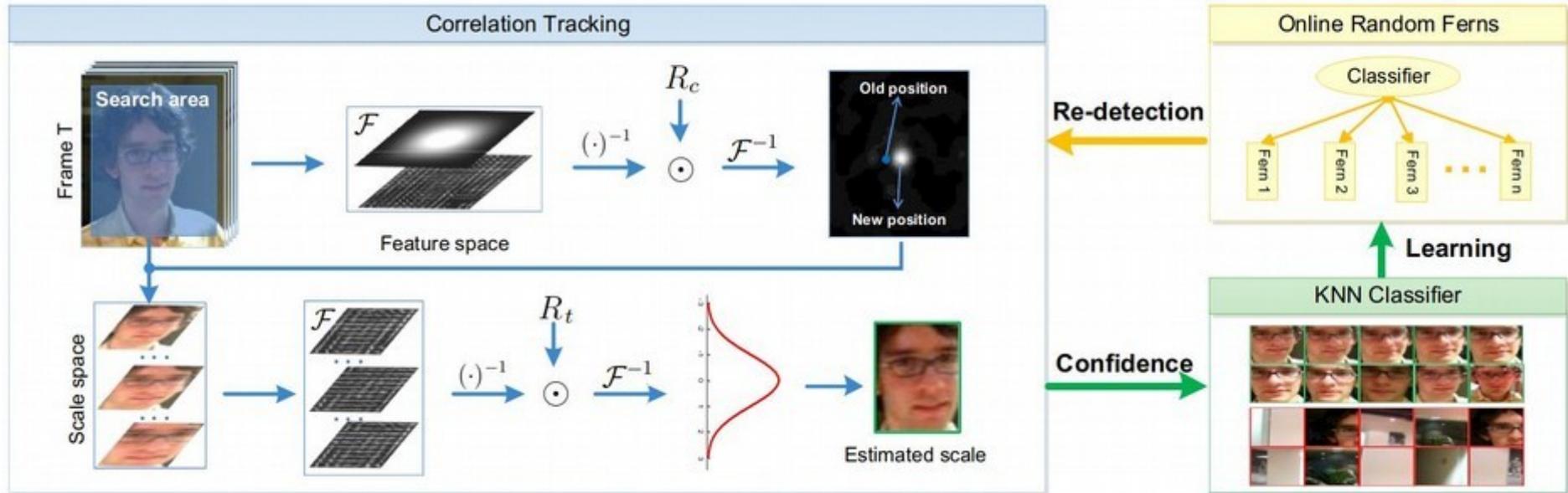


Single Shot Multibox Detector

- Single shot : tâches de localisation et de classification des objets sont effectuées en un seul passage du réseau
- Multibox : différentes boîtes pour différentes échelles
- Suppression des non-maxima



Suivi de corrélation





Caffe

- Paramètres :
 - Options de compilation des bibliothèques
 - Nombre maximal de visages détectables
 - Résolution de la caméra
 - Interface de communication de la caméra (CSI, USB)
 - Taille de l'image
 - Modèle de réseau de neurone



Plusieurs options testées :

- Architecture matérielle :
 - caméra sur USB ou sur bus CSI
 - CPU/GPU/NPU sélectionné(s)
- Architecture logicielle :
 - détection seule
 - détection + tracking
- OpenCV a été compilé dans OpenVINO avec des optimisations spécifiques pour chaque plateforme matérielle RBP ou Intel

Fixer des paramètres communs aux différentes plateformes

- Taille d'image réduite à 500 pixels de large avant prise en compte par le réseau
- Conditions d'éclairage identiques
- Un seul visage apparaît devant la caméra
- Algorithme identique
- Librairies opencv identiques
- 10 tests répétitifs (30 secondes/1 test) pour des tests du nombre d'images par seconde (FPS)
- 1 minute / test pour des tests d'énergie consommée (analyseur de puissance N6705B)



Raspberry Pi 3 B ----- USB Caméra (FPS)

Test	Sans Intel Movidius Neural Compute Stick				Avec Intel Movidius Neural Compute Stick			
	openCV 4.1.0		openCV 4.1.0-openVINO		openCV 4.1.0		openCV 4.1.0-openVINO	
	detection	detection & tracking	detection	detection & tracking	detection	detection & tracking	detection	detection & tracking
1	0.57	4.34	1.39	5.68	X	X	5.64	6.25
2	0.56	4.24	1.37	5.59	X	X	5.87	6.35
3	0.54	4.2	1.39	5.79	X	X	5.75	6.31
4	0.57	4.38	1.34	5.71	X	X	5.87	6.16
5	0.55	4.15	1.38	5.66	X	X	5.86	6.3
6	0.57	4.52	1.37	5.71	X	X	5.88	6.31
7	0.55	4.34	1.37	5.51	X	X	5.82	6.26
8	0.57	4.2	1.36	5.74	X	X	5.88	6.27
9	0.57	4.25	1.36	5.62	X	X	5.8	6.55
10	0.56	4.26	1.37	5.64	X	X	5.86	6.34
Moyenne	0.56	4.29	1.37	5.67	X	X	5.82	6.31

X : On ne peut pas utiliser OpenCV 4.1.0 avec NCS2 car il ne peut pas être compilé avec l'option Deep Learning Inference d'Intel.

Raspberry Pi 3 B ----- Pi Caméra (FPS)

Test	Sans Intel Movidius Neural Compute Stick				Avec Intel Movidius Neural Compute Stick			
	openCV 4.1.0		openCV 4.1.0-openVINO		openCV 4.1.0		openCV 4.1.0-openVINO	
	detection	detection & tracking	detection	detection & tracking	detection	detection & tracking	detection	detection & tracking
1	0.68	5.64	1.63	7.88	X	X	8.75	8.92
2	0.68	5.65	1.65	7.93	X	X	8.65	8.97
3	0.68	5.67	1.66	7.81	X	X	8.58	8.79
4	0.68	5.56	1.65	7.86	X	X	8.6	8.96
5	0.67	5.7	1.65	7.93	X	X	8.66	8.91
6	0.68	5.48	1.65	7.9	X	X	8.63	8.8
7	0.68	5.6	1.66	7.99	X	X	8.7	8.64
8	0.68	5.7	1.65	7.91	X	X	8.57	8.9
9	0.68	5.64	1.64	7.79	X	X	8.68	8.88
10	0.68	5.58	1.66	7.88	X	X	8.7	8.97
Moyenne	0.68	5.62	1.65	7.89	X	X	8.65	8.87

Test	RBP - CPU - caméra Pi		RBP - CPU - caméra USB		RBP - NCS2 - caméra Pi		RBP - NCS2 - caméra USB	
	FPS	Énergie consommée	FPS	Énergie consommée	FPS	Énergie consommée	FPS	Énergie consommée
Detection	1.65	3.75	1.37	4.14	8.65	4.67	5.82	5.07
Detection & Tracking	7.89	3.26	5.67	3.8	8.87	4.53	6.31	4.92

CAMERA Pi et algorithme « détection »

Pour 24,5 % de consommation en plus avec le Neural Stick (et le surcoût matériel), on a 524 % de performances en plus !!!

→ A priori intéressant !

CAMERA Pi et algorithme « détection + tracking »

Pour 39 % de consommation en plus avec le Neural Stick (et le surcoût matériel), on a « seulement » 12,5 % de performances en plus

→ est-ce vraiment nécessaire d'ajouter le neural stick ?

Test	Jetson NANO - Caméra USB - Detection		Jetson NANO - Caméra USB - Detection & Tracking	
	FPS	Énergie consommé	FPS	Énergie consommé
CPU	3.11	6.39	2.12	5.67
GPU	15.78	6.63	2.26	5.69

En effectuant les calculs de « détection » complets dans le GPU, le nombre de FPS traités est 5 fois + élevés que sur le CPU: les GPU ont plus de cœurs et surtout les techniques spéciales que NVIDIA a développées sont complètement adaptées à ce type de traitement.

Dans les tests « detection & tracking », la différence entre CPU et GPU est faible : cela est du à l'asynchronisme entre le framework Caffe et la bibliothèque Dlib (le suivi de corrélation de la bibliothèque Dlib s'exécute dans CPU), ainsi qu'au transfert de données entre le GPU et le CPU lors du calcul.

Pour un algorithme d'intelligence artificielle vision :

- Les performances du/des processeur(s) pré-déterminent les performances des algorithmes, mais pas seulement !
 - Il faut adapter les architectures matérielle et logicielle pour optimiser les ressources
- L'ajout du co-processeur peut être utile pour booster les performances, mais pas dans toutes les circonstances
- Les GPUs sont une alternative intéressante et performante, il faut cependant prêter attention aux transferts de données entre processeurs de traitements qui peuvent être très impactant

- Travaux CRESITT en 2023 pour étendre ces tests à d'autres plateformes, notamment sur les FPGA
 - C'est quoi un FPGA ?
 - Dossier sur : <https://cresitt.com/veille-technologique/>
 - Formation possible au CRESITT : <https://cresitt.com/formations/>
- Sujet de veille et séminaire en 2023 : nous suivre sur **LINKED IN** + **newsletter** !

Et aussi :

Financement : appel à projet en cours IA embarquée !

<https://www.bpifrance.fr/nos-appels-a-projets-concours/appel-a-projets-maturation-technologique-et-demonstration-de-solutions-dintelligence-artificielle-embarquee>

Elisabeth PATOUILLARD / Christophe ALAYRAC

CRESITT Industrie, Lab'O, 1 avenue du Champ de Mars, CS 30019,
45074 Orléans Cedex 2

02 38 69 82 60 / 06 95 12 51 76 / 07 67 29 56 40

Elisabeth.patouillard@cresitt.com / Christophe.Alayrac@cresitt.com

Le CRT CRESITT est soutenu par :

