ARM and Machine Learning

ARM

Jem Davies
ARM Fellow & VP Technology
Media Processing Group, Imaging & Vision Group

December 12 2016

Who am I?

- Jem Davies - ARM Fellow and VP of Technology - Media Processing and Imaging & Vision Groups
  - Working on multimedia, computer vision and machine learning
- 13 years at ARM
- Mathematician turned chemist, turned software engineer…
  - … turned architect
  - … turned tech future predictor
- Glider pilot instructor, fireworks lighter and scuba diver…
- … what next?

ARM

Machine Learning overview

- Machine Learning makes smart connections to previously encountered concepts
- It's useful when:
  - We don't have algorithms…
  - but we do have a lot of data

Image recognition

Robotics

Home security

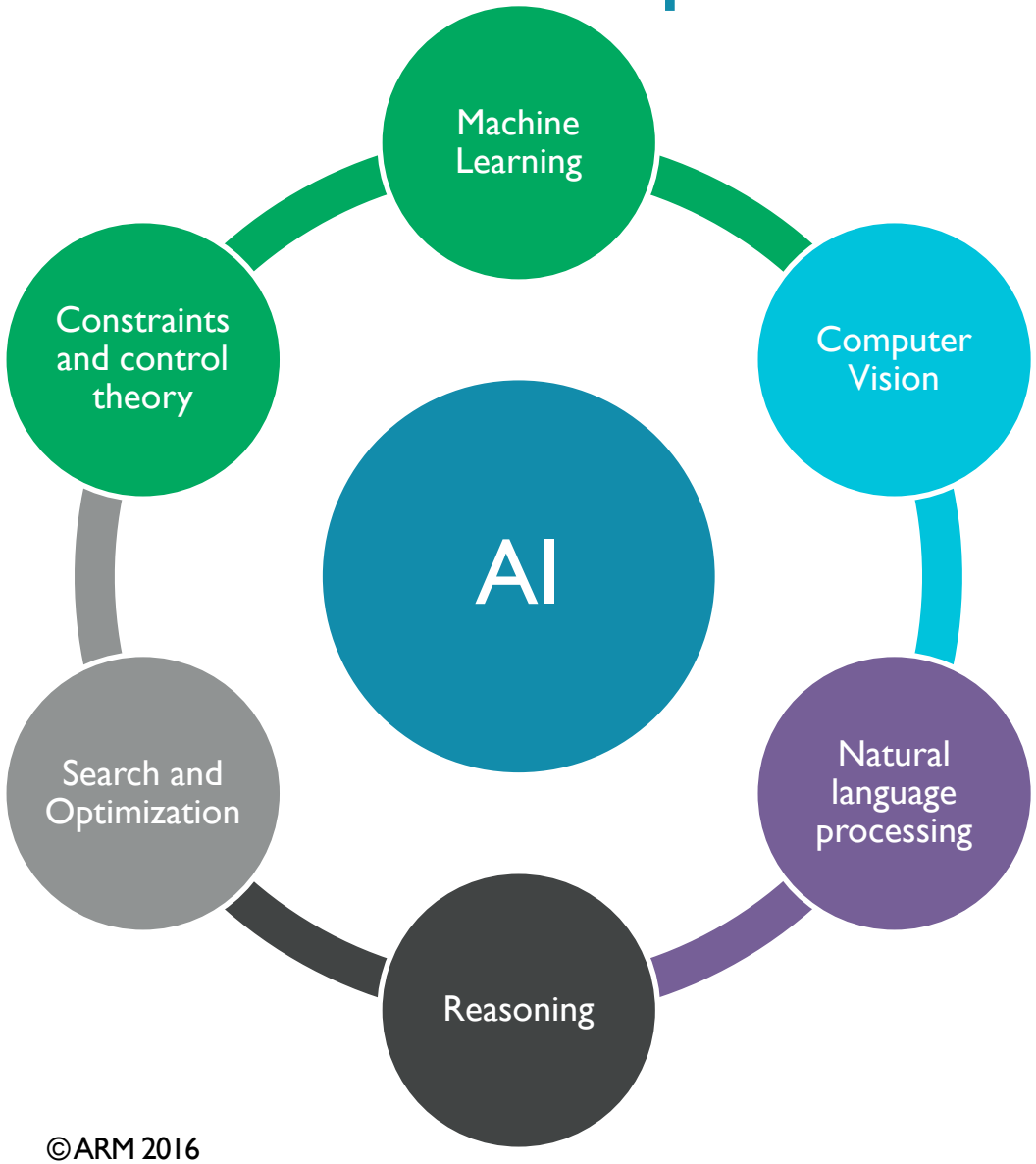
Speech recognition

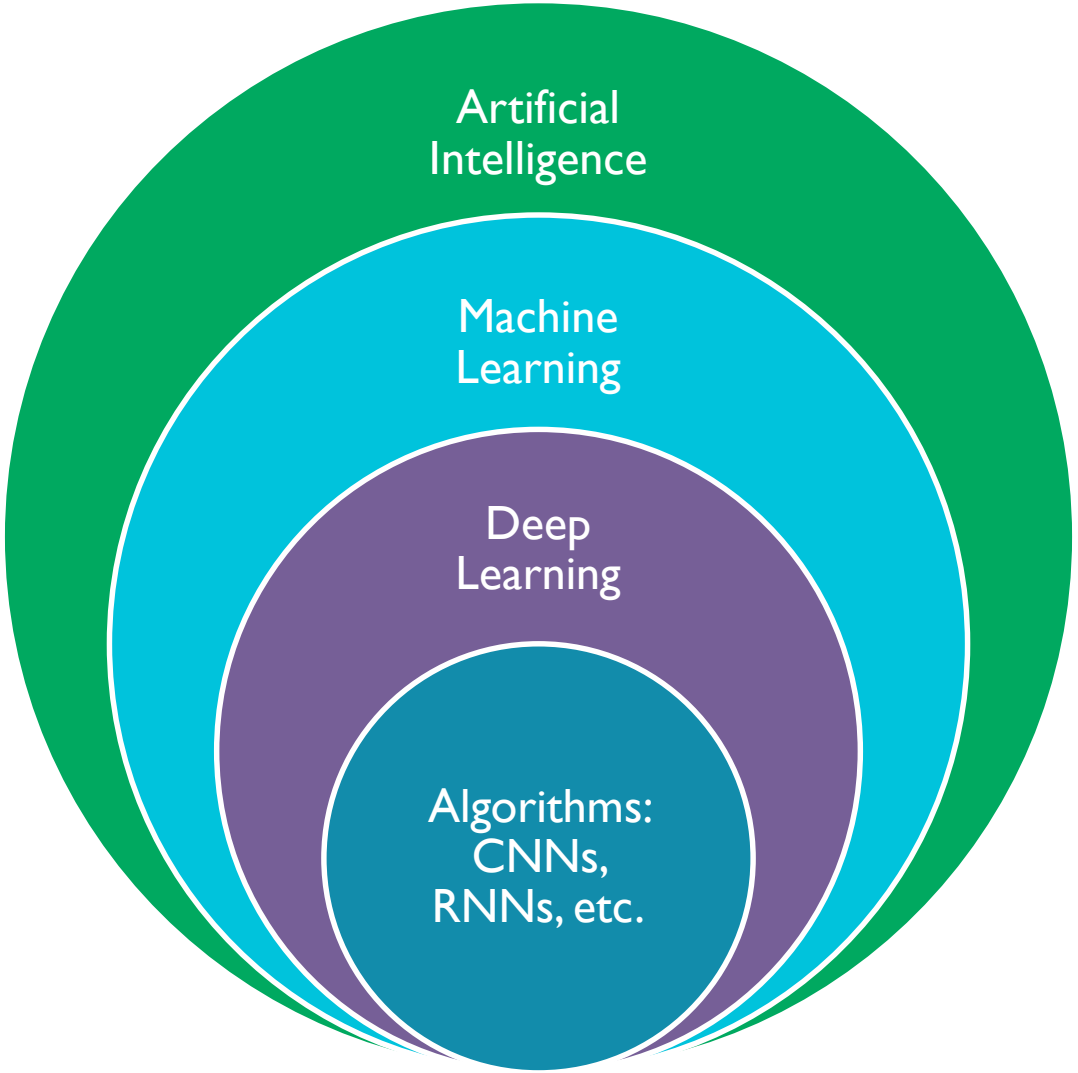
We use Machine Learning technology every day

ARM

Definitions and recent developments

The AI landscape





ARM

Key terms and definitions

Artificial Intelligence
- The broadest term - applying to any technique enabling computers to mimic human intelligence

Machine Learning
- A subset of AI including techniques enabling computers to improve at tasks with experience. Includes deep learning

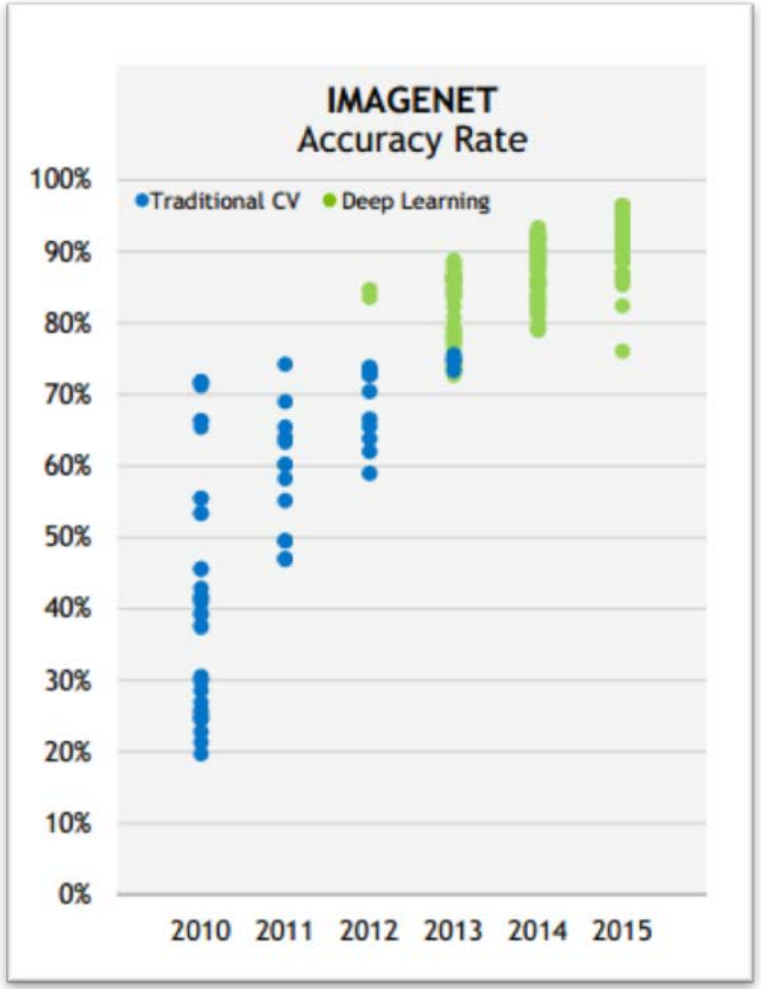
Deep Learning/Neural Networks
- A branch of machine learning that attempts to model real life ideas in data by using a deep graph with multiple processing layers

Algorithms
- DNNs, CNNs, RNNs, SGEMM etc.

Computer Vision
- An interdisciplinary field that allows computers to gain understanding from digital images or videos. Many computer vision applications use ML algorithms.

Recent developments in deep learning

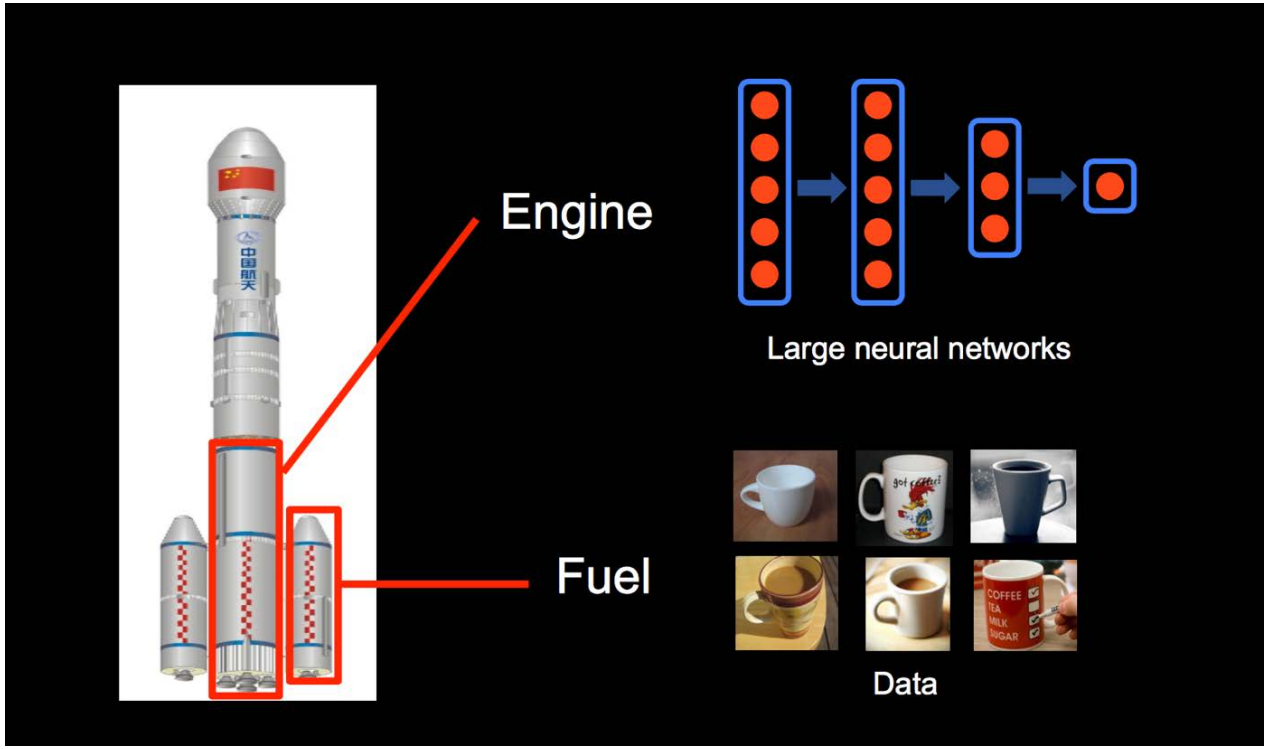
- The image detection benchmark ImageNet has reached nearly 100% accuracy
  - Largely due to Neural Networks
- Research groups feel it's not useful to work on ImageNet anymore (it's a solved problem)
  - GoogLeNet, Inception, etc.
- The successful approach used for ImageNet has spread to other domains, yielding great improvements

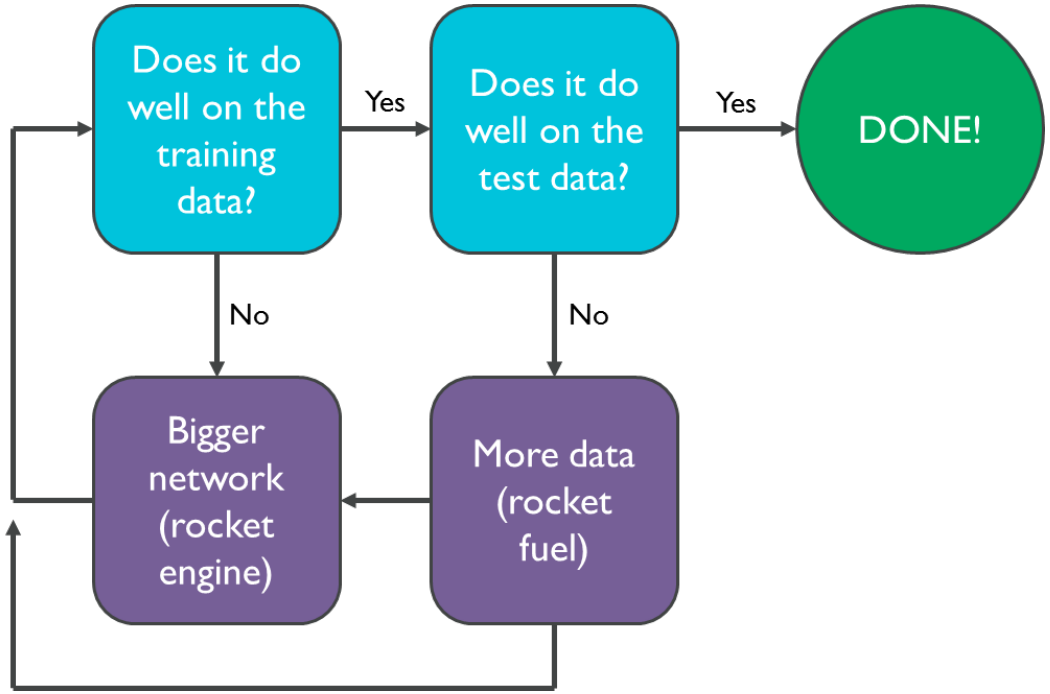


ARM

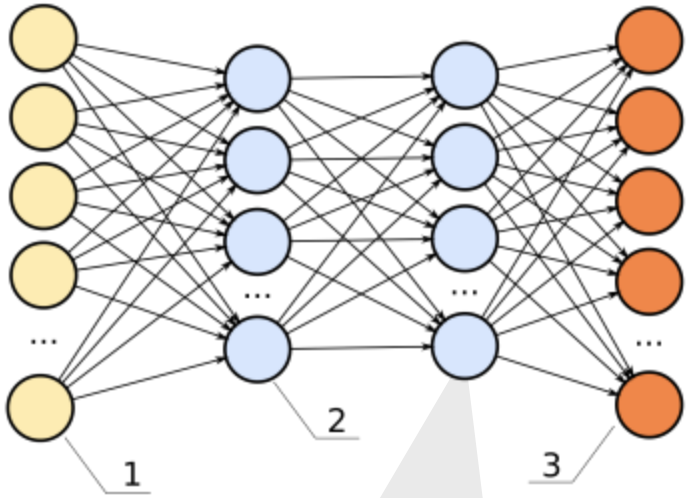
Why did this happen ?

The Machine Learning learning loop

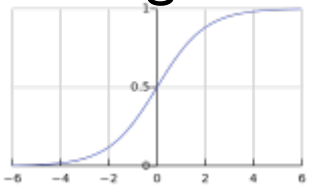




Basic neural network



Squashing function



Sigmoid: y=1/(1+e^(-z))

[Inputs]

x_1 x_2 x_3 x_{n-1} x_n

w_1 w_2 w_3 w_{n-1} w_n

$$z = \sum_{i=1}^{n} w_i x_i$$

$$y = H(z)$$

y [Output]

ARM

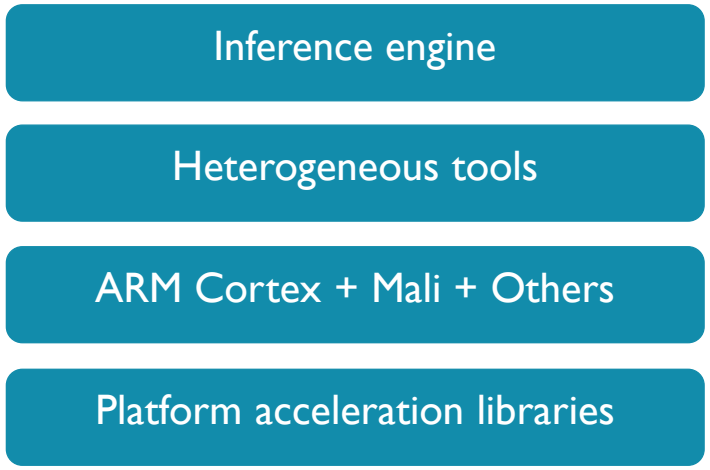
Machine Learning computation in the cloud and on-device

Dissecting the Machine Learning process

Cloud side

Generate inference engines

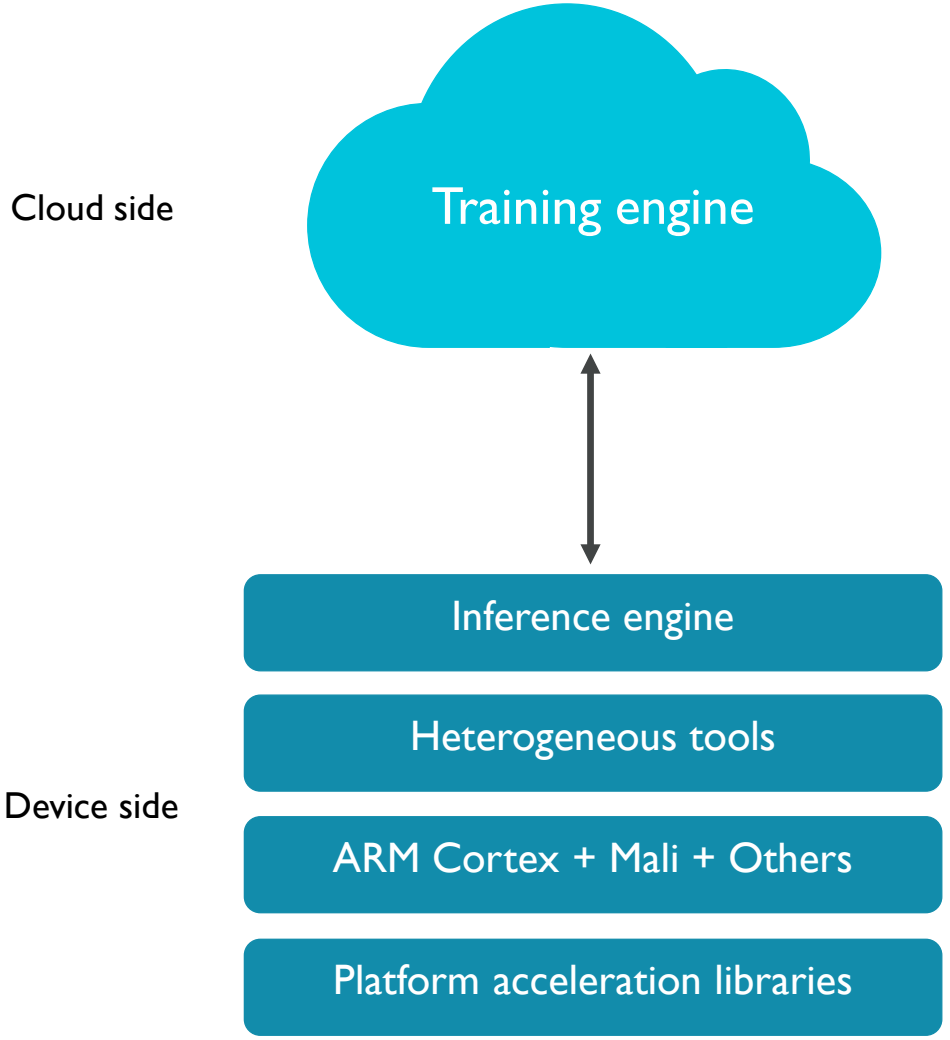
Device side

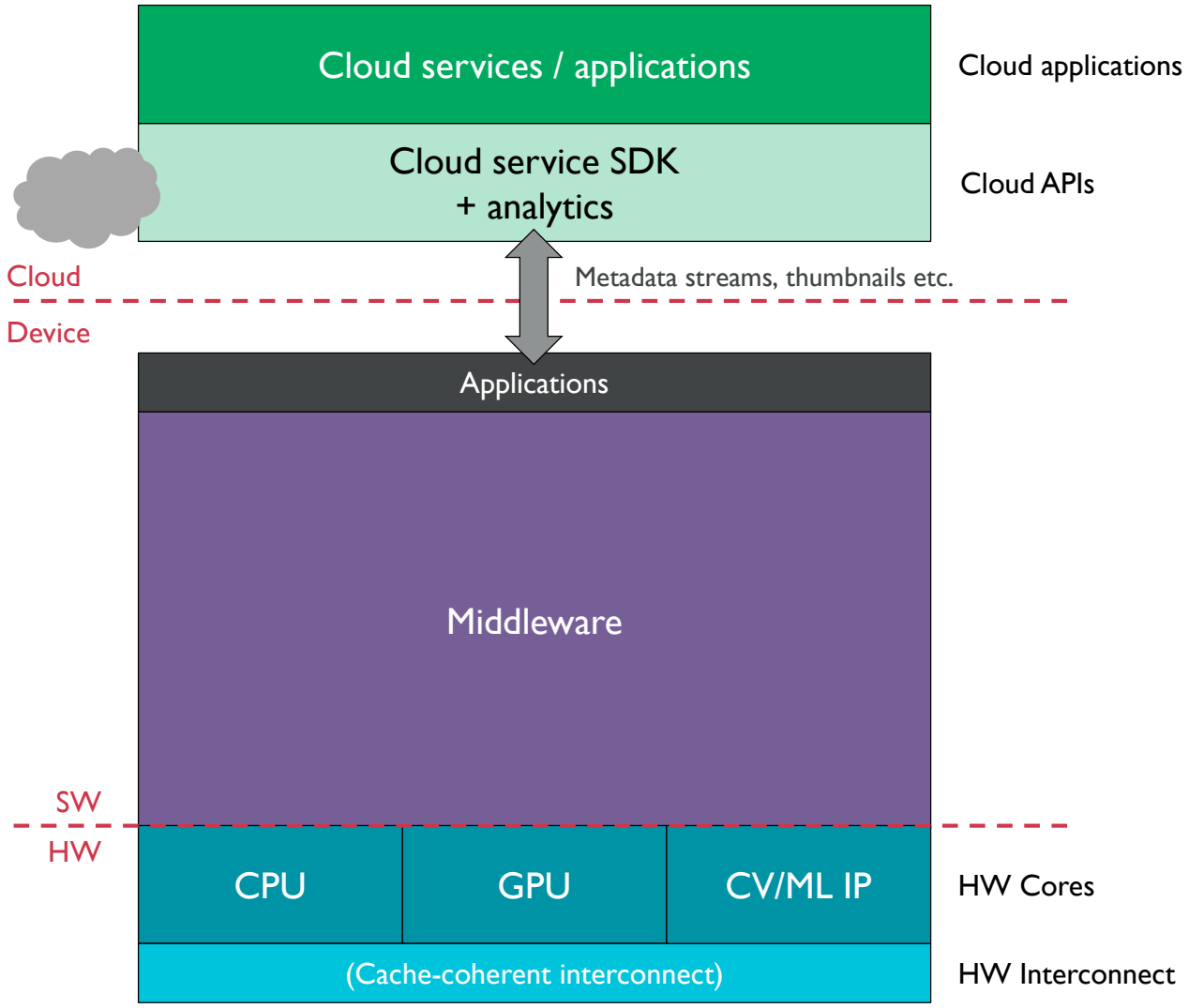


Distribute & optimize

ARM

Machine Learning process on ARM





ARM

It is easy to recognise speech
It is easy to wreck a nice beach

Automatic speech recognition is not easy

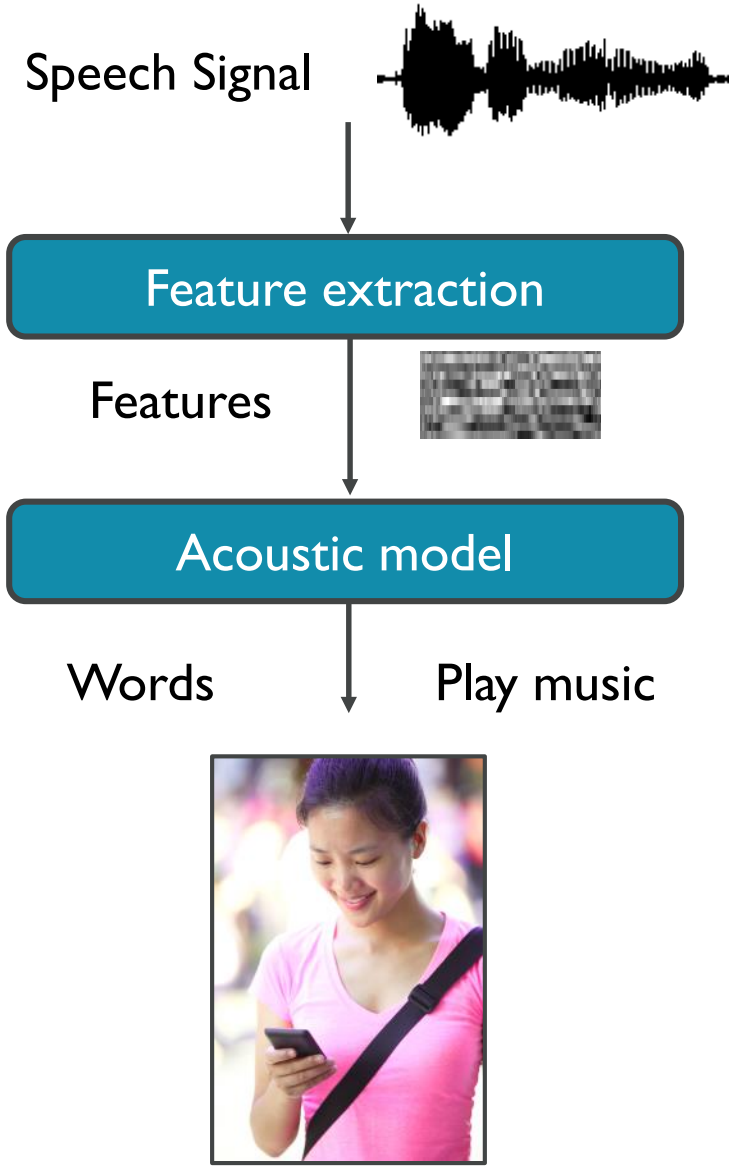
- Active research area for over 50 years!
  - Significant attention recently due to large amount of cloud computing
  - Neural Networks has improved performance
  - Siri, Google Now, Cortana, Amazon Echo now creating usable applications
- Applications
  - Safety and security
  - Interactions with smaller or hands free devices (e.g. wearables)
  - Improved accessibility for hearing-impaired
  - Allows indexing of spoken words

ASR use cases and Machine Learning

- Large Vocabulary Continuous Speech Recognition (LVCSR)
  - Dictation/transcription, virtual assistant
  - Requires dictionary, knowledge of grammar
- Keyword spotting of simple commands
  - "OK Google", "Set alarm for 7"
- Sound monitoring
  - Early/automatic anomaly detection

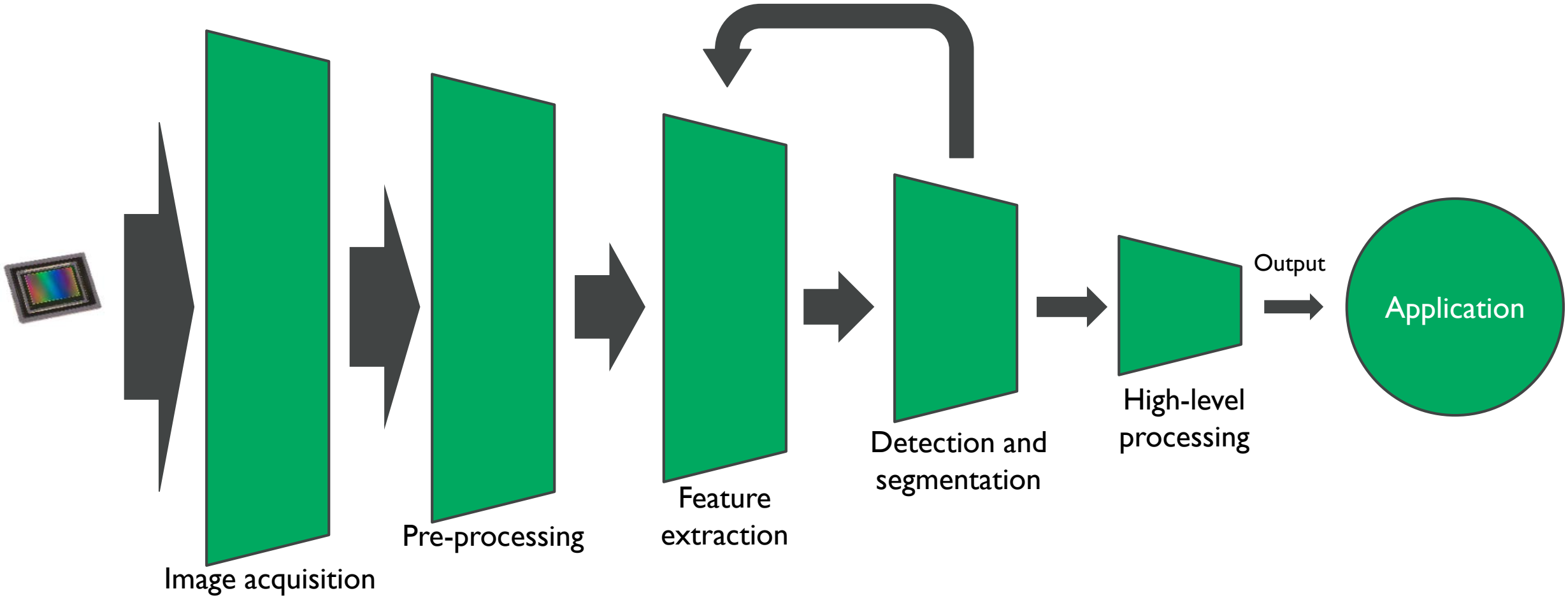
Keyword spotting

- Listen for certain words/phrases
  - Only need to learn certain words
  - "Okay Google"
  - "Play music"
- Simpler algorithm
  - No knowledge of grammar
  - Only needs acoustic model
- Algorithm must run locally on device
  - Potentially always listening
  - Power consumption vitally important



Computer Vision

Computer vision pipeline



ARM

ARM and Machine Learning

Machine Learning runs on ARM-powered client devices today

LUXOFT

ARM

Caffe framework deep learning on ARM

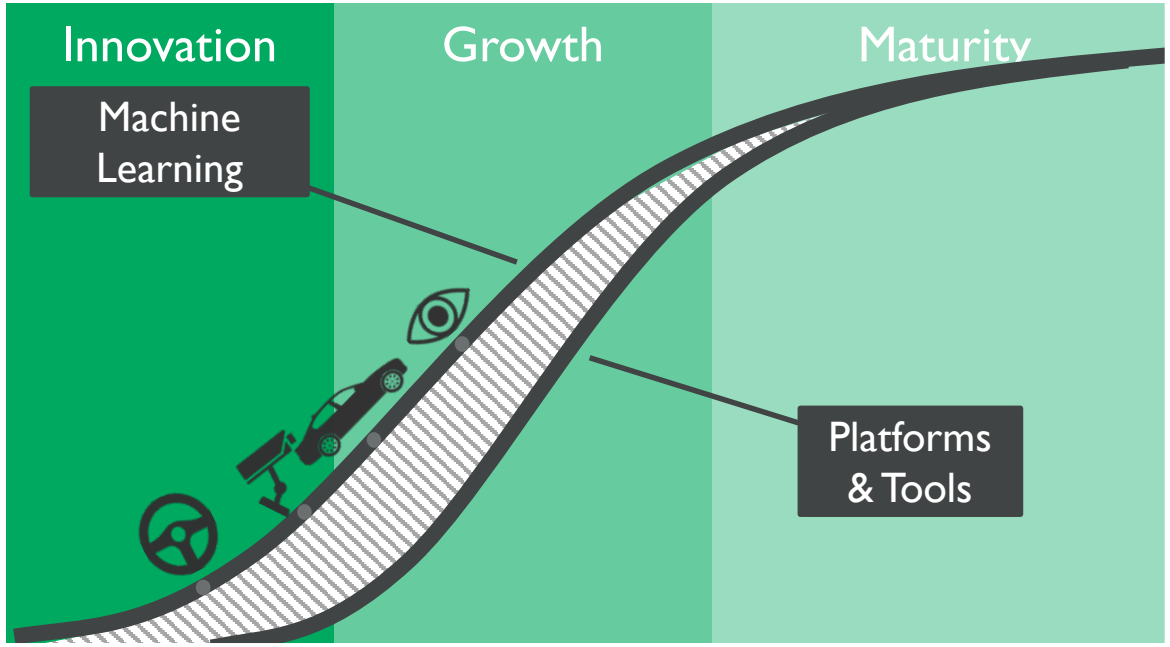
Video hosted on youtube

https://www.youtube.com/watch?v=k4ovpelG9vs&t=13s

Deep learning frameworks on ARM

- DNN used for real-time detection of objects
- Based on the Caffe deep learning framework
- Detection entirely on the mobile device – not cloud based
- Optimized for ARM Cortex CPU or Mali GPU
- Running Machine Learning on the GPU frees up the CPU for other tasks
- Optimised libraries also being developed for TensorFlow and OpenVX

Conclusions



- The rapid uptake in Machine Learning is going mainstream, affecting compute everywhere
- Machine Learning dramatically increases compute demands
- As much as possible, Machine Learning workloads should run locally on device, not on remote servers
- Machine Learning is driving demand for advanced ARM processors and accelerator IP
- Machine Learning is having a significant impact on ARM's roadmap for future processors and architectures

ARM

Thank you for listening

ARM