

Face detection on MAXIM78000

Lionnus Kesting

Tuesday, June 6, 2023



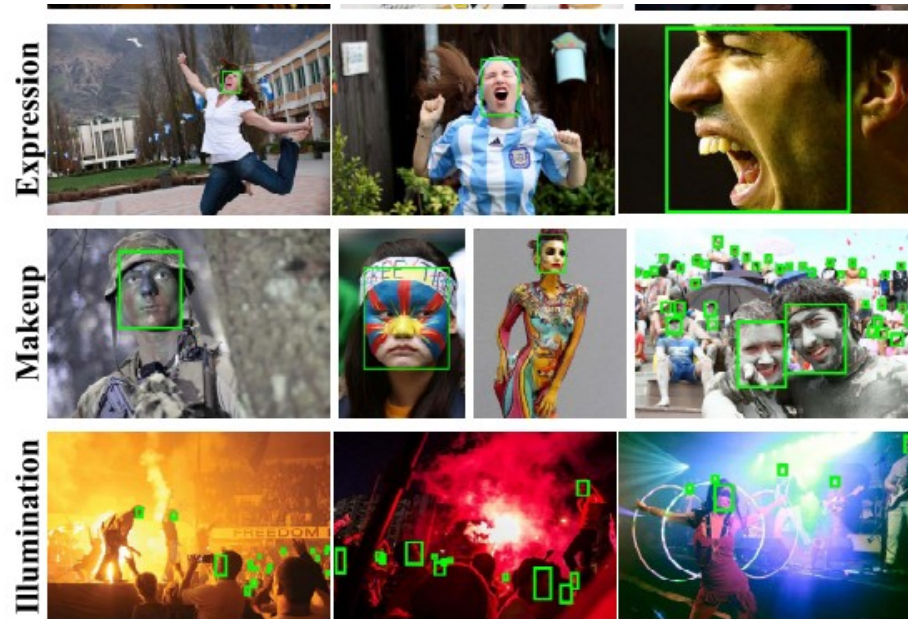
Contents

1. Overview
2. Dataset
3. MCU
4. Network
5. Results
6. Demo/Conclusion

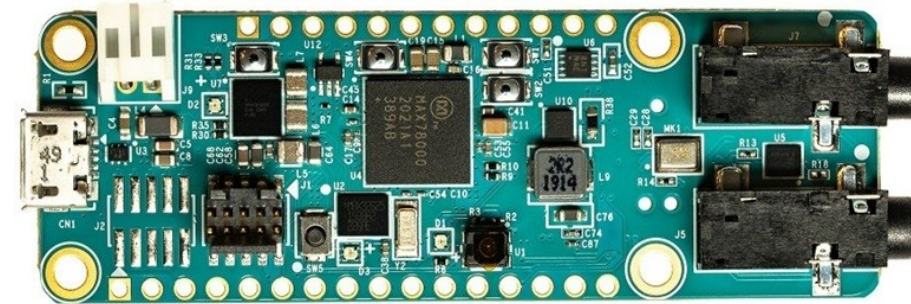
Overview

Dataset: WIDER Face

(<http://shuoyang1213.me/WIDERFACE/>)



MCU: MAXIM78000



Goal: Designing and deploying a ML network that can perform face tracking using the hardware CNN accelerator of the MAXIM78000

WIDER FACE Dataset

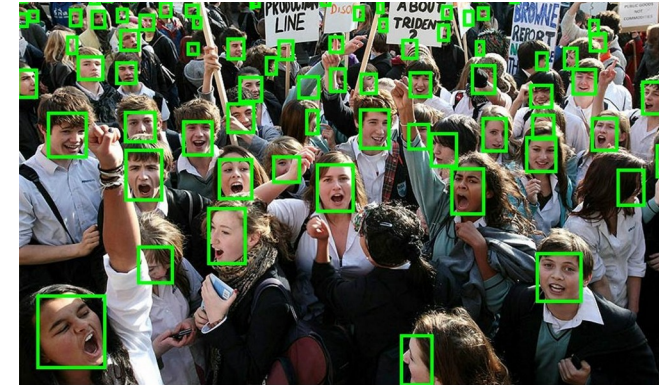
Contains 32,203 images with **393,703** labeled faces

Pre-processing for lightweight network:

1. Filter the dataset to contain only **images with a single face** →
2. Generate **samples without any faces** from images with multiple faces →
3. Images have arbitrary aspect ratios: **resize both images and boundary boxes** for the network structure

After pre-processing:

- Contains 4,631 images and correspondingly **4,631** labeled faces
- Balanced dataset with **same amount of images with/without faces**

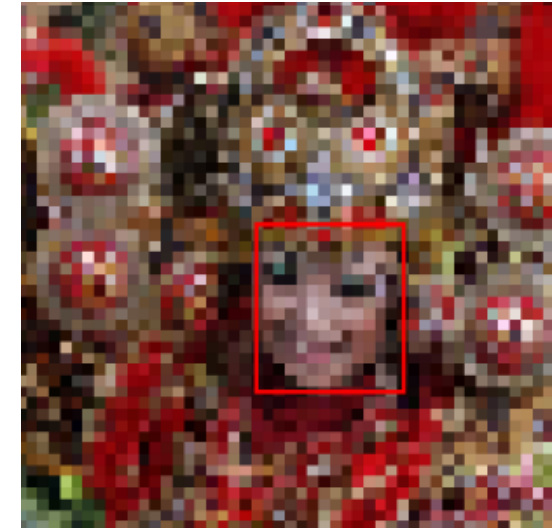
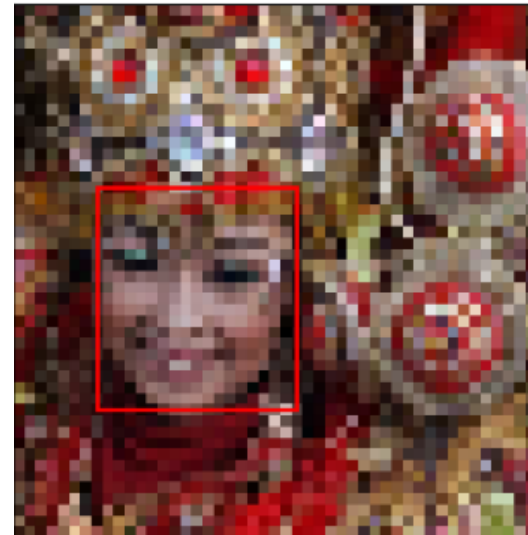


WIDER FACE Dataset

Faces too small for the input image size (48x48), aspect ratio not maintained when just resizing



Crop square image with random size, with random spacing around boundary box



Final dataset

Train set length: **7435 pictures** (3780 with face, 3655 without)

Test set length: **1781 pictures** (891 with face, 890 without)

MAXIM7800 Neural Network Accelerator capabilities

- 64 parallel physical channel processors
- **442k 8-bit weights**
- Max. **1024x1024 input image**
- Network depth up to **64 layers**
- Maximum channel **width of 1024 channels**



Network structure of WIDER FACE paper

MTCNN Network

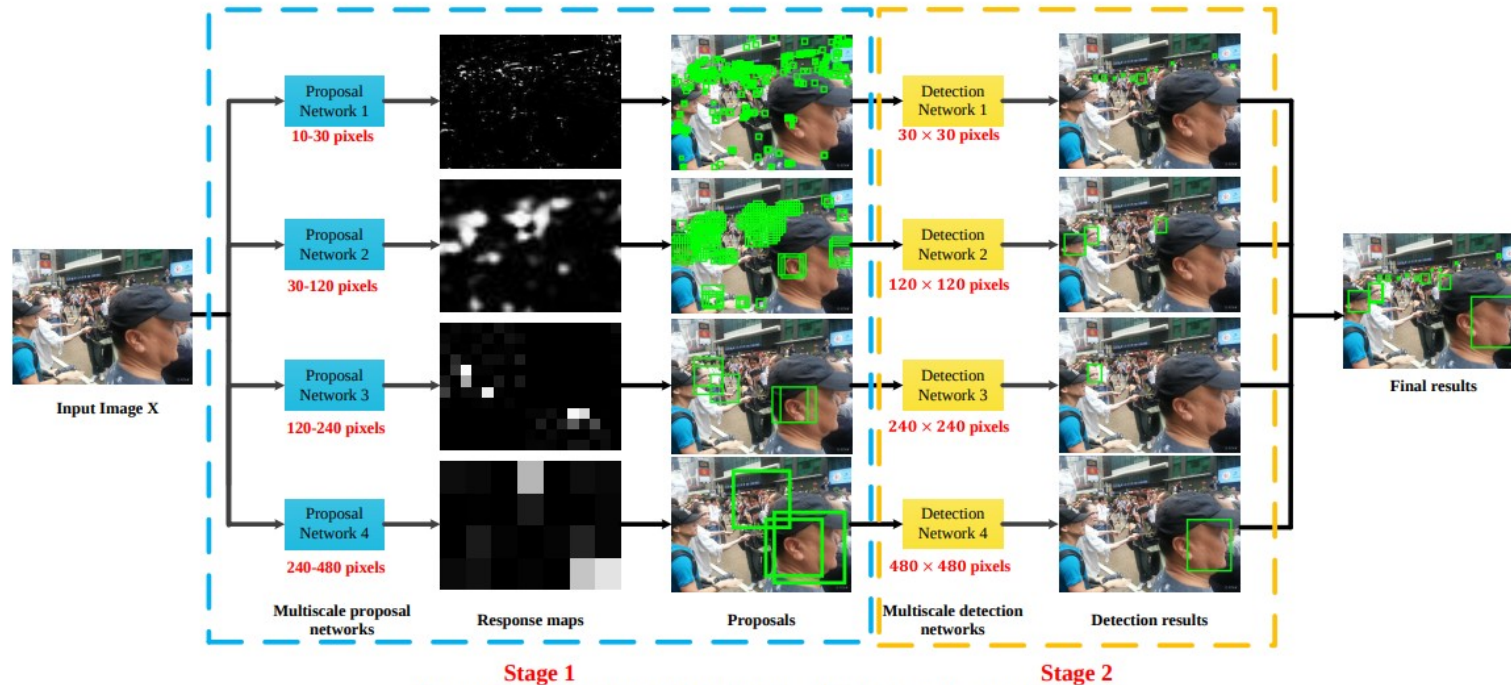


Figure 5. The pipeline of the proposed multi-scale cascade CNN.

→ Use small part of the MTCNN network

Image source: [1] S. Yang, P. Luo, C. C. Loy, and X. Tang, "Wider face: A face detection benchmark," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. doi:10.1109/cvpr.2016.596

Multi-task Cascaded CNN (MTCNN)

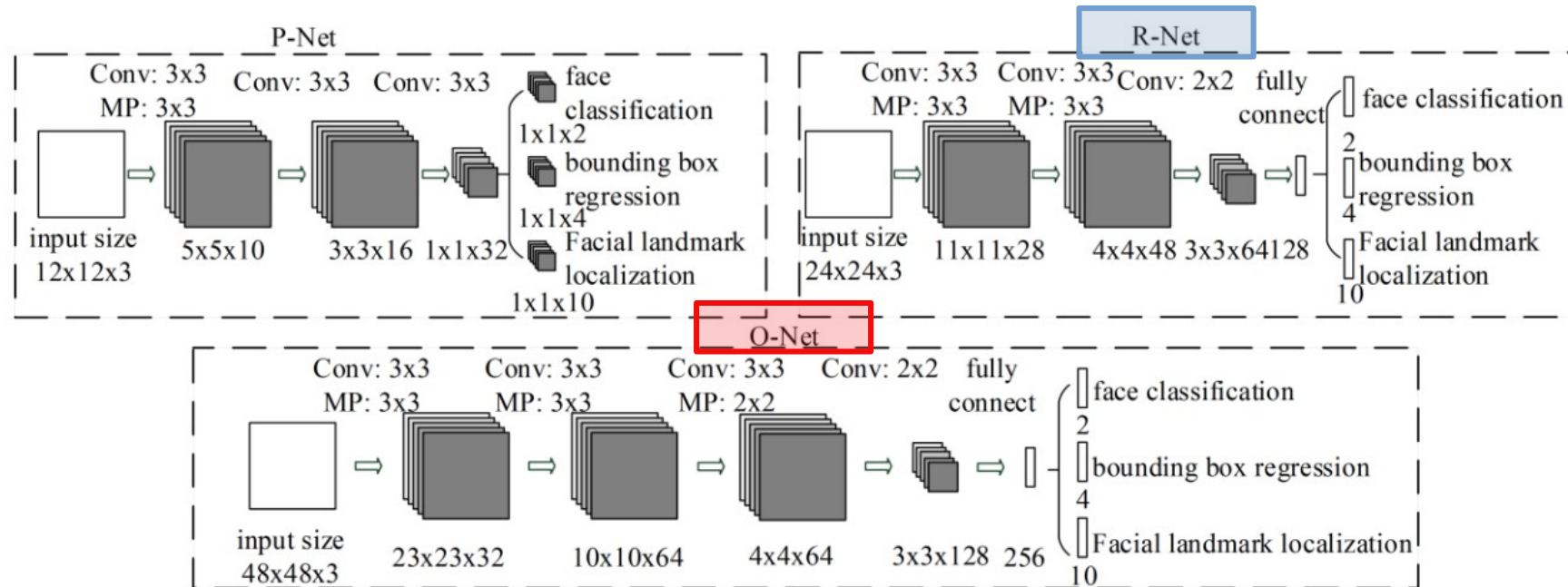


Image source: K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499–1503, 2016. doi:10.1109/lsp.2016.2603342

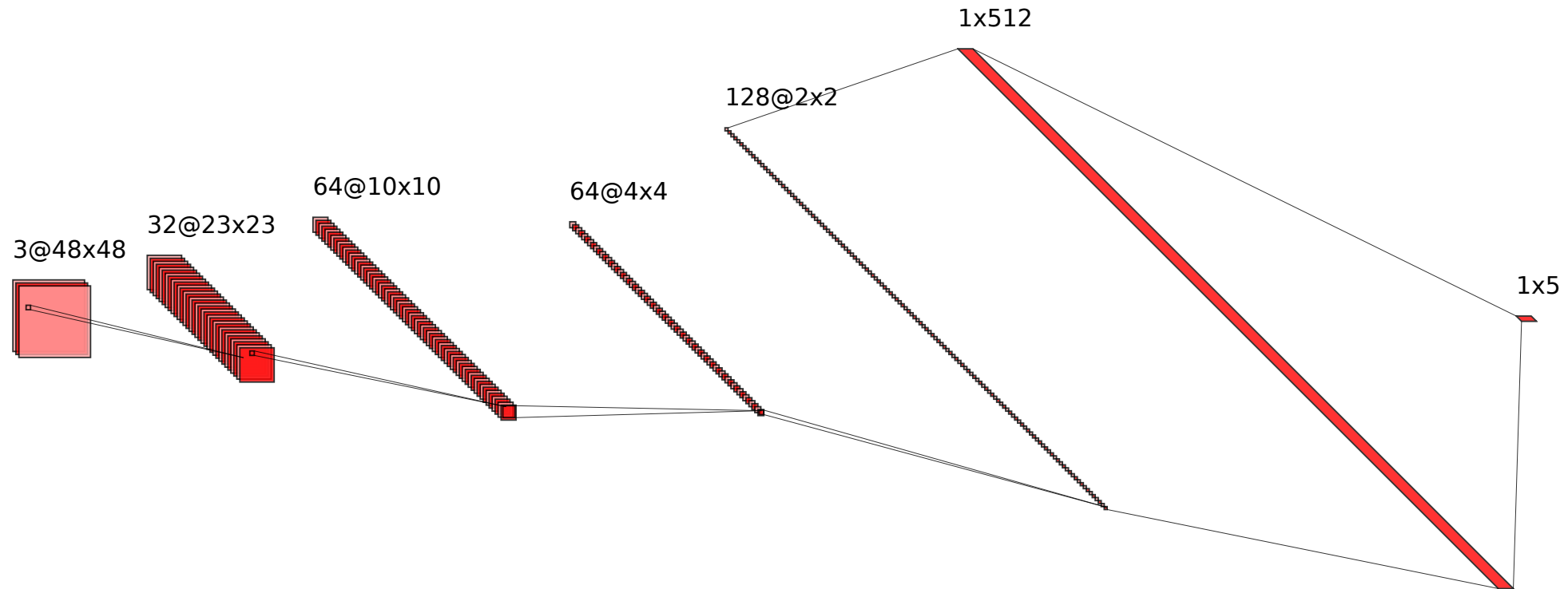
Network structure ONet

O-Net of MTCNN for lightweight implementation

Some adjustments needed for MAXIM78000:

- **Dense layer has too many channels (1152 instead of the maximum of 1024) → Output of convolutional layer changed from 256 to 128 channels**
- **Convolutional kernel size of 2x2 not supported → changed to 3x3**

Network structure ONet



Conv + ReLU

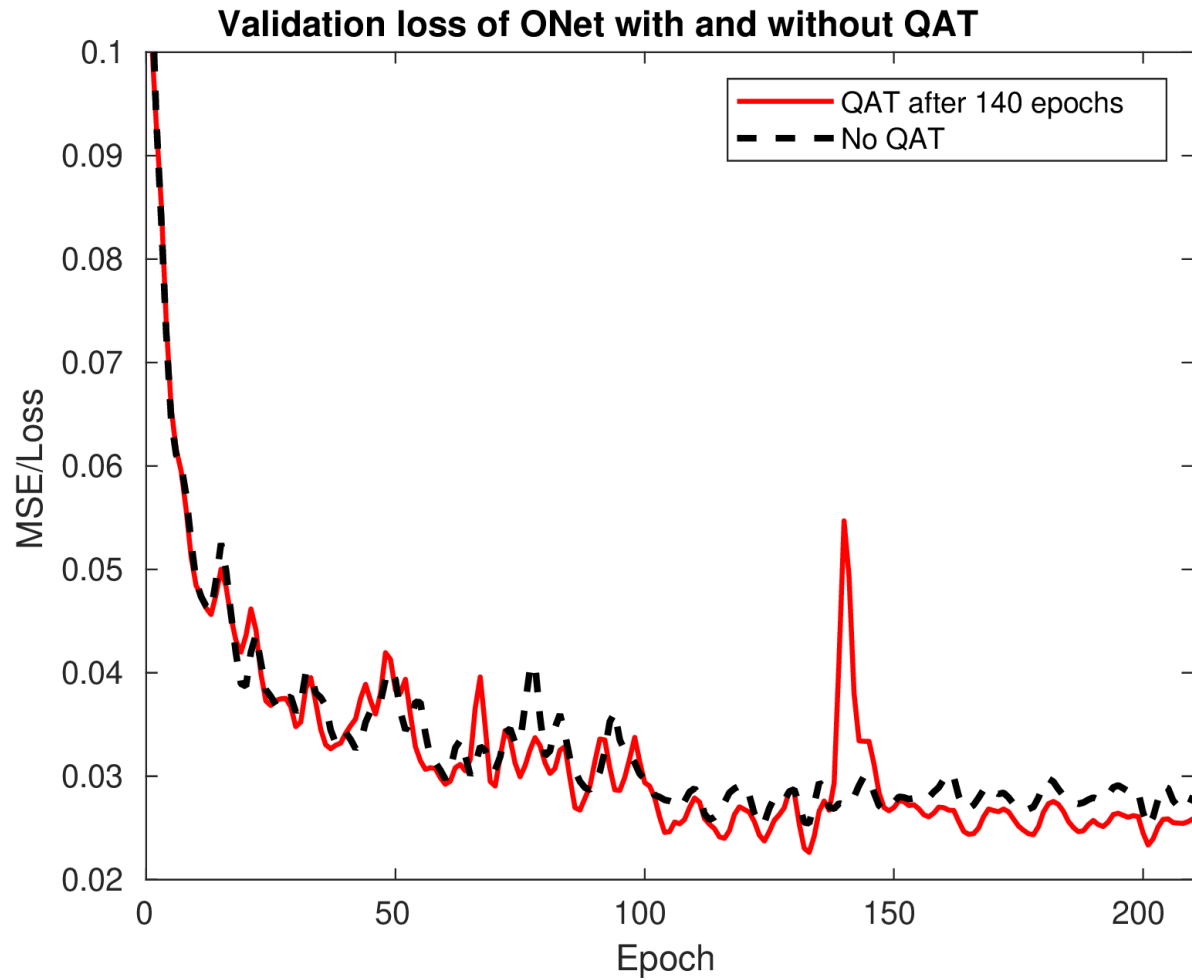
MP + Conv + ReLU

MP + Conv + ReLU

MP + Conv + ReLU

Flatten

QAT ONet



With QAT → **11% decrease** in loss

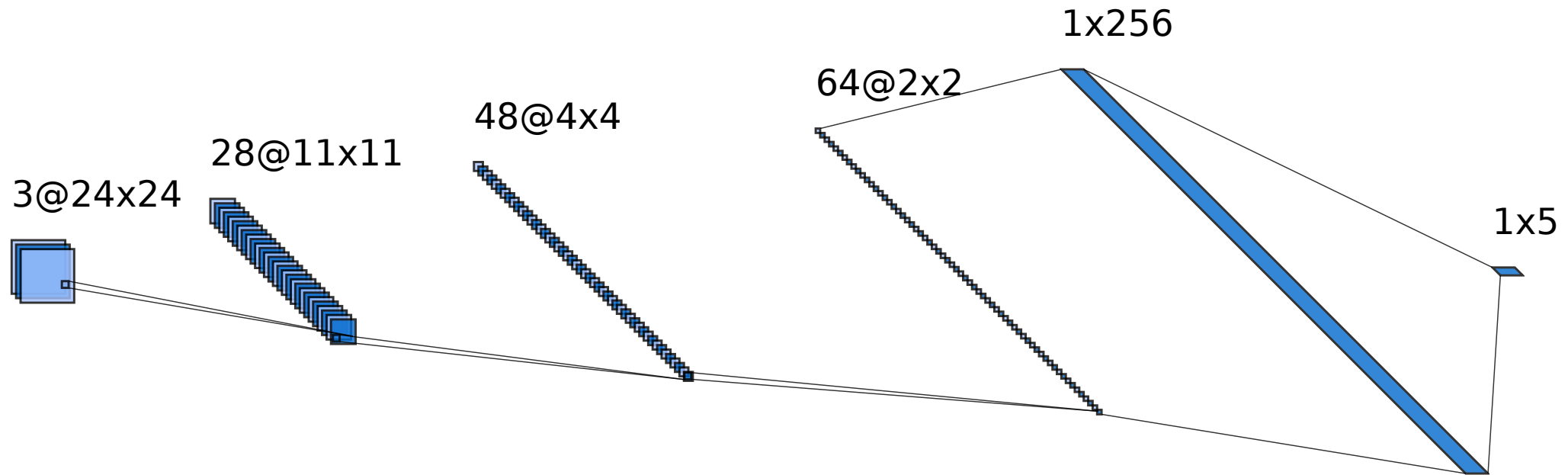
Network structure RNet

Even smaller than Onet, input image of 24x24 pixels

Adjustment needed for MAXIM78000:

Changed kernel size of last convolutional layer from 2x2 to 3x3

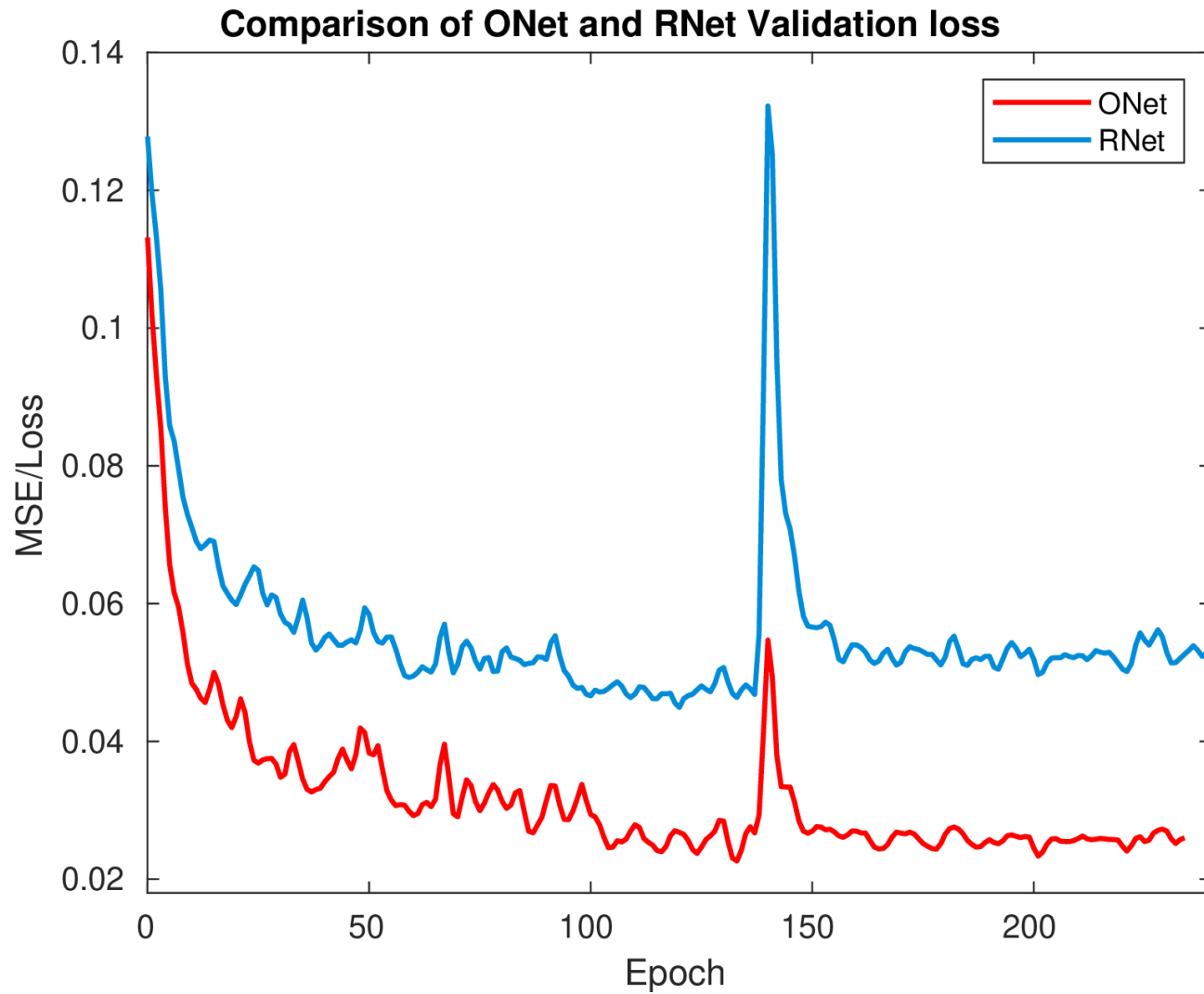
Network structure RNet



Conv + ReLU MP + Conv + ReLU MP + Conv + ReLU Flatten

Results

Onet vs RNet



Results

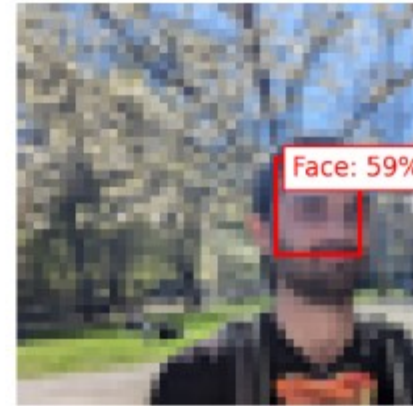
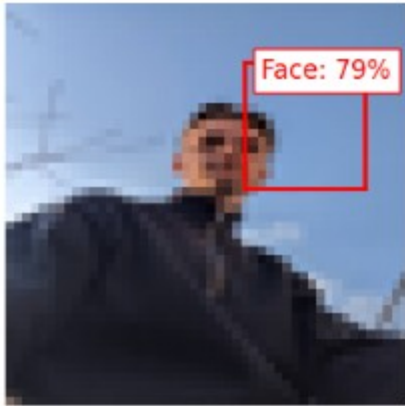
Onet vs RNet

Network	Number of parameters	Final validation MSE loss	Inference time	MACC
ONet	262,240 (59%)	0.026	2416 us	12,905,728
RNet	73,908 (17%)	0.052	518 us	1,559,232
	RNet (3.5x smaller)	ONet (2x better)	RNet (4.7x faster)	RNet (8.3x less)

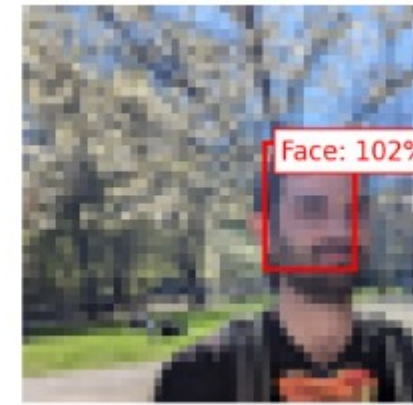
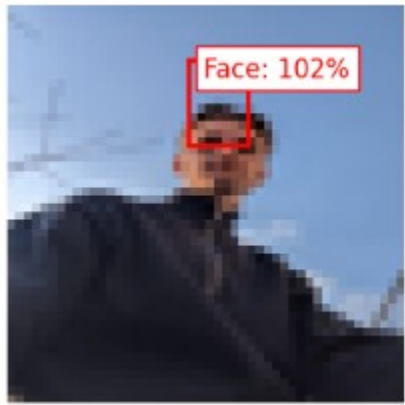
Results

Trained ai8x model on test images

ONet



RNet



Made with the test.npy files and the generated sampleoutput.h files

Live demo

All code available on:

<https://github.com/lionnus/maxim7800-face-detection/>

Lionnus Kesting
Lkesting@student.ethz.ch