

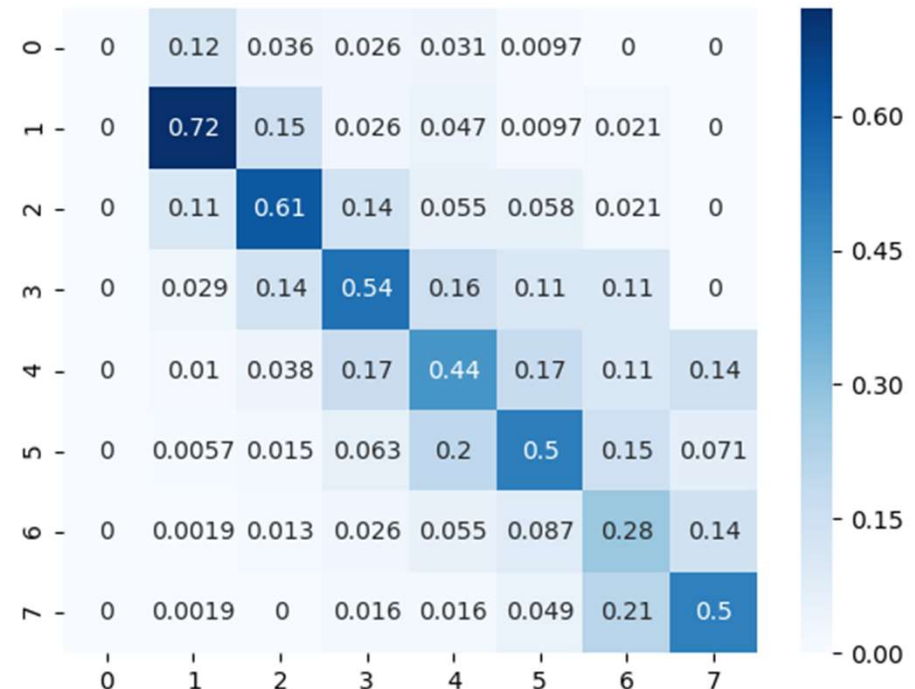


Lightweight Face Detection on ARM Cortex-M7

Edoardo Mello Rella, Davide Menini

Ideal Algorithm

- Pretrained Multi-Task CNN
 - 3 cascaded CNNs, for a total of 600k parameters
 - Works with RGB images of any shape
 - Various iterations at different scale
 - Needs a quite high resolution to work properly
 - Detects also face details (out of scope)
- Accuracy
 - 86.55% with IoU-F1 metric (tolerance ± 1)
- Great results, but definitely a dream for a MCU



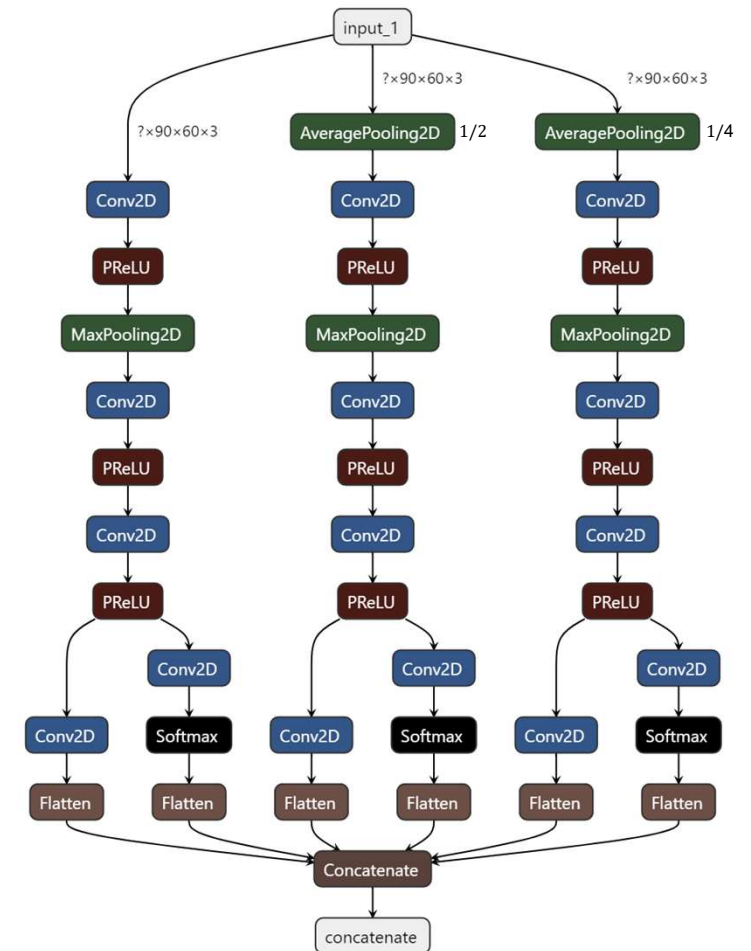
Real Implementation

- Image Preprocessing on Host Device
 - Gaussian smoothing to spread information over neighboring pixels ----- accuracy
 - Downscaling the max dimension to size 90 ----- memory, accuracy
 - Padding to size 90x60x3 using boundary values ----- MCU constraint
 - Normalization of intensities ----- accuracy
- Original CNN lightened and flashed on Cortex-M7 ----- memory, accuracy
- Postprocessing on Host Device
 - CNN's outputs transformed from predictions to boxes
 - Clean-up to reduce false positives ----- accuracy

Real Implementation - CNN

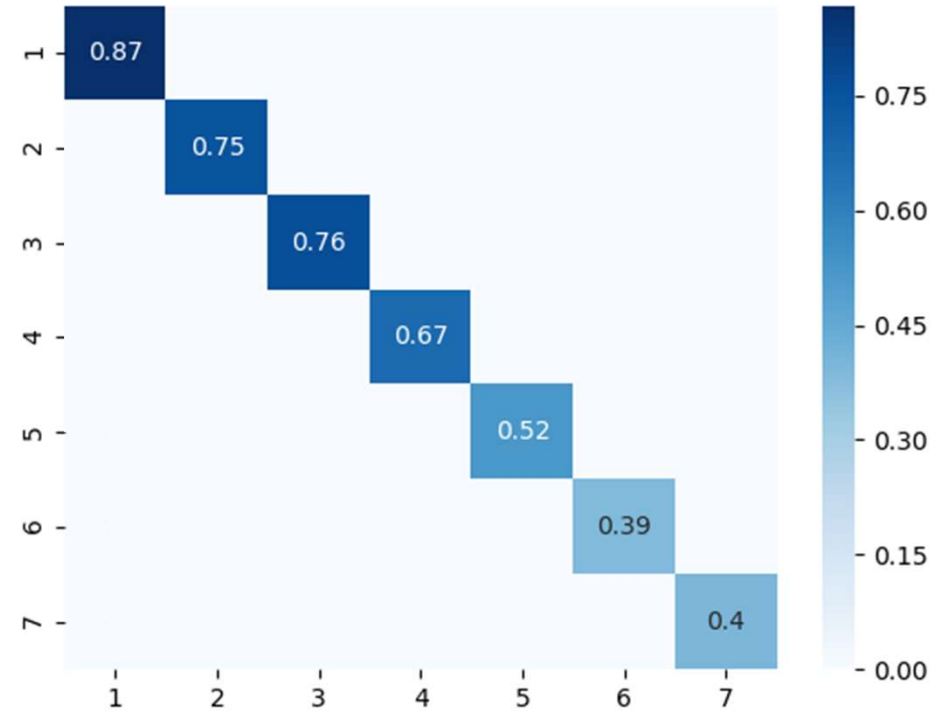
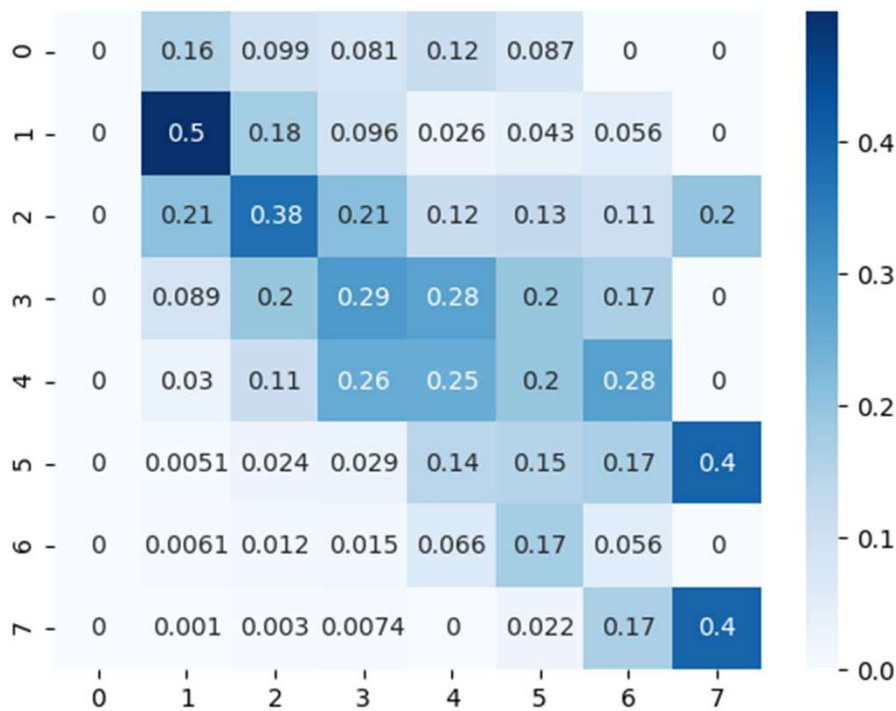
- Only 1st CNN (*pnet*)
- 3 iterations at different input scale (1, 1/2, 1/4) unrolled and concatenated
- Single input and single output to be compliant with X-CUBE-AI validation process

Parameters	19'896
RAM	301 kB (205 kB + 96 kB)
ROM	81 kB
MACs	9'974'344



Results - Accuracy

- Accuracy: 76.62% with IoU-F1 metric (tolerance ± 1)



Results - Profiling

- Profiling on STM32F756 @192MHz
- Results from X-CUBE-AI ('validate on target'), double checked by using the DWT_CYCCNT register to count cycles at runtime.

Duration (on average)	363.6 ms
CPU cycles	69'811'225
MACs	9'974'344
Efficiency	7 cycles/MAC (float32)
Bottleneck	159 ms for 3 rd conv2D

Notes & Future Work

- 7 cycle/MAC is not optimal. We could accelerate it using some advanced DSP extensions of Cortex-M7 (e.g. CMSIS-DSP library)
- Training the network by ourselves would allow some improvements:
 - Higher accuracy
 - Using grayscale images to reduce memory occupation and computation time
 - Using float16 instead of float32
- Postprocessing in C was complicate and would have required more time

Thank you for your attention!