Objective: **Video analytics** on edge and IoT devices

Networks must be:

- **Efficient**: Low power consumption for battery powered devices
- **Lightweight**: Able to run in real time on MCUs and edge devices
- **Adaptable**: Easy to scale to different resource constraints

**XiNet** + **Yolo-pose**

Building blocks: XiNet Convolutions, built from a set of maximally efficient operators.

- **Efficient**: From real-world measurements
- **Lightweight**: >80% fewer ops than Conv2D
- **Adaptable**: Allow for Hardware Aware Scaling

Hardware Aware Scaling: three hyperparameters allow to disjointly optimize FLASH, RAM, MAC.

- **Alpha**: Width factor, sets MAC
- **Beta**: Shape factor, sets FLASH
- **Gamma**: Compression ratio, sets RAM

**Results**

Networks scaled using Hardware Aware Scaling to fit different classes of devices:

- **MCU** STM32H7 - 100MMAC/s, 2MB Flash, 1MB Ram
- **TPU** K210 - 1GMAC/s, 16MB Flash, 5MB Ram
- **MPU** Raspberry Pi 4B - 16GMAC/s, 16GB SD, 4GB Ram

**Raspberry Pi 4B**

- Network Configuration:
  - \(\alpha = 1.0\)
  - \(\beta = 1.0\)
  - \(\gamma = 4.0\)
- Resource Utilization:
  - MMAC = 6112 M
  - Parameters = 3.2 M
  - RAM = 1.2 M
- Performance: 72.6 mAP
  - Energy: 1989 mJ/frame

**Kendryte K210**

- Network Configuration:
  - \(\alpha = 0.75\)
  - \(\beta = 1.0\)
  - \(\gamma = 4.0\)
- Resource Utilization:
  - MMAC = 859 M
  - Parameters = 1.8 M
  - RAM = 622 K
- Performance: 71.2 mAP
  - Energy: 410 mJ/frame

**ST STM32H743**

- Network Configuration:
  - \(\alpha = 0.33\)
  - \(\beta = 0.8\)
  - \(\gamma = 5.0\)
- Resource Utilization:
  - MMAC = 625 M
  - Parameters = 1.2 M
  - RAM = 180 K
- Performance: 70.1 mAP
  - Energy: 33.4 mJ/frame