PhiNets

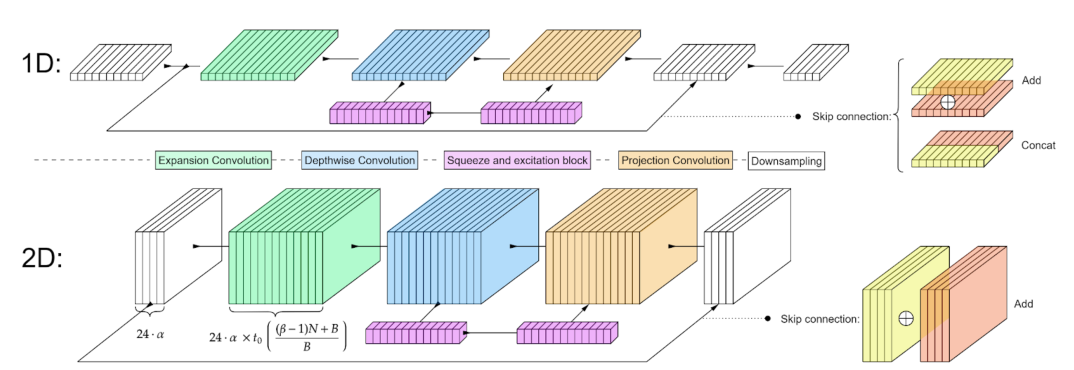
Francesco Paissan*, Alberto Ancilotto*, Elisabetta Farella
Fondazione Bruno Kessler (FBK) – Energy Efficient Embedded Digital Architectures Unit

What is the main contribution of PhiNets?

PhiNets:

- are small footprint, scalable neural networks;
- have a computationally efficient convolutional block (modified version of the inverted residual block);
- can exploit hardware-aware scaling.

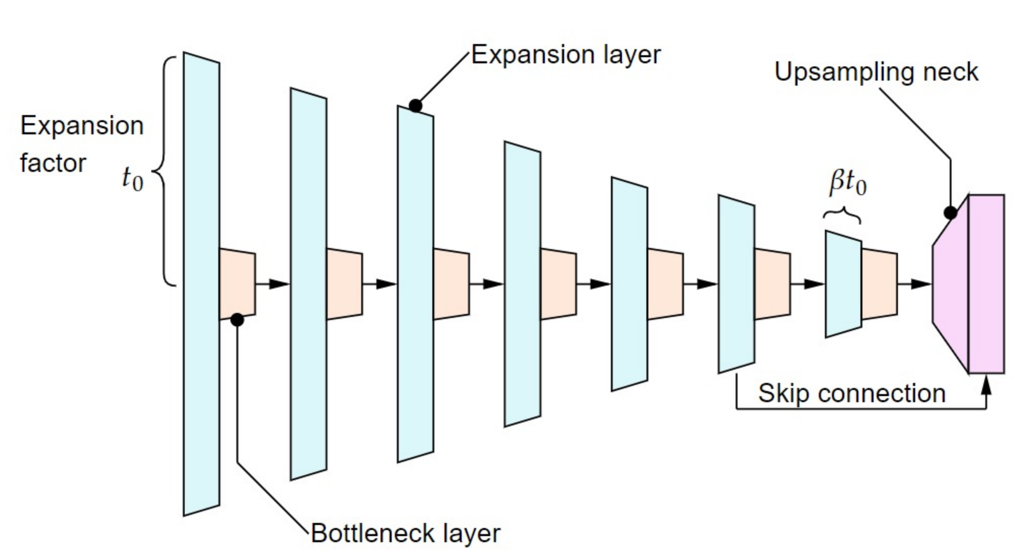
In particular, with PhiNets we propose to invert the Hardware Constrained Scaling paradigm and replace it with the **Hardware Aware Scaling** paradigm, which allows for one-shot generation of the neural architectures given the MCU's computational constraints.



PhiNets convolutional block

The PhiNets hyperparameters are:

- the **number of convolutional blocks** (B) used in the backbone;
- the **width multiplier** (α), which controls the number of channels in the feature maps;
- the **base expansion factor** (t0), which controls the expansion ratio inside the convolutional blocks;
- the **shape factor** (β) which helps in fine-tuning the memory used by the network;

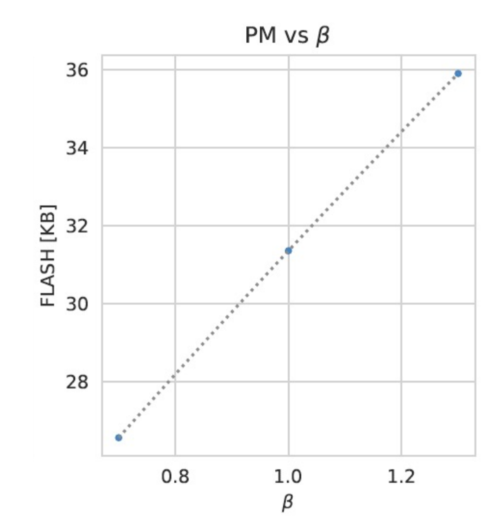


PhiNets backbone depicted with its hyperparameters

PhiNets resource usage

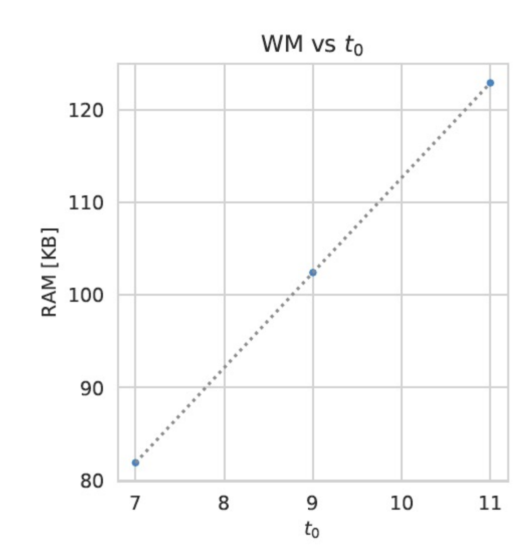
FLASH usage

The FLASH usage is determined by the parameter count of the network. In particular, for PhiNets, this scales linearly with the shape factor.



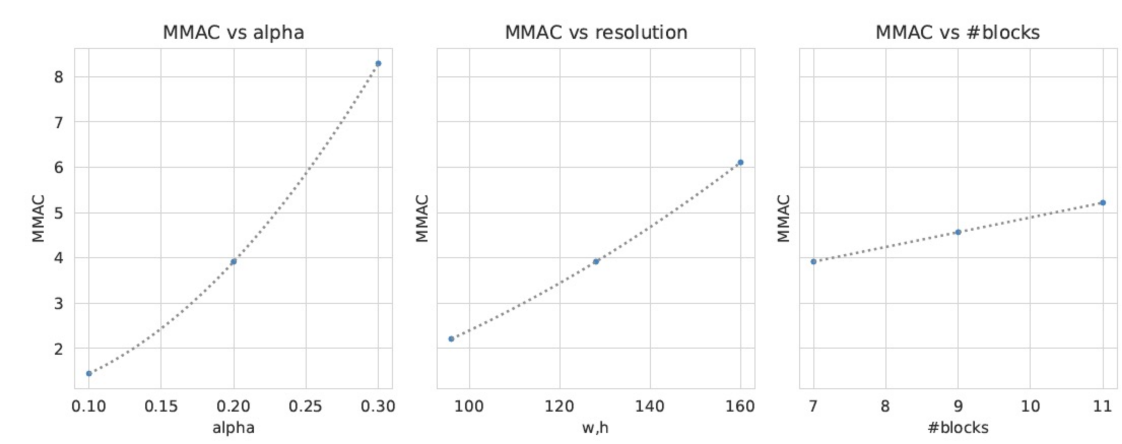
RAM usage

The RAM usage is determined by the biggest tensor in memory during inference. For PhiNets, this scales linearly with the expansion factor.



Multi-add count

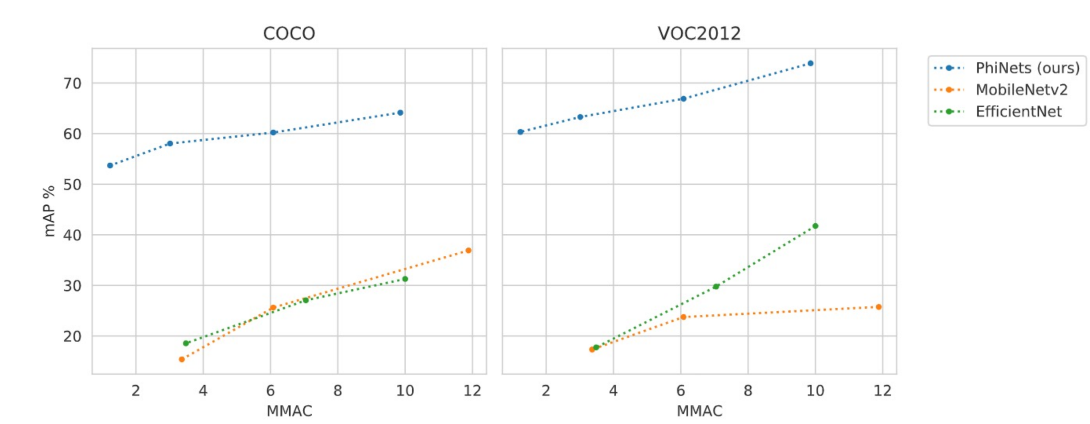
The multi-add count is by many factors during one inference step. In particular, this controls the inference rate of the PhiNets-based pipelines. As depicted, this scales quadratically with the width multiplier and input resolution and linearly with the number of blocks.

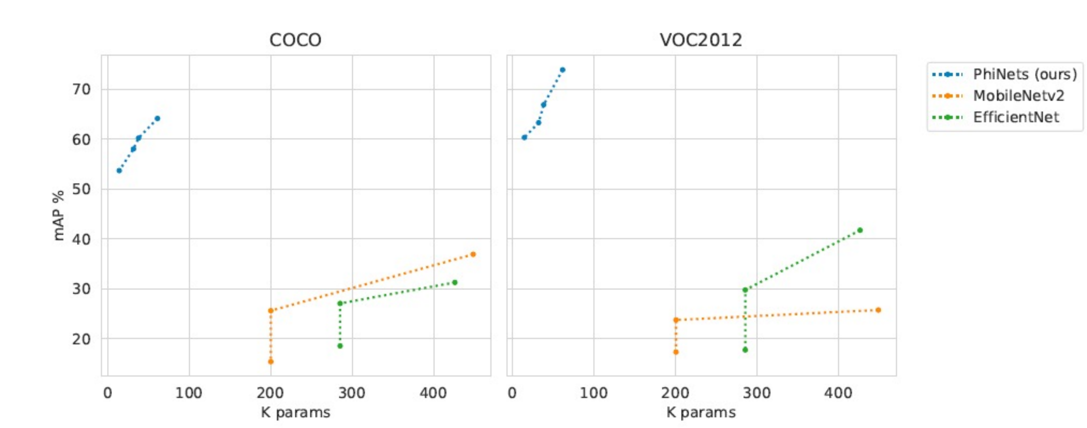


PhiNets for multimedia analytics

Multi-Object Detection

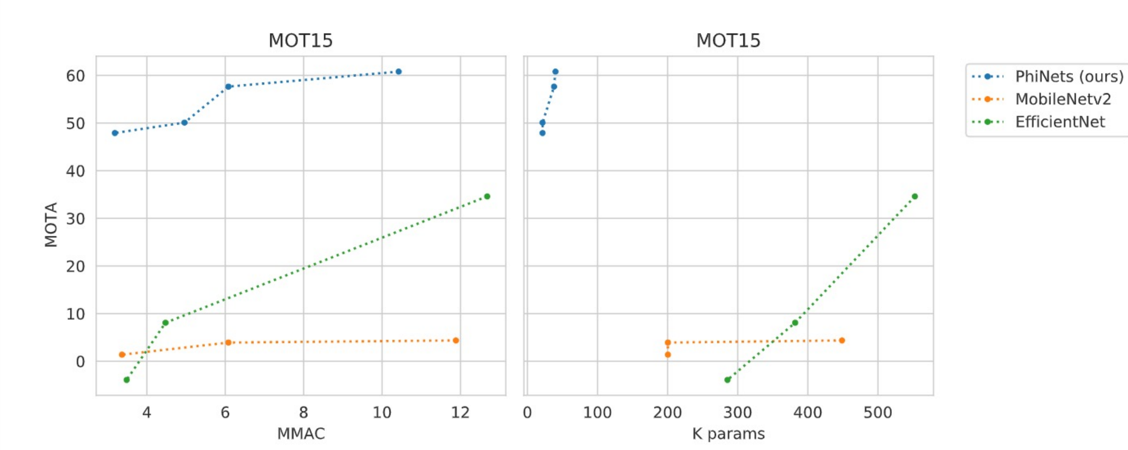
Consists of detecting and classifying objects in video streams. The proposed Multi-Object Detection pipeline is composed of a PhiNet backbone coupled with a YOLOv2 detection head. PhiNets have state-of-the-art performance in object detection in the MCU range.



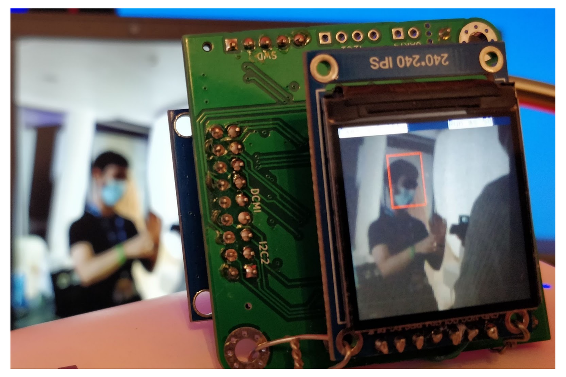


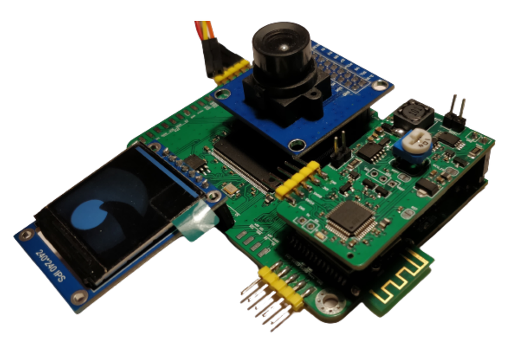
Multi-Object Tracking

We applied a simple online realtime tracker (SORT) on top of the detection pipeline above. As expected, the performance are in line with the tracking ones.



Demo device





Sound Event Detection

Performs detection of events from audio monaural signals. In this setup, the PhiNet backbone is coupled with a classifier for the 10-classes classification.

Input	Model	Params (K)	10-fold acc
Spectrogram	PICZAKCNN [10]	26 000	73.7
	SB-CNN [11]	241	73.11
	VGG [9]	77 000	70.74
	Cerutti M_{20k} [25]	30	69
	Cerutti M_{200k} [25]	200	72
	Cerutti M_{2M} [25]	2 000	75
	Cerutti M_{20M} [25]	70 000	76
	PhiNets M40	27.1	**76.3** ± 5.6
	PhiNets M15	32.2	76.1± 5.0
	PhiNets M5	3.80	68.8± 3.1
	PhiNets M3	2.18	65.3± 1.6
	PhiNets M1.5	2.00	62.3± 3.9

Input	Model	Params (K)	10-fold acc
Waveform	AudioCLIP [5]	>30 000	90.01
	ENVNET-V2 [14]	101 000	78
	W11-NET-WL [16]	1 806	68.47± 4.914
	W18-NET-WL [16]	3 759	65.01 ± 5.431
	W34-NET-WL [16]	4 021	66.77± 4.771
	1DCNN [13]	453	62± 6.791
	W-1DCNN-WL [16]	458	62.64± 4.979
	PhiNets 1D M1	**11.5**	59.3± 3.7
	PhiNets 1D M0.5	5.91	56.4± 6.4
	PhiNets 1D M0.2	2.11	48.4± 2.5
	PhiNets 1D M0.1	1.15	46.3± 4.2
	PhiNets 1D M0.07	0.766	43.3± 2.6

Image-to-image translation

We applied PhiNets to port Generative Adversarial Networks on MCUs for the task of face-swapping. Our approach uses only 76K params and 40M multi-adds and was successfully implemented on a K210 MCU at 20fps.

Conclusions

- PhiNets advanced scaling principles enable porting application to varying hardware platforms without the need of training many neural networks and select the feasible ones;
- PhiNets have state-of-the-art or near-SoA performance in many multimedia analytics tasks for both video and audio;
- With PhiNets, we were able to port GANs on microcontroller units;

[1] Francesco Paissan*, Alberto Ancilotto*, and Elisabetta Farella. "PhiNets: a scalable backbone for low-power AI at the edge." – ACM TECS 2022

[2] Francesco Paissan*, Alberto Ancilotto*, Alessio Brutti, and Elisabetta Farella. "Scalable neural architecture for end-to-end environmental sound classification" – Accepted to ICASSP 2022