How to develop a biometric system on Arm board

William Gao, AI computing Architect

OPEN AI LAB

Self Introduction

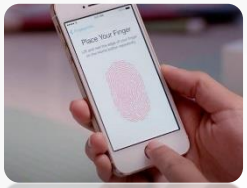
- **William Gao**
  - 2 years experience in AI product and developing — OPEN AI LAB
  - 3 years technical support of Arm Cortex-A/M CPUs — Arm
  - 3 years experience in CPU design — C-sky Microsystem
  - Master degree in computer architecture — Zhejiang University

- **OPEN AI LAB**
  - was established in 2016 to provide integrated AI open infrastructure software and hardware platform, computing operating system and application level solutions for partners in the AIoT industrial chain.

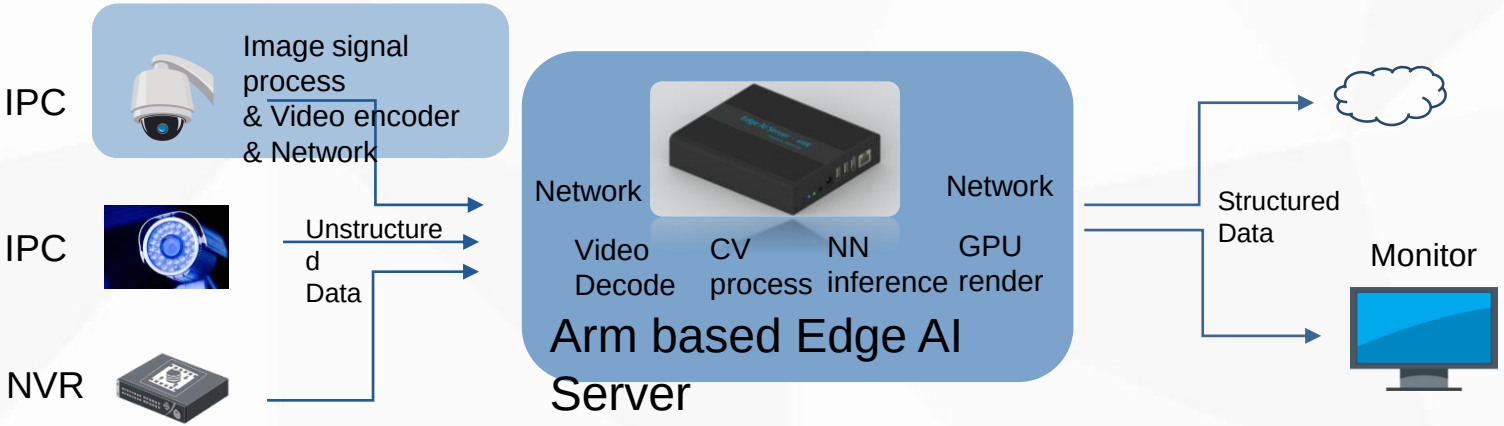
Outline

- Challenge of developing biometric system on Arm board
- Face recognition system deployment on Arm board
- Use Tengine to improve system efficiency

Widely used Arm devices in Bio-system



Challenge of developing on Arm Board

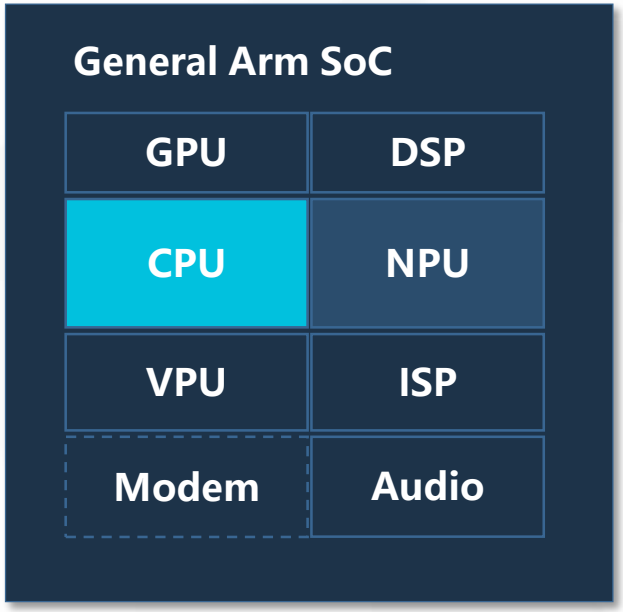


Performance **Cost** **Power**

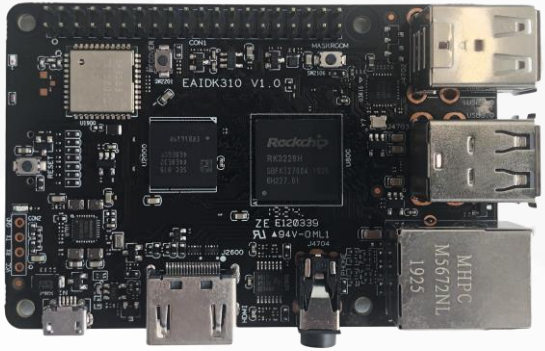
Easy to program

Domain Specific Architecture

- Balanced performance, cost and power
- Different chips have different spec
  - provide rich options
  - also introduce fragment



EAIDK310 -- Example Arm board



SoC RK3228H
CPU Arm Cortex-A53 MP4 · up to 1.3GHz
GPU Arm Mali-450 MP2 GPU OpenGL ES 1.1/2.0
LPDDR3 1GB
Storage 8GB emmc extern MicroSD up to 128GB
Ethernet RJ45 · 10/100M
WIFI 802.11 ac/a/b/g/n, 2.4G/5GHz
Bluetooth 5.0
USB 1xUSB3, 3xUSB2, 1xMicro-USB
HDMI 2.0, 1xType-A, 4Kx2K@60Hz

CV accelerator： resize, crop · jpg codec
H264/H265 decoder：4K@60fps
H264/H265 encoder：1080p@30fps

Face access control system on Arm board

2016-2017

10Tflops $3000

2018-2019

48Gflops $30

Less Bandwidth

1:N N=10,000 face database
0.5s response time

System design and algorithm choose

Pic resize &
yuv->rgb
640x480

<40m
s

Face Detection

Pic transform&
crop & resize
112x112

100 - 300ms

Feature Exaction

<20m
s

Features comparison

Pass

<v0,v1,v2,v3,…,v127>

Real Time requreiment 25fps

Get 5 pints landmark of Largest Face

112x112 Face - > 128d vector

Find match vector in 10K database

NN detection + CV Trace

MTCNN
YuFace

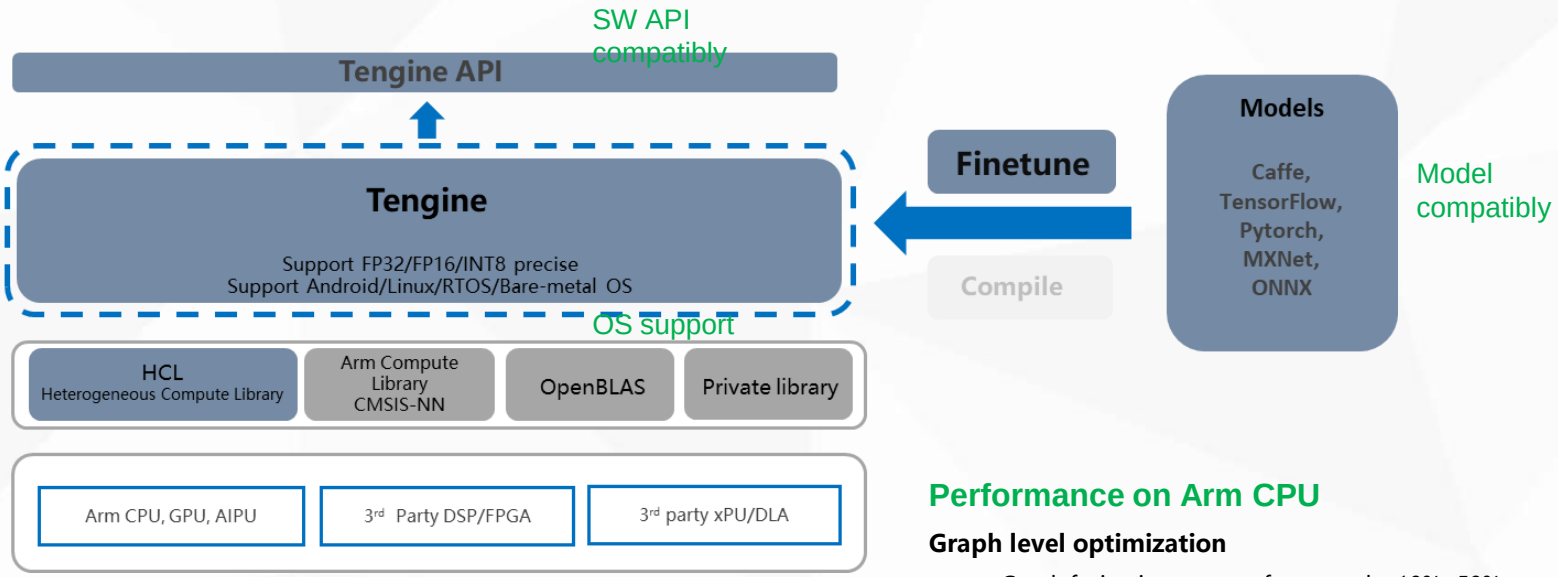
Neural Network inferencing

MobileFaceNet
ResNet50

Vector multiply matrix
Read 5MB data and compare

Cosine Similarity

Use Tengine to accelerate NN deployment



Performance on Arm CPU

Graph level optimization

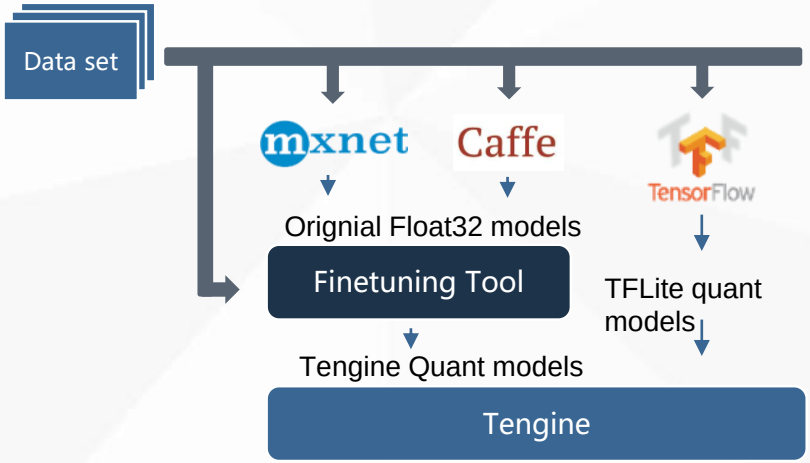
Graph fusing improve performance by 10%~50%

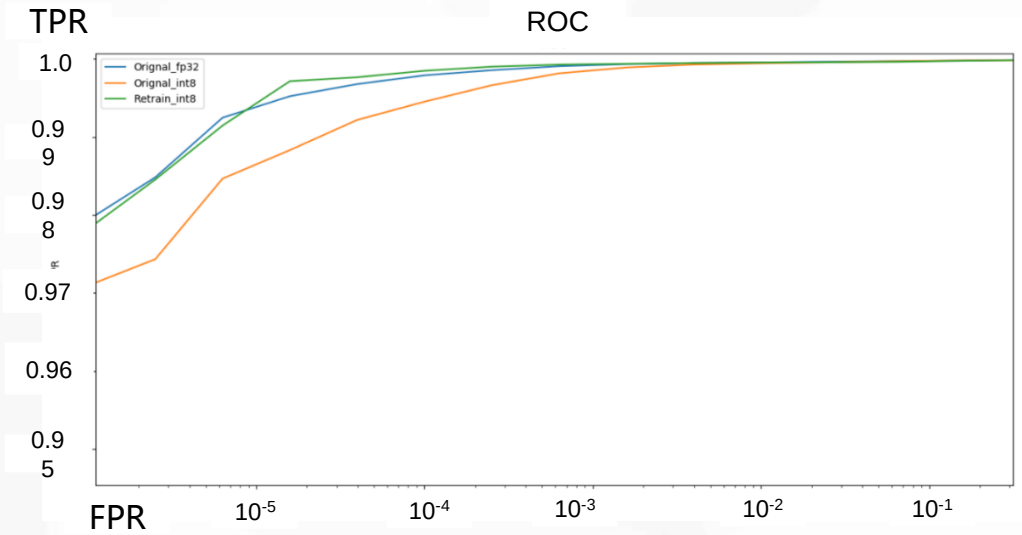
HCL(Heterogeneous Compute Library)

state of art optimization at micro-arch level for all Arm Cortex-A processors

Support end2end/mix mode FP32/FP16/INT8 inferencing, INT8 mode is 50-90% faster than FP32 mode

Quantization Finetuning Tools ensure accuracy





1200W 1:1 testing result on LFW

INT8 inferencing + Quant Finetuning

Successfully enlarge face database from 20K to 50K without performance and accuracy penalty

Make use of NPUs

Performance boosting
0.1T -> 4-8T ops
Yolov3 2s -> 100ms

Private architecture & Toolchain
Un-supported model &operators
Uncontrollable Precise drop

RK3399Pro 3T
RK1808 3T

Hi3559av100 4T
Hi3519av100 2T
Hi3516dv300 1T
Hi3516cv500 0.5T
…

865 15T
855 7T
845 3T

Edge TPU

More applications powered by Tengine

Support partners to build the best and cost-effective solution

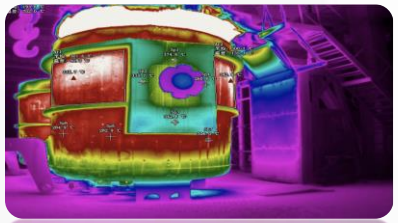
Face Capture Recognition
Pan Security Case

Helmet Detection
Case of Safety Supervision

Water Quality Detection
Intelligent Transformation of
Beacon Light

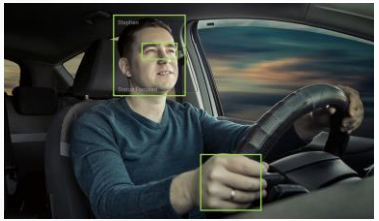
Warehousing Case

Assembly Line Sorting Case



Metallurgical Case

Agriculture and Animal
Husbandry Case



Smart Cockpit Case

OPEN AI LAB

Q & A

OPEN AI LAB · **AI** Anything

OPEN AI LAB Official Wechat

market@openailab.com

www.openailab.com

3 / F, building B8, No.188, Yizhou Road, Xuhui District, Shanghai

8 / F, building B, Tsinghua Ziguang information port, No. 13, Langshan Road, Nanshan District, Shenzhen

9 / F, ideal international building, No. 58, Beisihuan West Road, Haidian District, Beijing

5 / F, South Building, building C, financial science information center, No. 2, Xueyuan South Road, Haidian District, Beijing