

CRESITT INDUSTRIE

Centre de Ressources
Technologiques en Électronique



SÉMINAIRE

Intelligence Artificielle et
Électronique Embarquée

17 octobre 2019

À Polytech Orleans



Séminaire gratuit, inscription obligatoire sur www.cresitt.com



Évaluation de performance d'algorithmes IA sur des plateformes embarquées

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.



- 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 port USB 2.0
- Port HDMI
- Port de caméra CSI
- Port Ethernet
- Ram 1Go
- OS GNU Linux
- 31€



Intel Neural Compute Stick 2 :

- VPU Intel Movidius Myriad X
- Monoprocasseur (NPU)
- Traitement d'images et de vidéos
- Stick ~ USB
- IoT
- Toolkit OpenVINO
- 72€



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€



PC portable :

SAMSUNG R780

- intel core i5-430M (dual core 64bits) 2.26 Ghz
- RAM 4Go
- Carte graphique : NVIDIA GeForce GT 330M
- OS GNU Linux



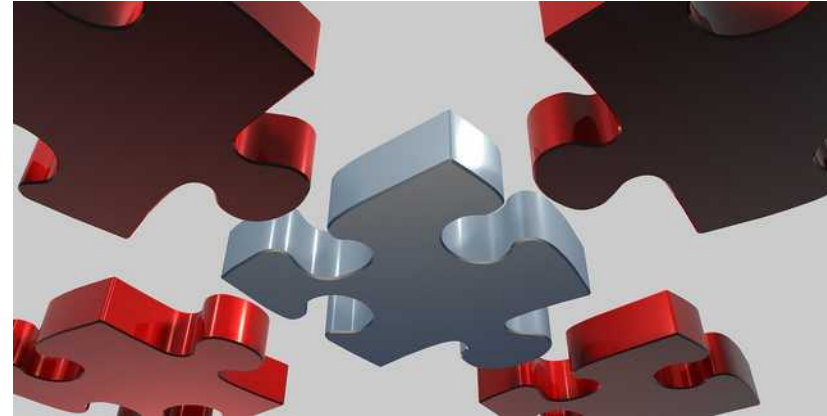
PC tour :

DELL precision T1500

- Intel core i7 (Quad core 64 bits) 2.93 Ghz
- RAM 8Go
- Carte graphique : AMD Radeon RV620
- OS GNU Linux



- Configurations possibles
 - Monoprocasseur (CPU)
 - Multiprocesseurs (CPUs)
 - Multiprocesseurs (CPU) + GPU
 - Multiprocesseurs + accélérateur NPU
 - 100 % GPU



Algorithmes choisis : traitement d'image

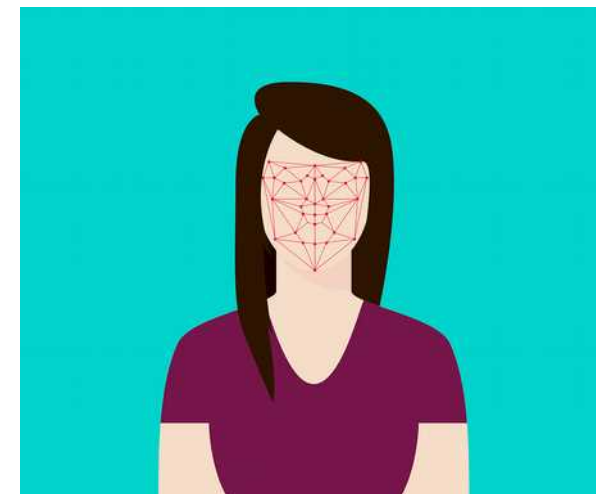
=> Détection de visages par Single Shot Multibox Detecteur (SSD) :

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



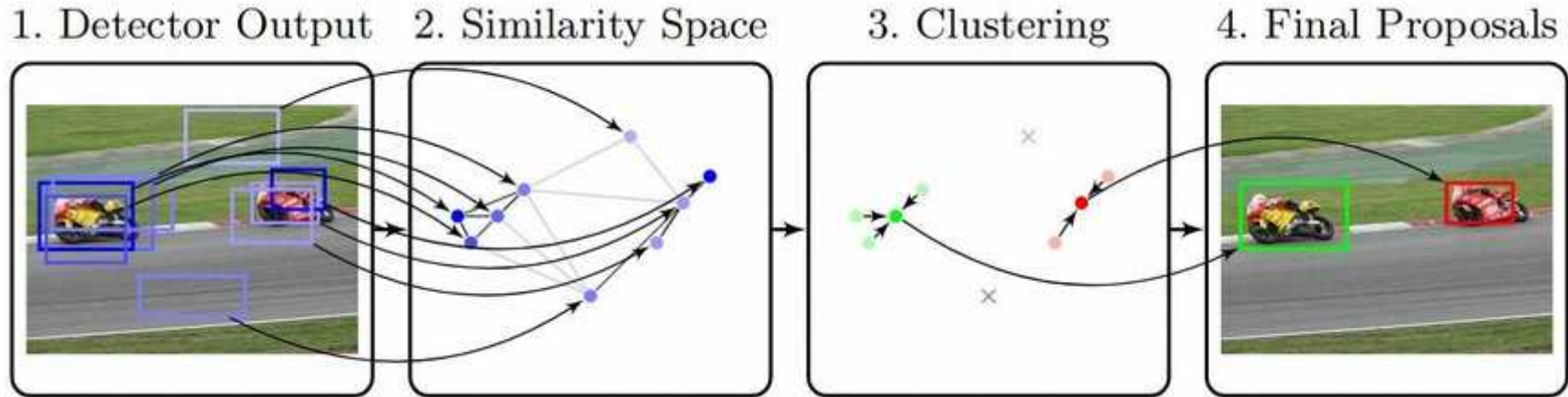
=> Suivi des visages dans une scène : suivi de corrélation

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

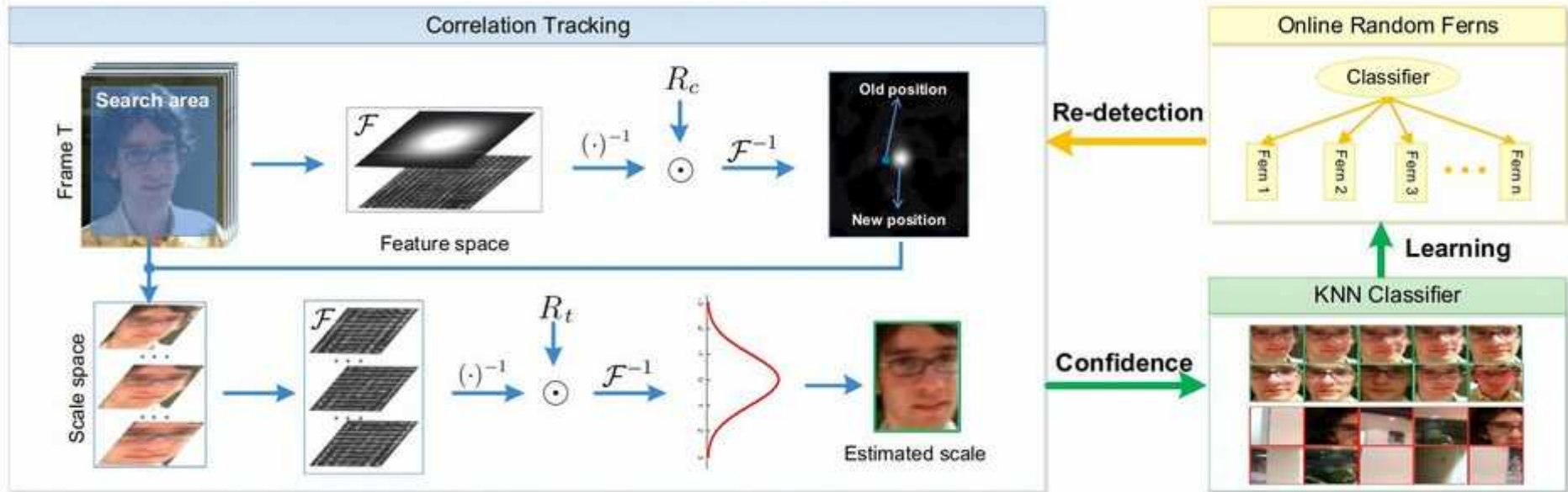


Single Shot Multibox Detector

- 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

Librairie de vision



- Licence BSD 3 clauses
- Interfaces : C++, Python, Java et MATLAB
- Manipulation de fichiers images et vidéos
- Traitement d'images
- Interface graphique
- Détection d'objets
- Fonction de filtrage
- Module de réseau de neurone profond
- Module de Machine learning
- Module de Clustering
- Nombreux modules externes
- >2500 algorithmes optimisés

Librairie d'apprentissage automatique (Machine Learning)



- Licence : Boost Software License
- Interface : C++
- Algorithmes de Machine learning
 - Ressources pour la classification, la régression, la labellisation, détection d'objets
 - Algorithmes de clustering
- Traitement d'images
- Gestion des threads
- Gestion de la compression

Convolution Architecture For Feature Extraction

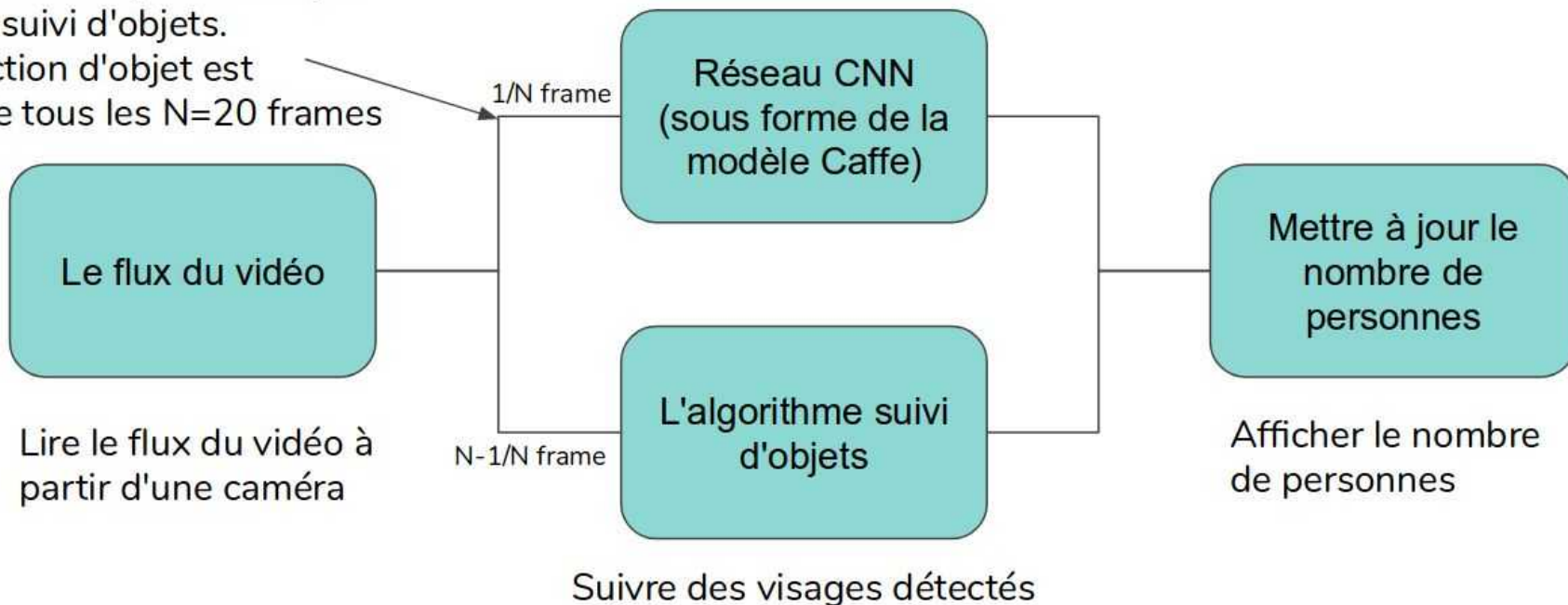


- Licence BSD 2 clauses
- Open framework, modèles, et exemples de deep learning
- 600+ citations, 100+ contributeurs, 7,000+ étoiles, 4,000+ forks
- Spécialisé vision et +
- C++ / CUDA architecture pour le deep learning
- Ligne de commande Python, interfaces MATLAB
- Rapide, fiable
- Outils : modèles de références, démos, et recettes
- Transition CPU / GPU

Slide credit: Evan Shelhamer, Jeff Donahue, Jon Long, YangqingJia, and Ross Girshick

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



- 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

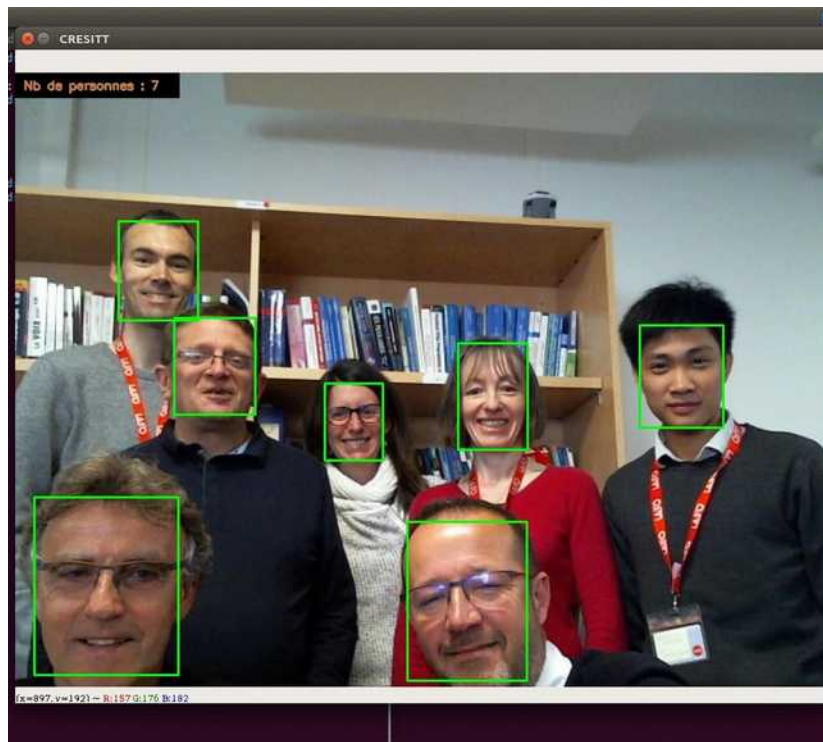


Fixe des paramètres communs aux différentes plateformes

- 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)



- Analyse des résultats obtenus (détection et détection & suivi)



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	6.34

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	Laptop - Caméra USB - Detection	Laptop - Caméra USB - Detection & Tracking	PC - Caméra USB - Detection	PC - Caméra USB - Detection & Tracking
	FPS	FPS	FPS	FPS
Avec NCS2	8.28 ↓	22.19	19.01 ↓	83.93
Sans NCS2	8.53 ↓	23.3	88.57 ↓	103.25

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

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

Pour un algorithme d'intelligence artificielle vision :

- Les performances du/des processeur(s) pré-déterminent les performances des algorithmes .
 - Plus il est rapide et plus il est performant
- L'ajout du co-processeur peut être utile pour booster les performances à moindre coût.
- Les GPUs sont une alternative intéressante et performante.
 - Avec de fortes ressources GPU, l'option cpu est moins performante
- L'inférence peut être exécutée sur des plates-formes légères, mais pas l'entraînement (trop coûteux en temps cpu).
 - Ex : cpu 8 cœurs 2.93Ghz, 8Go ram, 25000 images, 25000 itérations => 625 Heures (26j) 90 % de réussites. (Obtenu potentiellement en 5 Jours, asymptote)