### 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 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.

*Figure of the PhiNets convolutional block*

The PhiNets hyperparameters are:
- the number of convolutional blocks \(B\) used in the backbone;
- the width multiplier \(s\), which controls the number of channels in the feature maps;
- the base expansion factor \(d\), which controls the expansion ratio inside the convolutional blocks;
- the shape factor \(\beta\) which helps in fine-tuning the memory used by the network.

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

#### 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.

*Figure showing the multi-add count during one inference step.*

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.

---

#### 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.

*Figure showing the demo device.*

---

#### 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;

---
