

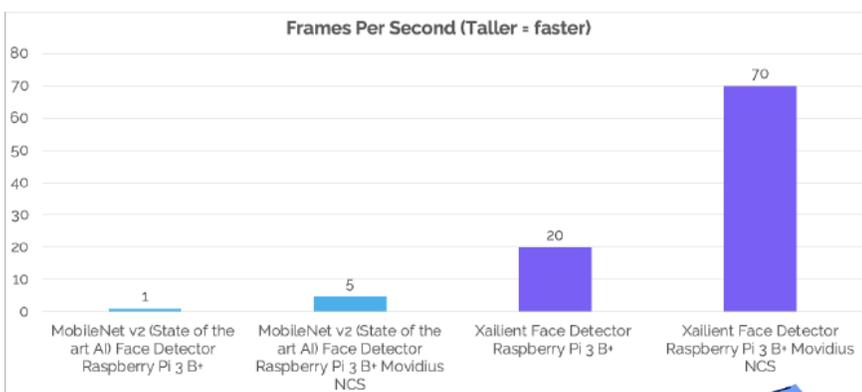
Xailient with Movidius on Raspberry 3B - Better Together

Synopsis

Xailient Detectum™ was created to fulfill the need for fast deep learning model inference at the Edge. Xailient with Movidius on Raspberry 3B+ gives you 70 fps with better software and an AI accelerator.

Key Outcomes

- ✓ Xailient **outperforms** the state-of-the-art MobileNet v2 model with and without the use of hardware accelerators.
- ✓ Xailient Detectum was created to fulfill the same real-time demands, making deep learning algorithms more efficient. The result is a **4 times higher frame rate** than the state-of-the-art MobileNet accelerated with the Movidius™ Neural Compute Stick.



Problem Statement

MobileNet v2 is a state-of-the-art architecture for mobile and embedded computer vision applications. Traditional AI models, like MobileNet v1 and MobileNet v2, are computationally intensive and require computational resources beyond the capabilities of low-power and low-cost devices.

Activity

We trained MobileNet v2 and Xailient Detectum using an open-source training dataset to create face detectors. These face detectors perform both localization and classification of faces in an image (i.e., generate bounding box data).

These models provided the baseline of state-of-the-art Edge-optimized AI, and their performance was measured in terms of inference speed in frames per second on Raspberry Pi 3B+.

Four different experiments were run: MobileNet v2 Face Detector without Movidius, MobileNet v2 Face Detector with Movidius, Xailient Face Detector without Movidius, and Xailient Face Detector with Movidius.

Results

The Baseline MobileNet v2 without Movidius NCS had an inference speed of 1 frame per second, and MobileNet v2 with Movidius NCS had an inference speed of 5 frames per second. At the same time, the Xailient model processed **20 frames per second** without the Movidius NCS and processed **70 frames per second** with Movidius.

	Computation	
	inference time in CPU (ms)	frames per second
MobileNet v2	1000	1.00
MobileNet v2 with Movidius	200	5.00
Xailient Detectum™	50	20.00
Xailient Detectum™ with Movidius	14	71.43

Next Steps

To meet the demand to run such deep learning models providing real-time, on-device inferences, AI hardware accelerators such as Movidius™ and Neural Compute Stick **significantly improve inference times.**

Discussion

The results were unprecedented when the Xailient Detectum™ algorithm was combined with the Movidius™ Neural Compute Stick hardware accelerator. Xailient achieved **70x faster inference** than MobileNet v2 and **14x faster** inference with MobileNet v2 combined with Movidius™ Neural Compute Stick.