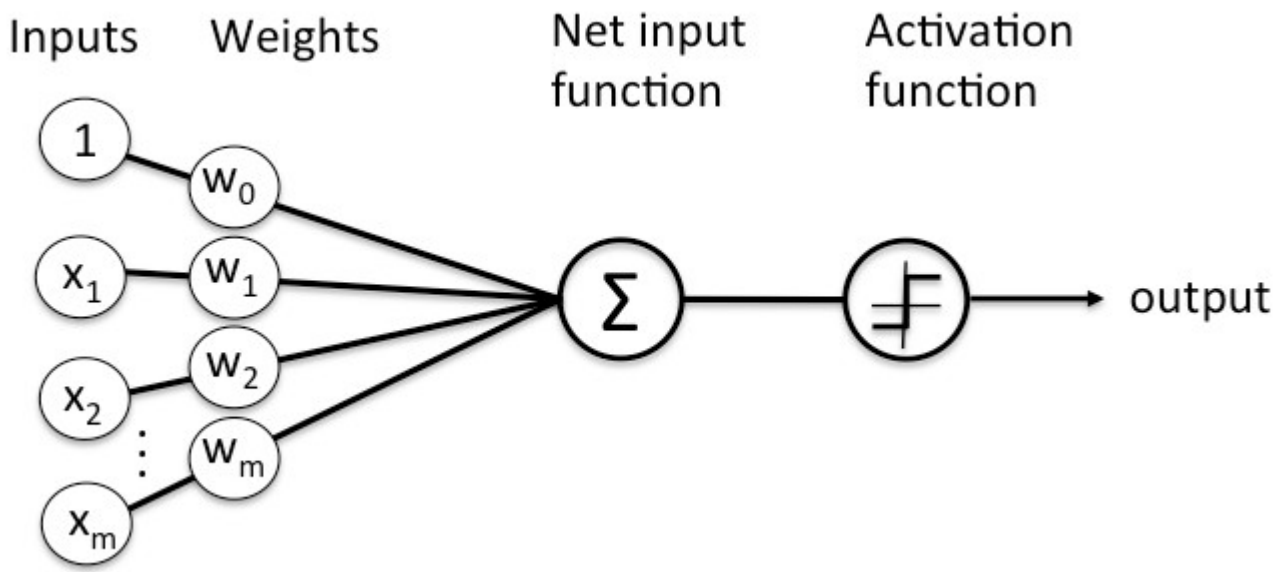
- AIoT & RISC-V
- Automatic speech recognition
- ASR demo

AIoT

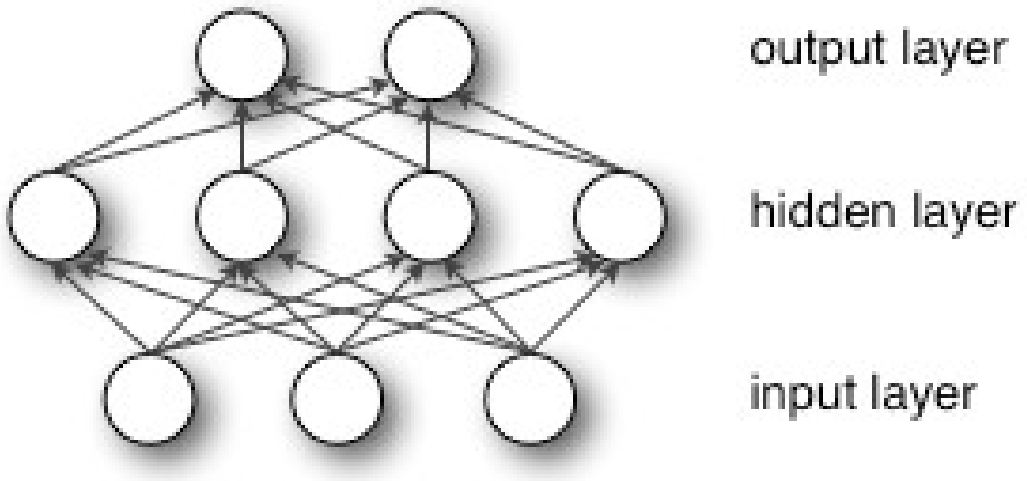
Artificial Intelligence of Things (AIoT) is the combination of artificial intelligence technologies with the Internet of Things

Neural Networks

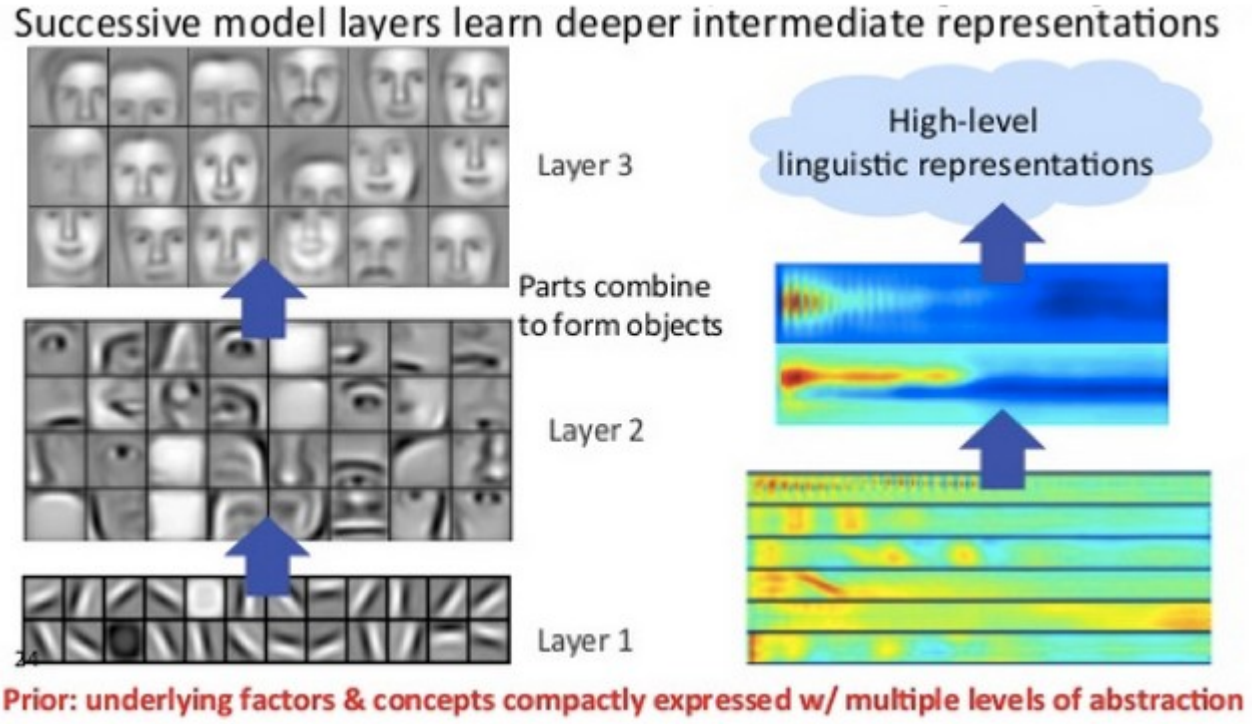


https://skymind.ai/wiki/neural-network

Hidden layers



Deep Neural Networks

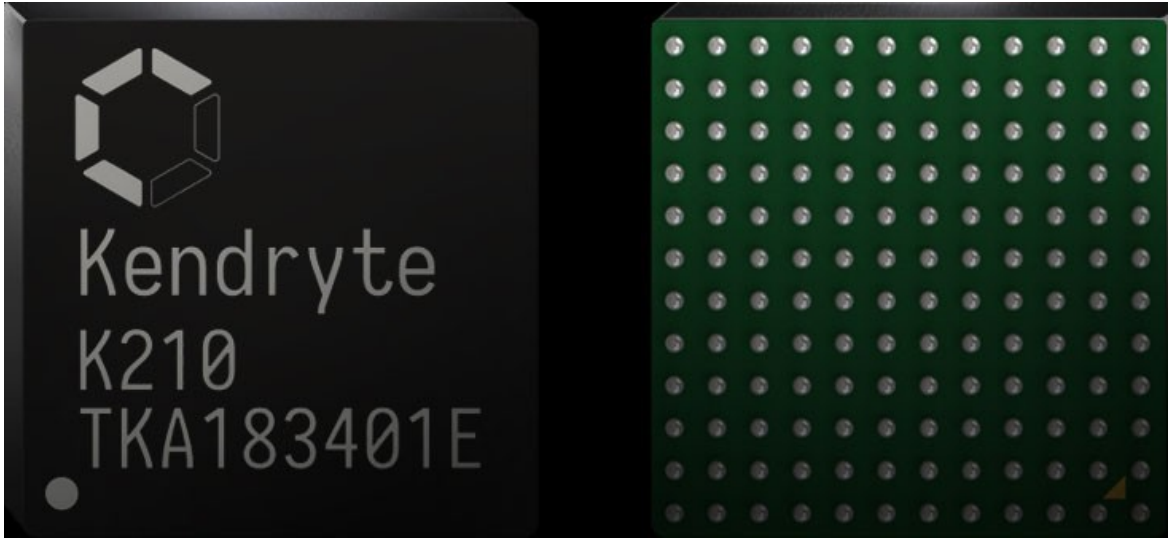


More than one hidden layer. Deeper understanding for each layer

SiPEED AIoT module

K210

https://kendryte.com/



K210

RISC-V Dual Core 64bit, with FPU
Typical application scenarios < 1W
Power consumption of chip < 300mW
OS: FreeRTOS
NN Model: TinyYOLOv2 (after pruned)
DNN Framework: TensorFlow/Keras/Darknet
Peripherals FPIOA/UART/GPIO/SPI/I²C/I²S/WDT/
TIMER/RTC, etc.

RISC-V

RISC-V
Is a opensource
Instruction Set Architecture (ISA) spec
Base
Extensions

RISC-V Base

- RV32I Base Integer Instruction Set, 32-bit
- RV32E Base Integer Instruction Set (embedded), 32-bit, 16 registers
- RV64I
- RV128I

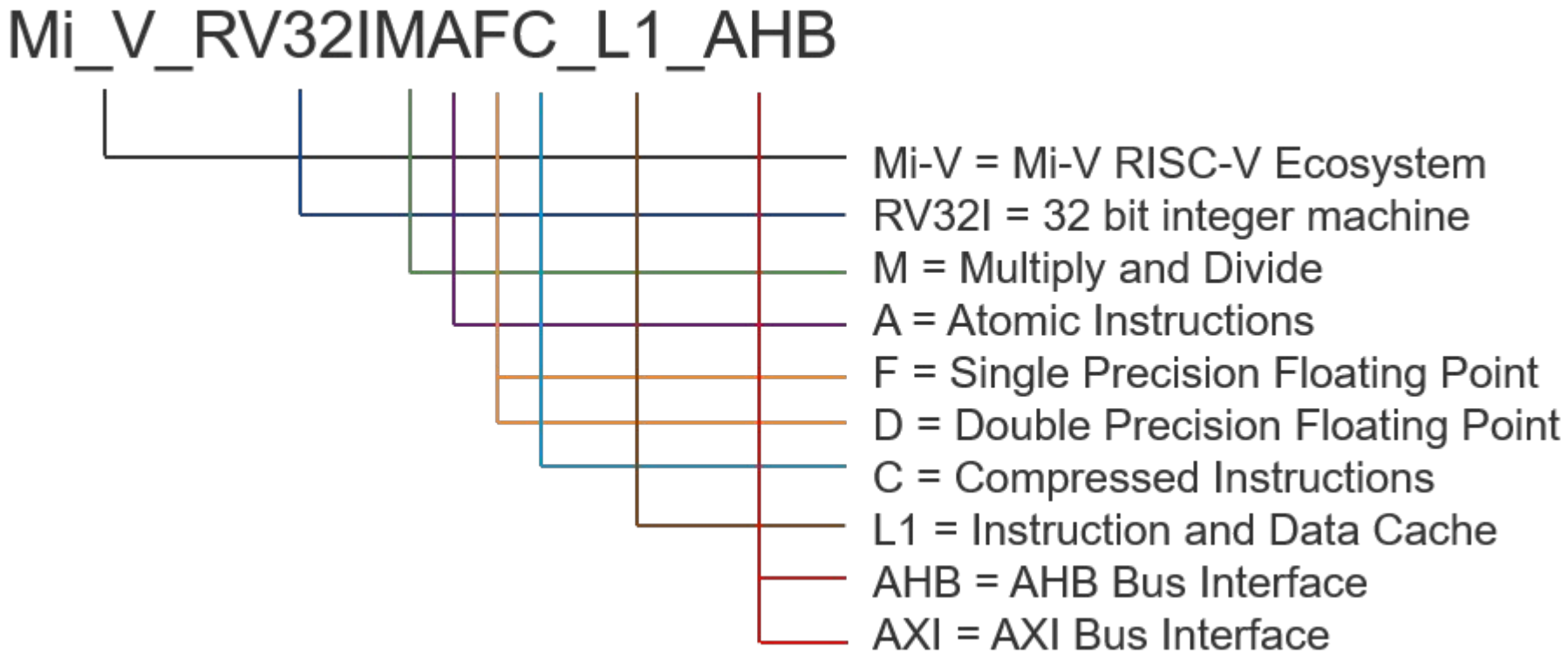
RISC-V Extension

C , Standard Extension for Compressed Instructions

F, Standard Extension for Single-Precision Floating-Point

M,A,D,G,Q,L,C,B,J,T,P...

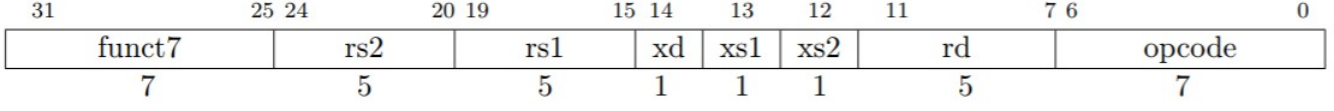
https://www.microsemi.com



ISA extensions

- Chapter 9 in the ISA manual
- 4 major opcodes set aside for non-standard extensions (Table 8.1)
  - Custom 0-3
  - Custom 2 and 3 are reserved for future RV128
- RoCC interface uses this opcode space
  - 2 source operands, 1 destination, 7 bit funct field
  - 3 bits(xd,xs1,x2) determine if this instruction uses the register operands, and passes the value in register rs1/2, or writes the response to rd

inst[4:2] inst[6:5]	000	001	010	011	100	101	110	111 (> 32b)
00	LOAD	LOAD-FP	*custom-0*	MISC-MEM	OP-IMM	AUIPC	OP-IMM-32	48b
01	STORE	STORE-FP	*custom-1*	AMO	OP	LUI	OP-32	64b
10	MADD	MSUB	NMSUB	NMADD	OP-FP	*reserved*	*custom-2/rv128*	48b
11	BRANCH	JALR	*reserved*	JAL	SYSTEM	*reserved*	*custom-3/rv128*	≥ 80b

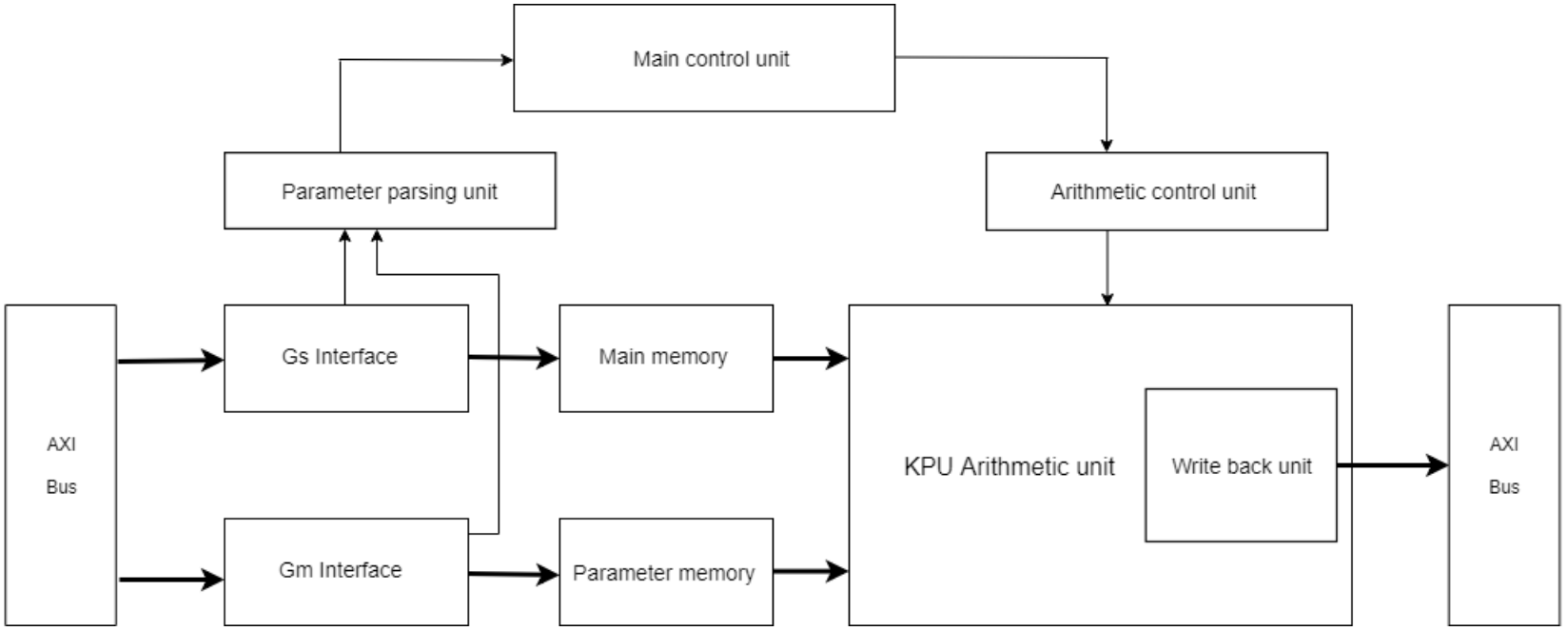


K210 standard extensions

RV64GC shorthand for RV64IMAFDC

C , Standard Extension for Compressed Instructions
F, Standard Extension for Single-Precision Floating-Point
D, Double-Precision Floating-Point

K210 KPU


Main control unit
Parameter parsing unit
Arithmetic control unit
AXI
Bus
Gs Interface
Main memory
KPU Arithmetic unit
Write back unit
AXI
Bus
Gm Interface
Parameter memory

Training and Inference

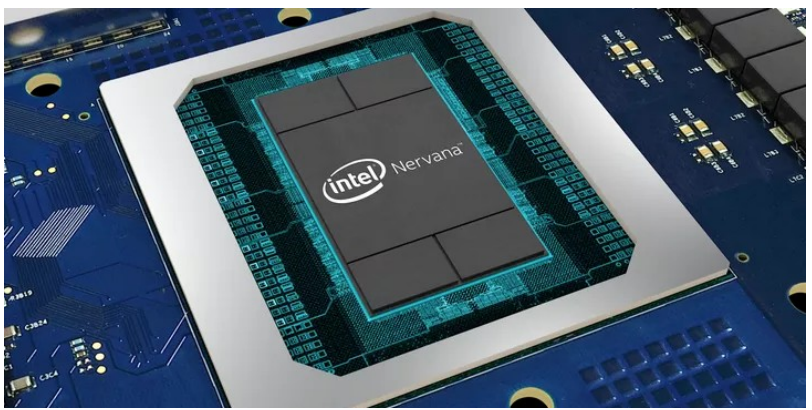
kendryte-model-compiler

```
python3 . --dataset_input_name input:0 \
 --dataset_loader dataset_loader/img_0_1.py \
 --image_h 240 --image_w 320 \
 --dataset_pic_path dataset/yolo_240_320 \
 --model_loader model_loader/pb \
 --pb_path pb_files/20classes_yolo.pb
--tensor_output_name yv2
```

KPU characteristics

- Supports the fixed training model that the common training framework trains according to specific restriction rules
- There is no direct limit on the number of network layers, which supports separate configuration of each layer of convolutional neural network parameters, including the number of input and output channels, input and output line width and column height.
- Support for two convolution kernels 1x1 and 3x3
- Support for any form of activation function
- The maximum supported neural network parameter size in real-time work is 5.5MiB to 5.9MiB
- Maximum support network parameter size when working in non-real time is (Flash capacity) - (software size)

Edge vs Cloud



CMSIS-NN

Cortex Microcontroller Software Interface Standard

- Convolution, depth separable convolution, fully-connected
- LSTM and GRU also possible.

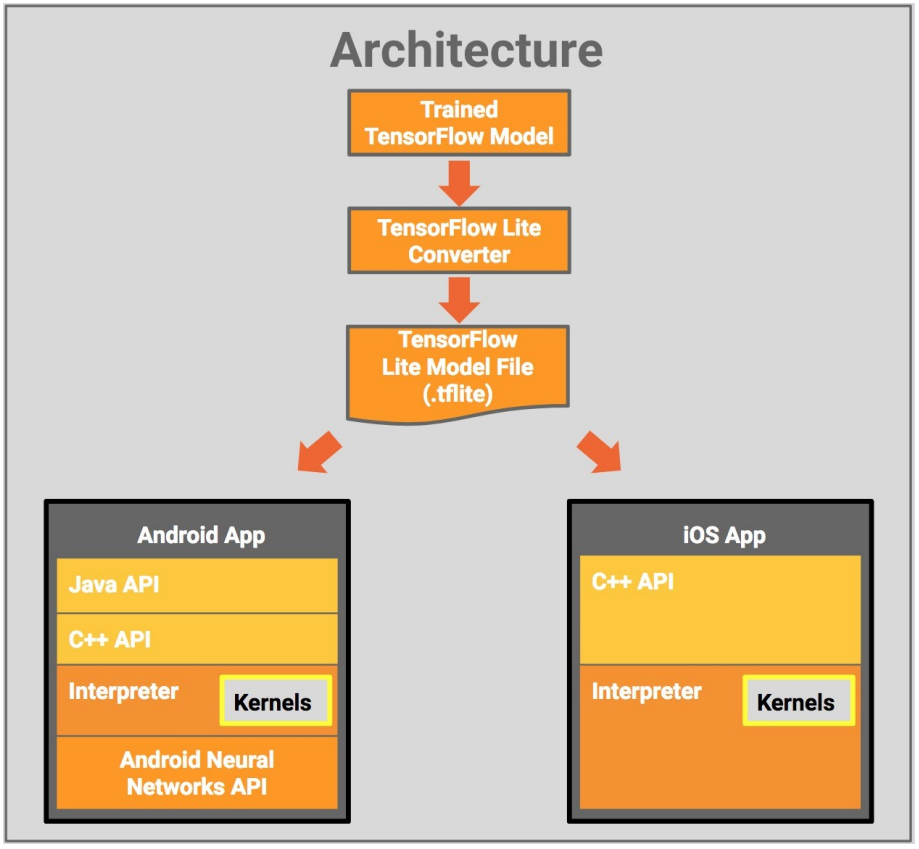
- Train with TensorFlow or Caffe then the weights and biases will first be quantized to 8 bit or 16-bit integers then deployed to the microcontroller for inferencing.
- 4.6X improvement in runtime/throughput
- Since version 5.5 of - Cortex Microcontroller Software Interface Standard
- All Cortex-M based systems that
- support SIMD instructions, especially 16-bit Multiply-and-Accumulate (MAC) instructions
- (e.g. SMLAD) which are very useful for NN computation

Sony Neural Network Libraries

NNabla

https://nnabla.org/

Tensorflow Lite

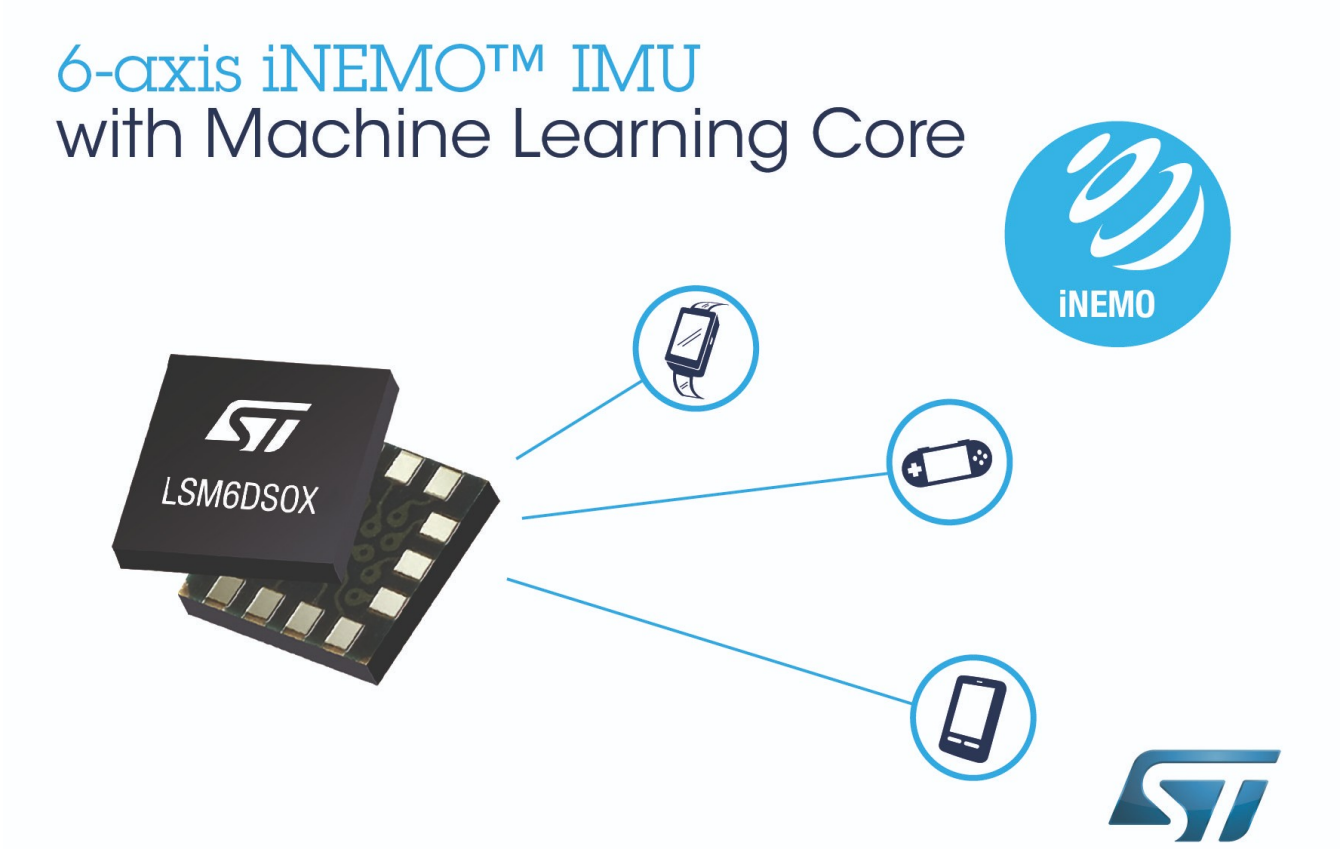


MEMS, iNemo

- ST LSM6DSOX
- inertial measurement unit (IMU), with Machine Learning

  Contains a machine-learning core to classify motion data based on known patterns. Relieving this first stage of activity tracking from the main processor saves energy and accelerates motion-based apps such as fitness logging, wellness monitoring, personal navigation, and fall detection.

iNEMO



Neural processors (China)

Alibaba Huwaei	Ali-NPU Ascend
Baidu	Kunlun
Bitmain	Sophon
Cambricon	MLU
Kendryte	K210

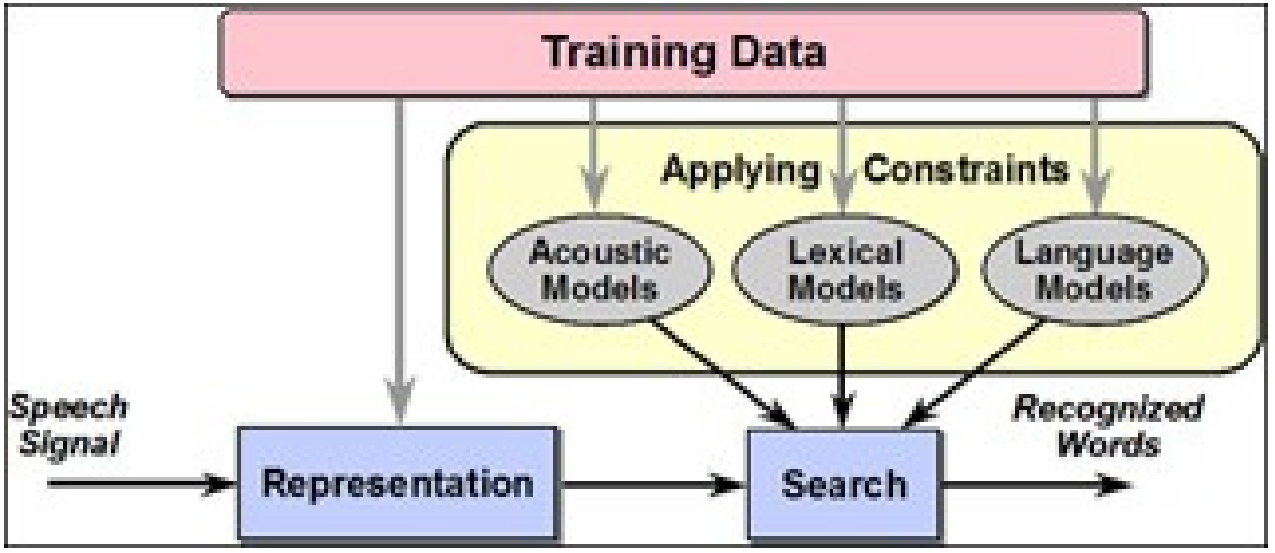
Neural processors (USA)

Amazon	AWS Inferentia
Google	TPU
Intel	NNP, Myriad, EyeQ,Nervana
Nvidia	NVDLA
Apple	Neural Engine

Neural processors (Other)

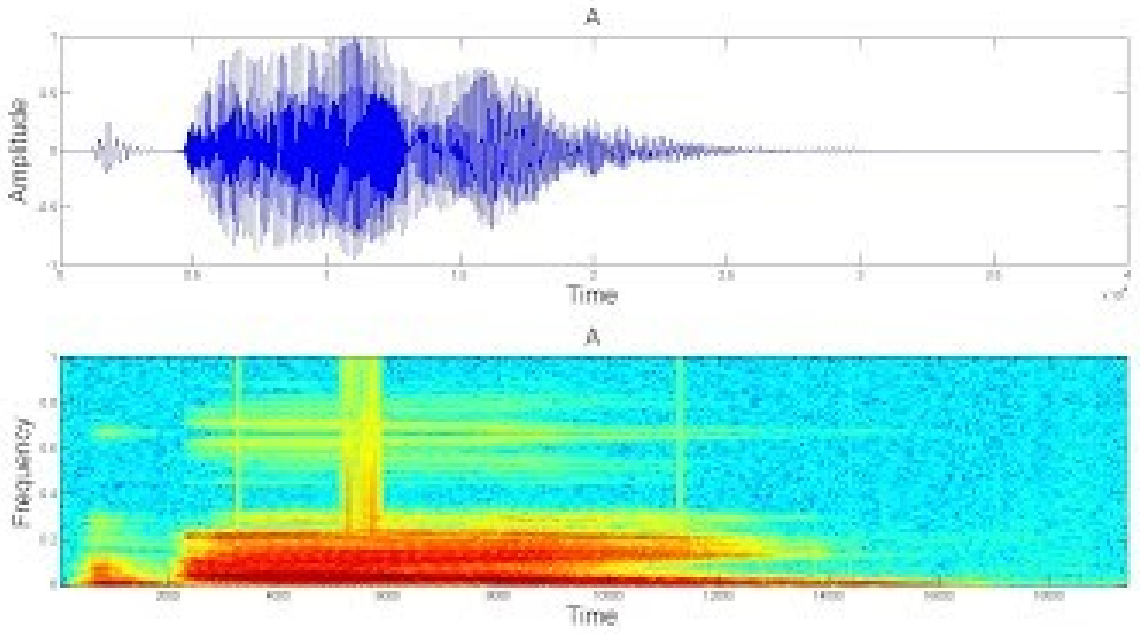
Graphcore	IPU
Samsung	Neural Processing Unit (NPU)
Groq (USA)	Supported by Google
CEVA (USA)	Neu Pro

Speech recognition

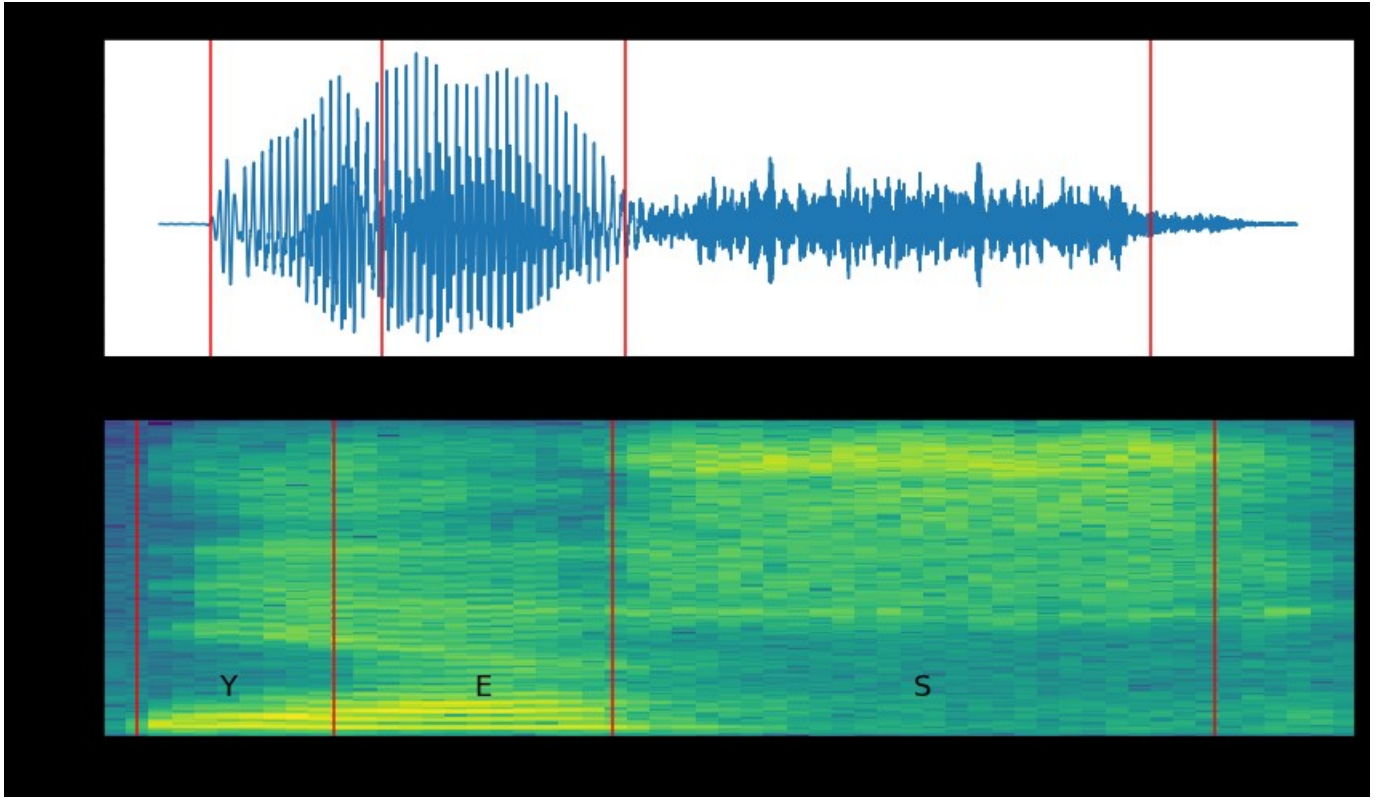


Mel Frequency Cepstrum Coef.

Mel scale is a scale that relates the perceived frequency of a tone to the actual measured frequency.

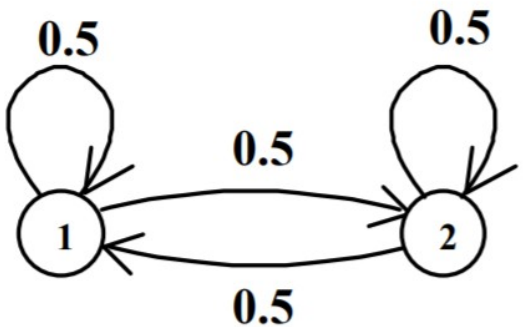


Spectrogram



HMM

Single Fair Coin



P(H) = 1.0 P(H) = 0.0

P(T) = 0.0 P(T) = 1.0

Pocket Sphinx

A version of Sphinx that can be used in embedded systems

git clone https://github.com/cmusphinx/sphinxbase.git

git clone https://github.com/cmusphinx/pocketsphinx.git

./autogen.sh
make

src/programs/pocketsphinx_continuous -inmic yes -hmm model/en-us/en-us -dict model/en-us/cmudict-en-us.dict -jsgf simple.jsgf

Grammar, Language model

```
#JSGF V1.0;
grammar all;
public <all> =   turn ( on | off )  light ( one |two | three | four | five | six | seven );
```

Demo

src/programs/pocketsphinx_continuous -inmic yes -hmm model/en-us/en-us -dict model/en-us/cmudict-en-us.dict -jsgf simple.jsgf

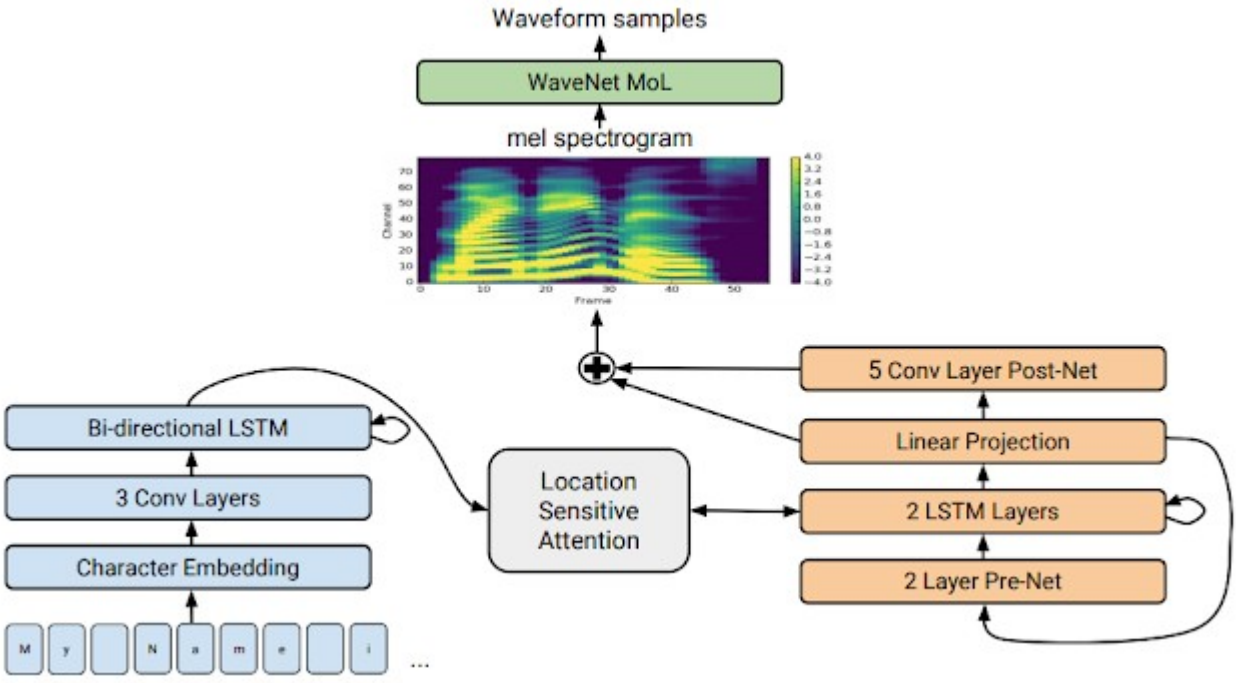
Text to speech

Is difficult

Text-to-phonetic or grapheme-to-phoneme conversion is difficult

Homographs, spelled the same way but they differ in pronunciation

Tacotron 2


Waveform samples
WaveNet MoL
mel spectrogram
5 Conv Layer Post-Net
Linear Projection
2 LSTM Layers
2 Layer Pre-Net
Location Sensitive Attention
Bi-directional LSTM
3 Conv Layers
Character Embedding
M
y
N
a
m
e
i
...

Tacotron 2 + Waveglow

Nvidia Waveglow

https://google.github.io/tacotron/publications/tacotron2/index.html

Google DNN for text to speech

Top AI/NN companies

Chinese: Baidu, Alibaba, and Tencent

American: Google, Amazon, Microsoft, Facebook, Intel, Nvidia

Demo

Demo on https://colab.research.google.com

Amazon Polly

https://d1.awsstatic.com/product-marketing/Polly/voices/LandingPage_OpeningParagraph_Brian.c584cee526350c7b2207fc39e356d9d57aea0772.mp3

Demo 2

https://voicebot.ai/2018/11/29/neural-text-to-speech-development-allows-alexa-to-sound-more-realistic/

https://towardsdatascience.com/ok-google-how-to-do-speech-recognition-f77b5d7cbe0b

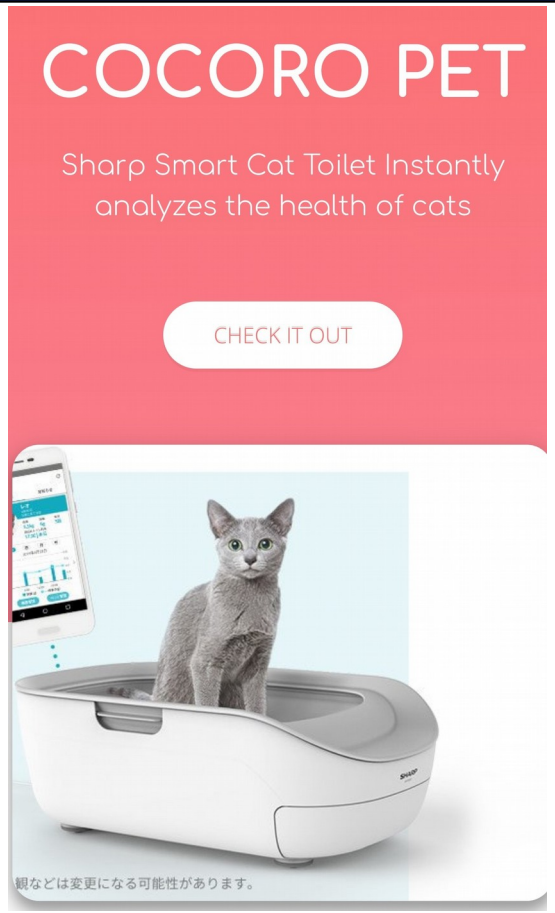
https://assistant.google.com/services/a/uid/0000007a9e6998fb?hl=en-US

Ok google, let me talk to Dungeon RPG

https://github.com/mitchellgordon95/DungeonForGoogleHome

Links

https://www.seeedstudio.com/Sipeed-M1-dock-suit-M1-dock-2-4-inch-LCD-OV2640-K210-Dev-Board-1st-RV64-AI-board-for-Edge-Computing-p-3211.html



https://affinelayer.com/pixsrv/

Comparing Griffin-Lim and WaveRNN Spectrogram Inversion

https://google.github.io/tacotron/publications/adv_tts/index.html

To be tested,

https://github.com/fatchord/WaveRNN

https://github.com/geneing/WaveRNN-Pytorch/tree/master

https://github.com/NervanaSystems/distiller