XiNet: Efficient Generative Networks for Edge Devices

Challenging the efficiency of depthwise convolutions for edge and tinyML
- Novel convolutional block optimizing latency and energy usage
- Benchmarking on multiple embedded platforms
- Hardware Aware Scaling: from hardware constraints to neural architecture

Object/Pose Detection

Typical pipeline:
- Video transmission
- Analysis (Large neural networks)
- Privacy problems: High bandwidth, Not scalable

On-device:
- Move processing to the edge device
- XiNet Pose
- Analysis (Simple neural networks)
- Privacy, Low bandwidth, Scalable

Reducing bandwidth by 3 orders of magnitude!

Efficient image generation on edge devices

Style transfer

Content Input

XiNet Style

Style transfer network

Style Input

One shot image generation can be used for anonymization while preserving semantic content - removing personal information for downstream tasks

Style loss

Content loss

Results

Networks scaled using Hardware Aware Scaling

- MCU: STM32 - 100MMAC/s, 2MB Flash, 1MB Ram
  - Speed: 63.6 fps
  - Power: 14.89 W

- TPU: K210 - 1GMAC/s, 16MB Flash, 5MB Ram
  - Speed: 31.7 fps
  - Power: 410 mW

- MPU: rPi 4B - 16GMAC/s, 16GB SD, 4GB Ram
  - Speed: 5.5 fps
  - Energy: 72.4 mW

Convolutions Block

Hardware-Aware Scaling

Three main computational constraints in different embedded devices:
- FLASH: stores network parameters
- RAM: stores intermediate tensors
- MAC/s: determines latency & energy

Designed from real world efficiency measurements on various platforms. Three hyperparameters:
- $\alpha$: sets MAC
- $\beta$: sets FLASH
- $\gamma$: sets RAM

Convolutional Block

Three hyperparameters:
- $\alpha$: sets MAC
- $\beta$: sets FLASH
- $\gamma$: sets RAM

XiNet Pose Analysis (Simple neural networks)

Privacy, Low bandwidth, Scalable

Video transmission

Expensive, power hungry

XimSwap

Single target face swapping in 4 steps using 3 networks:

1. Face Detection
   - XuNet + Yolo

2. Landmark Detection
   - XuNet + PFLD

3. Face Alignment
   - XuNet + PFLD

4. Face Generation
   - XuNet GAN

One shot image generation can be used for anonymization while preserving semantic content - removing personal information for downstream tasks