# BRING INTELLIGENCE TO THE EDGE WITH INTEL® MOVIDIUS™ NEURAL COMPUTE STICK

Darren Crews
Principal Engineer, Lead System Architect, Intel NTG

# LEGAL DISCLAIMER

# Agenda

- Motiviation to move intelligence to the edge
- Edge compute use cases
- Barriers to moving intelligence to the edge
    - Deep learning algorithms – can they run on an edge device?
- Movidius Neural Compute Stick (arch,usage, etc)
- Code walkthrough and demo

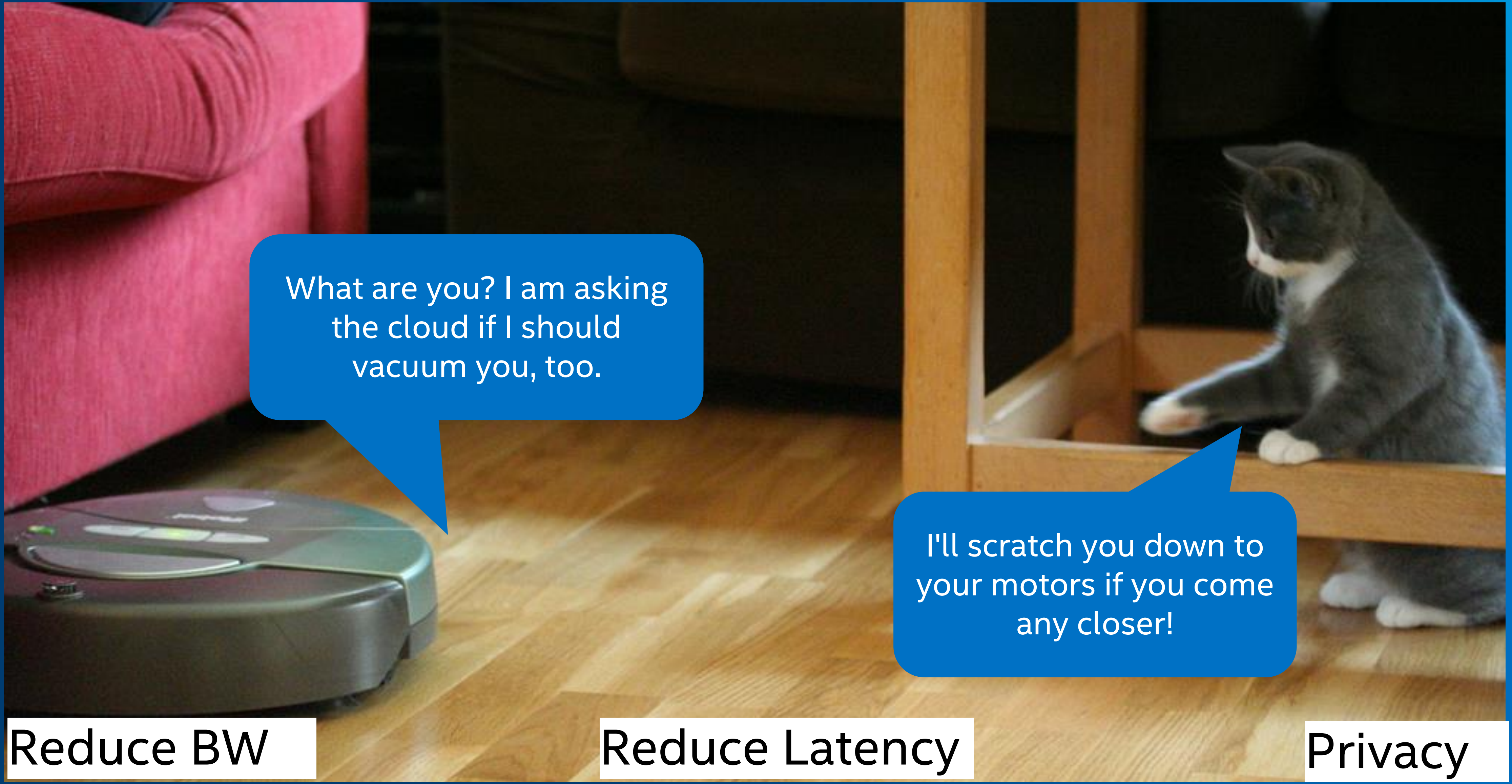
# Let's look at a larger scale…

20 billion connected devices by 2020[1]

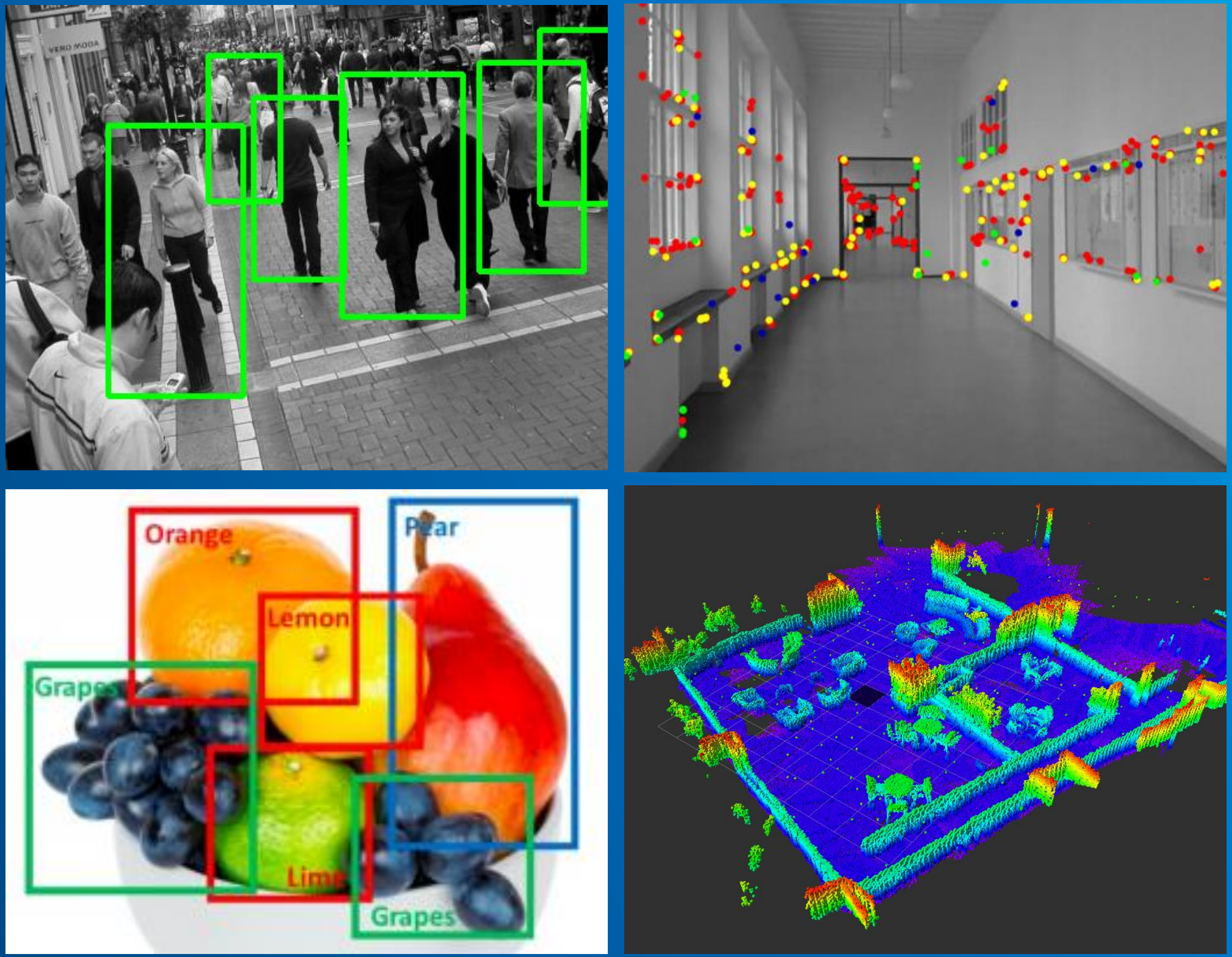
generating **billions of petabytes of data** traffic between devices & the cloud

1 Source: http://www.gartner.com/newsroom/id/3598917

# Why move intelligence from the cloud to the edge?



# Movidius
an Intel company


Computer vision and AI at the edge

# But what about Developers?

Complete HW + SW solution for developing Deep Learning application on the edge


+

**NC SDK**

Free download @ developer.movidius.com

# Use Case with the Neural Compute Stick

Little Ripper Lifesaver* UAV





https://newsroom.intel.com/editorials/data-most-important-force-society-today/
https://www.youtube.com/watch?v=QJOMfDyhUyo

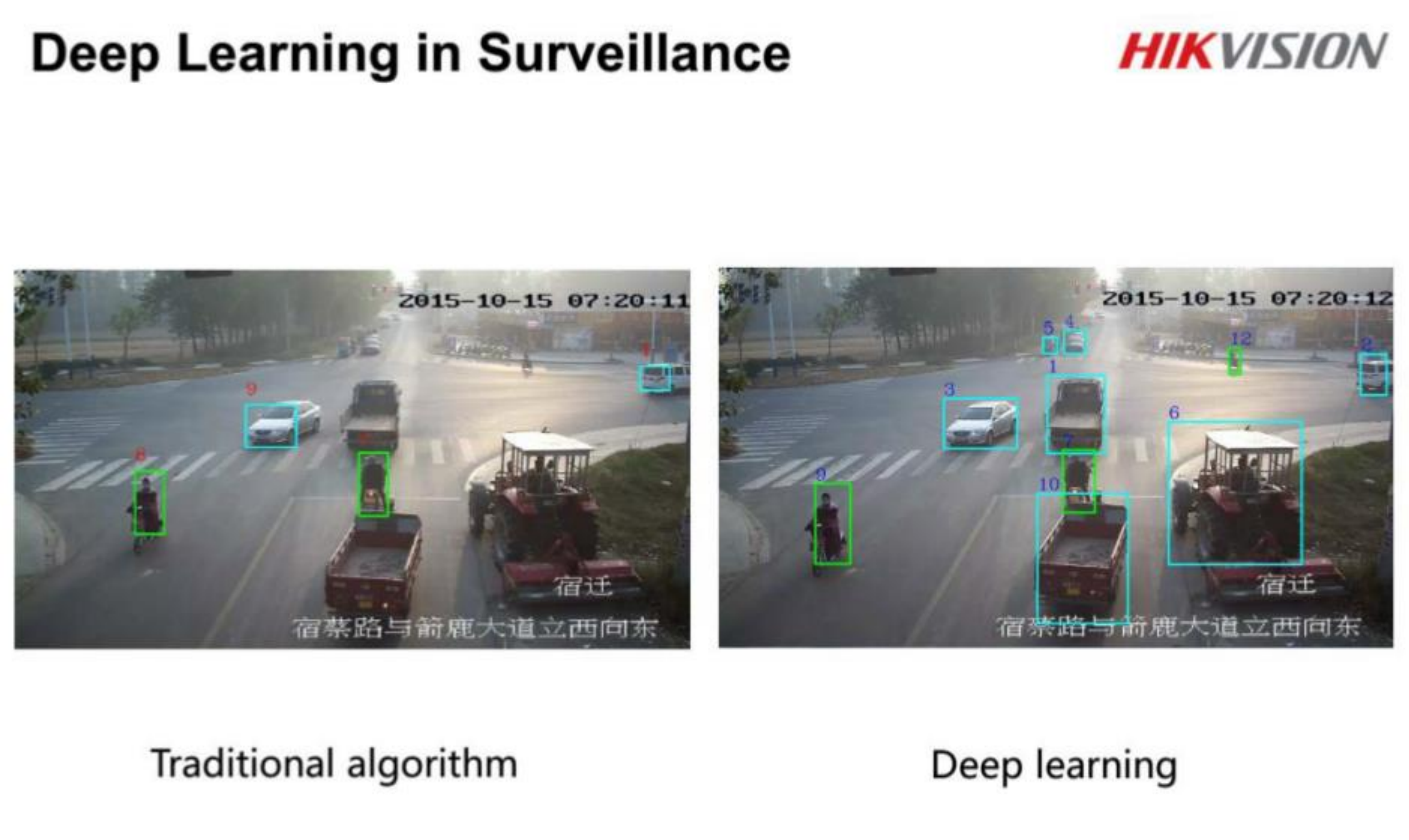
# Key Capabilities

Typical use cases could be:

- Robot
- Security camera
- Smart-home assistant

Key capabilities:

- Object detection
- Object  classification
- Facial recognition
- Natural language processing

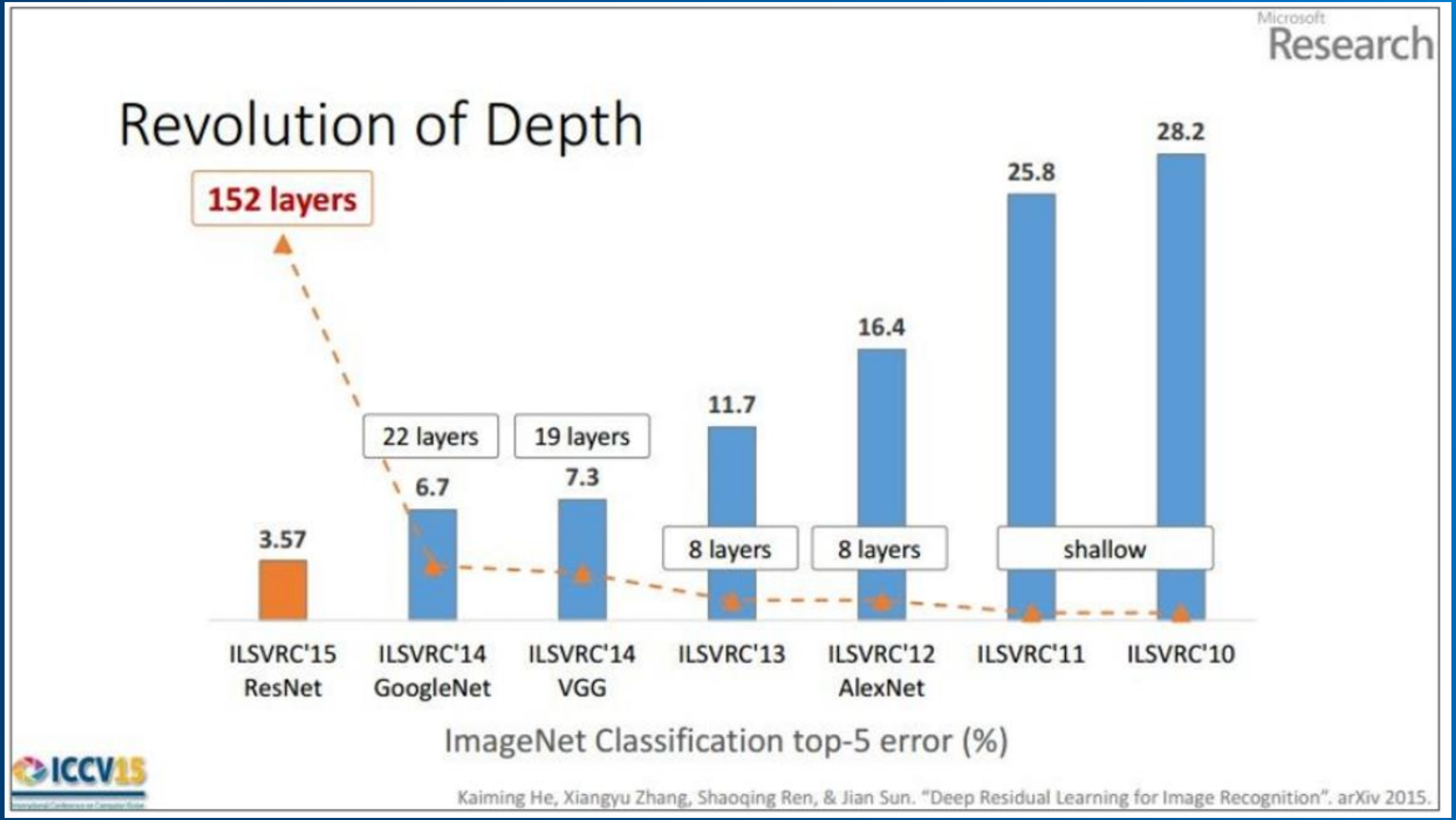


# Barriers to Moving Compute to the Edge

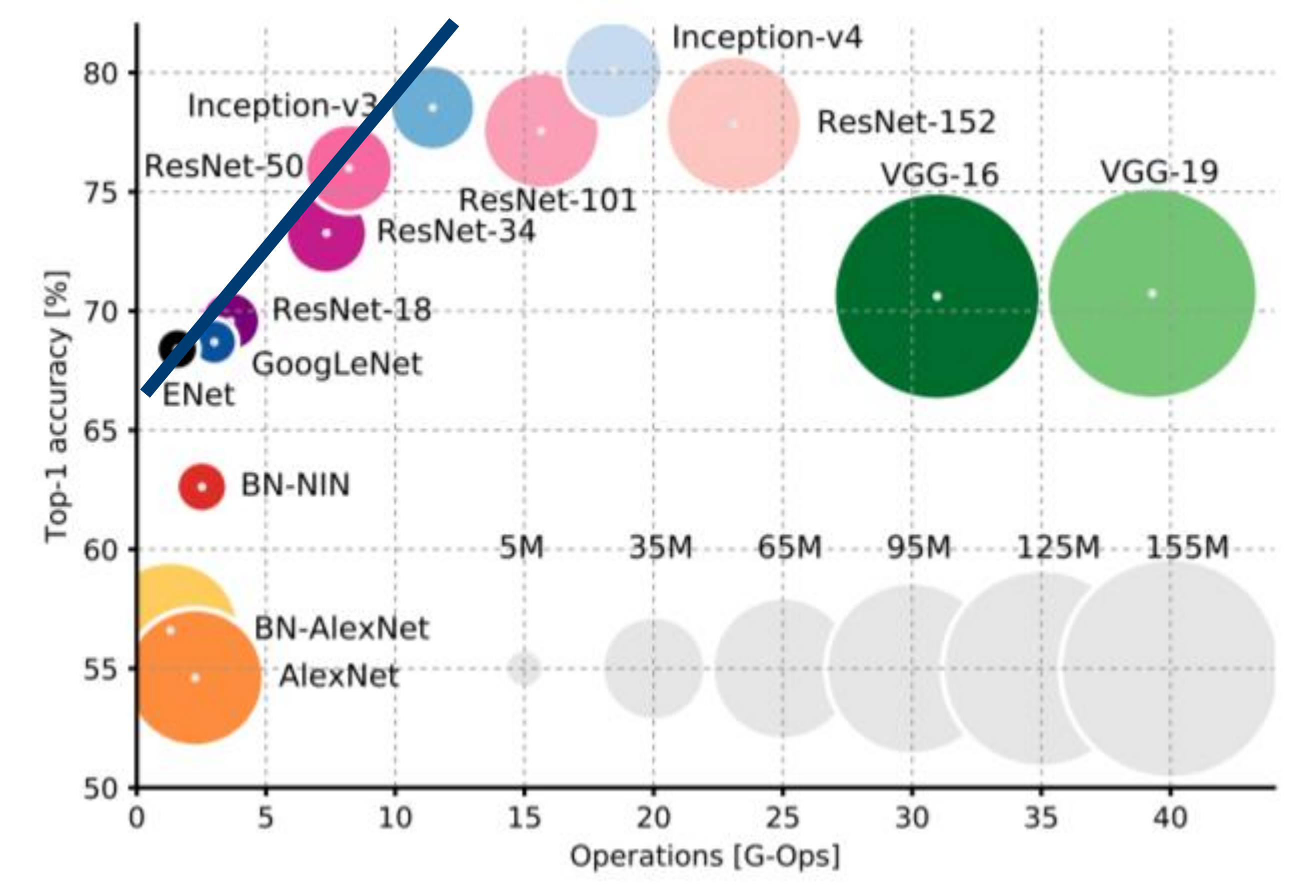
Move compute from the cloud to the edge:

- Accuracy
- Available compute
- Model efficiency
- Model size

# Image Classification: Getting more accurate and every year



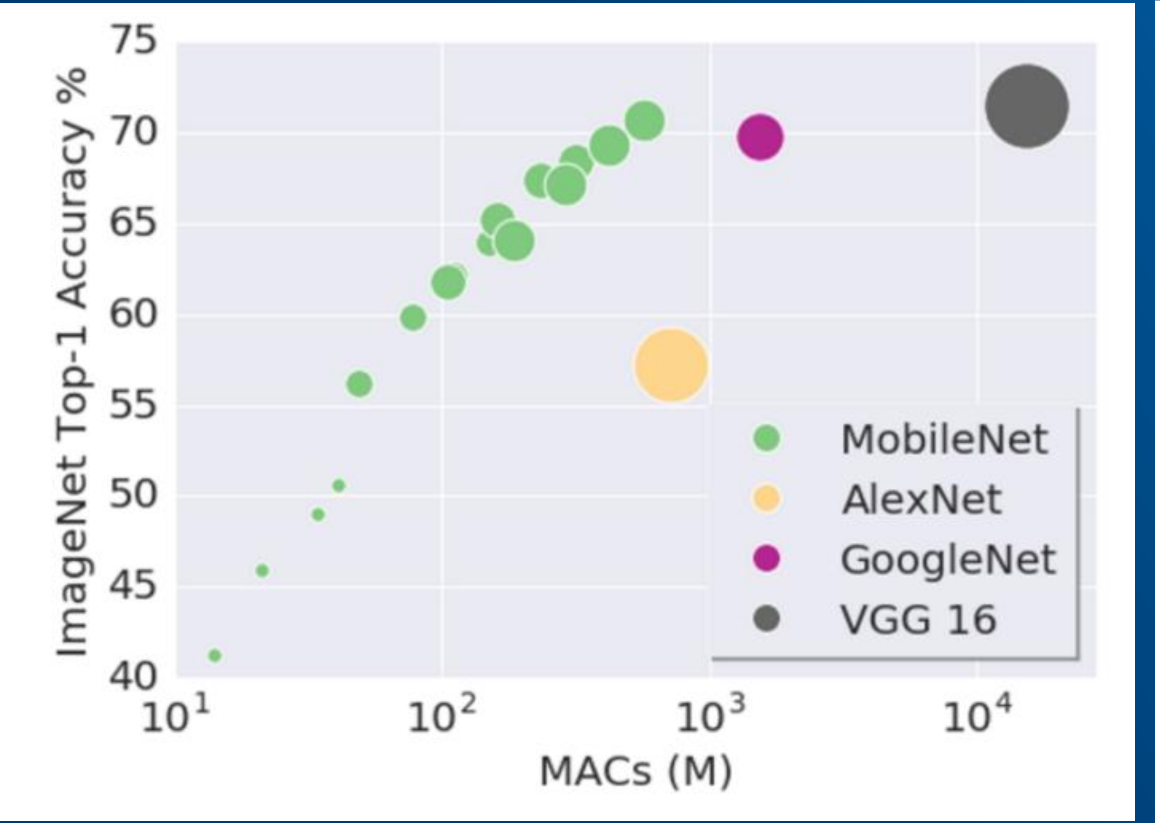
# Efficiency Key for Edge Devices

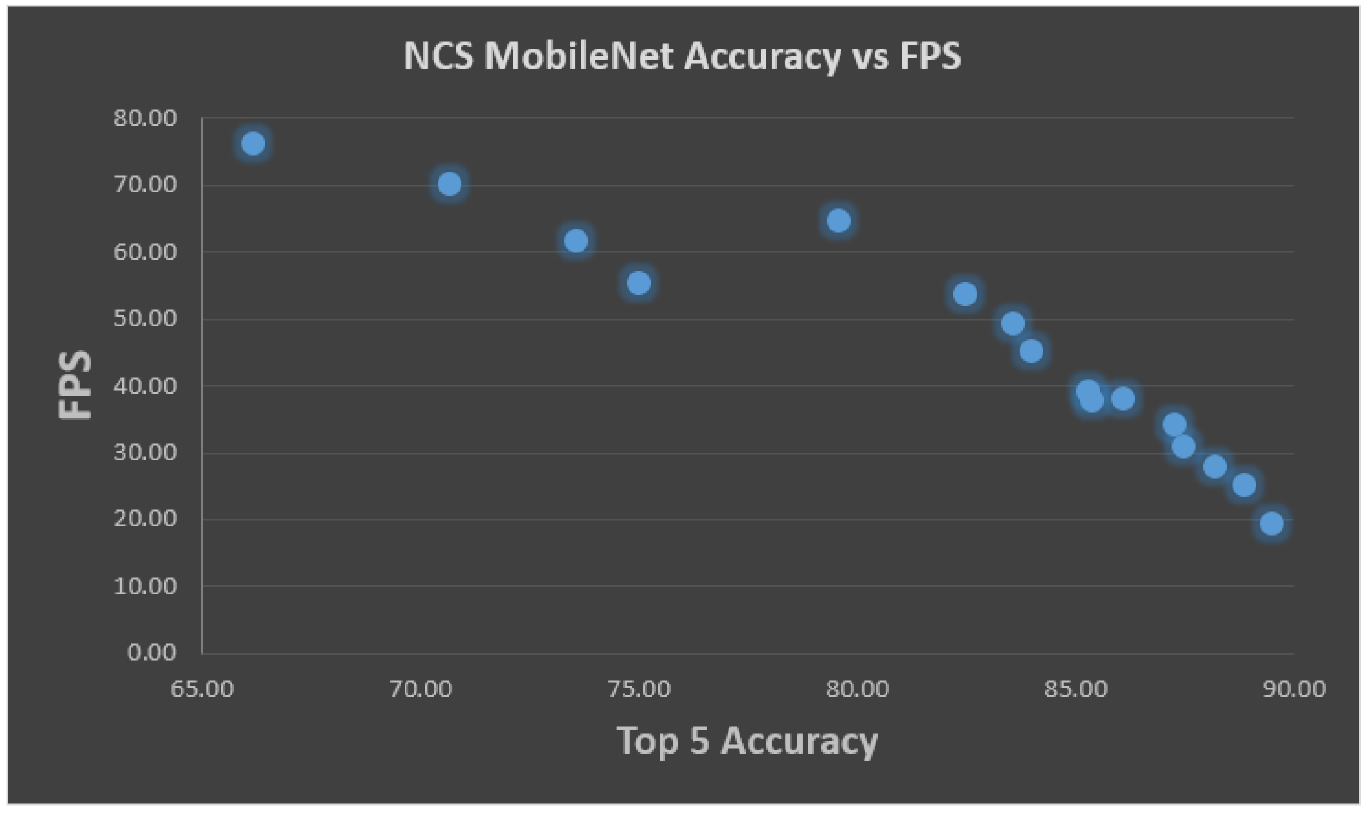



Alfredo Canziani, Eugenio Culurciello, Adam Paszke “AN ANALYSIS OF DEEP NEURAL NETWORK MODELS FOR PRACTICAL APPLICATIONS”

# MobileNet performance on NCS

Real time image classification on NCS

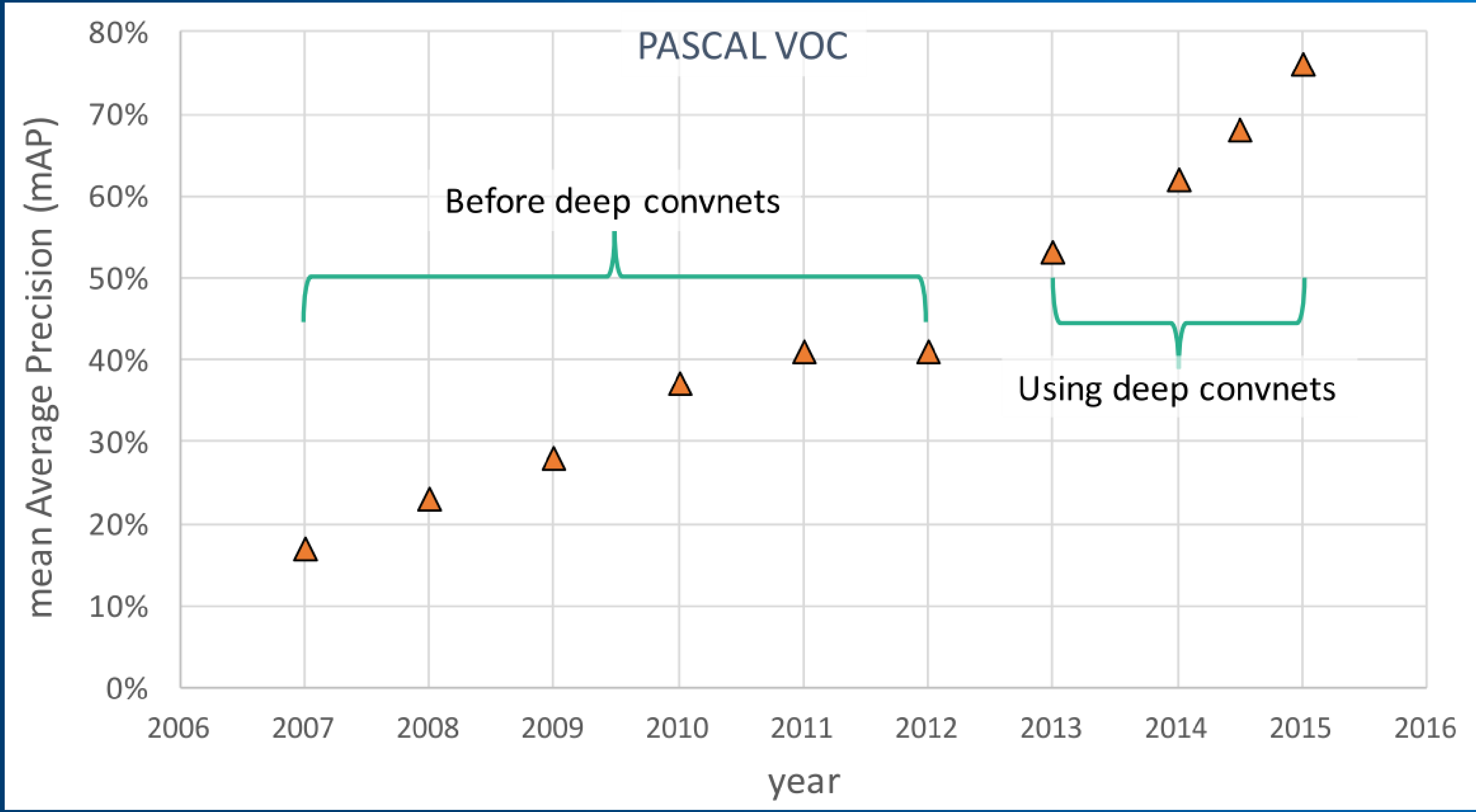



https://github.com/tensorflow/models/blob/master/research/slim/nets/mobilenet_v1.md

# Object Detection Benchmarks

Deep neural networks win in object detection in 2013.




**Ross Girshick,** IEEE International Conference on Computer Vision (ICCV), 2015

# Object Detectors – Speed matters!

Significant improvements in Object Detectors in the last two years enable Object Detection on edge devices like Intel Movidius NCS.

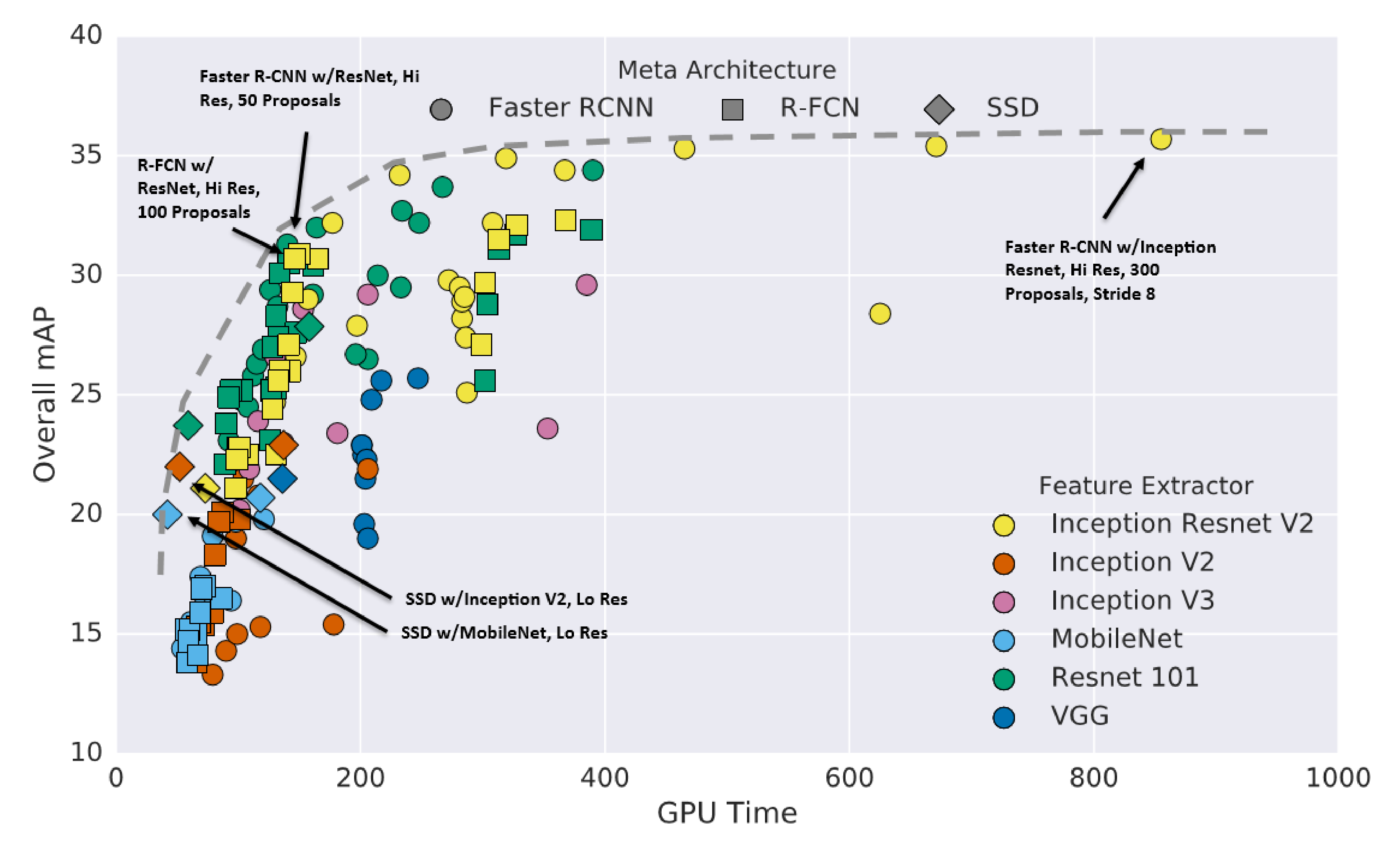
| | Pascal 2007 MaP | Speed |
|---|---|---|
| R-CNN | 66.0 | .05FPS |
| Fast R-CNN | 70.0 | 0.5FPS |
| Faster R-CNN | 73.2 | 7FPS |
| YOLO | 69.0 | 45FPS |
| SSD-Mobilenet | 72.7 | 11FPS |

GPU

NCS

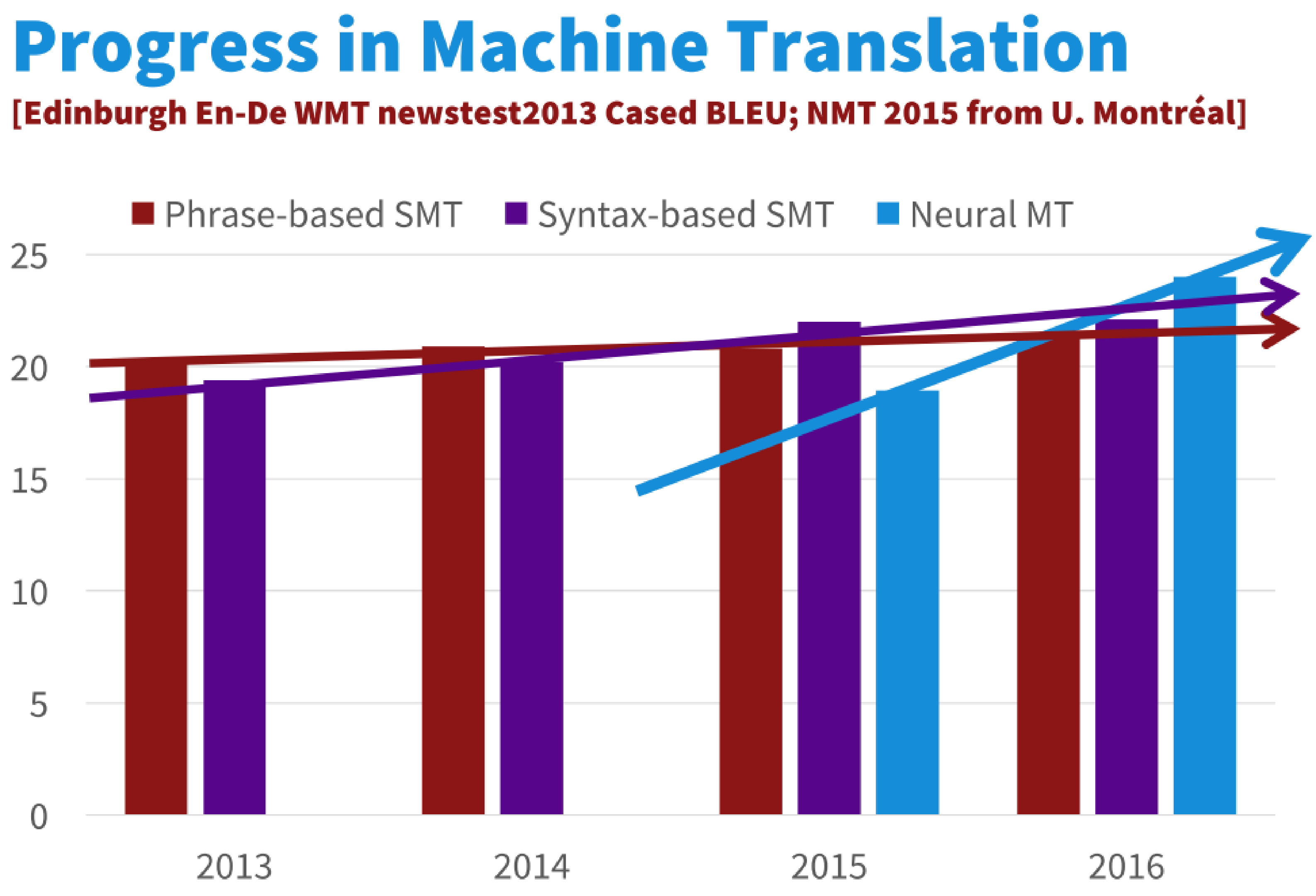
Jonathan Huang, et al, Speed/accuracy trade-offs for modern convolutional object detectors

# Object Detectors – lots of choices, choose wisely




Jonathan Huang, et al, Speed/accuracy trade-offs for modern convolutional object detectors

# Machine Translation



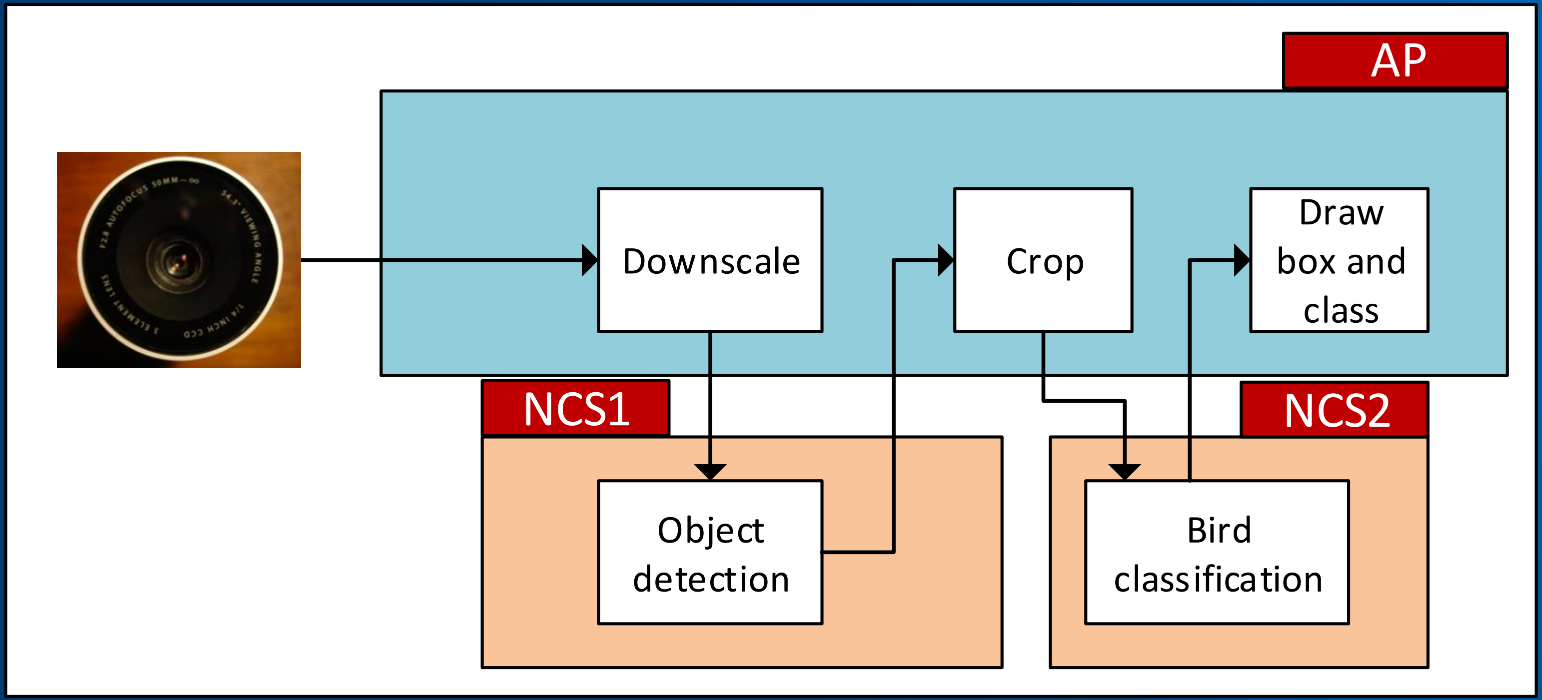
# Intel® Movidius™ Neural Compute Stick

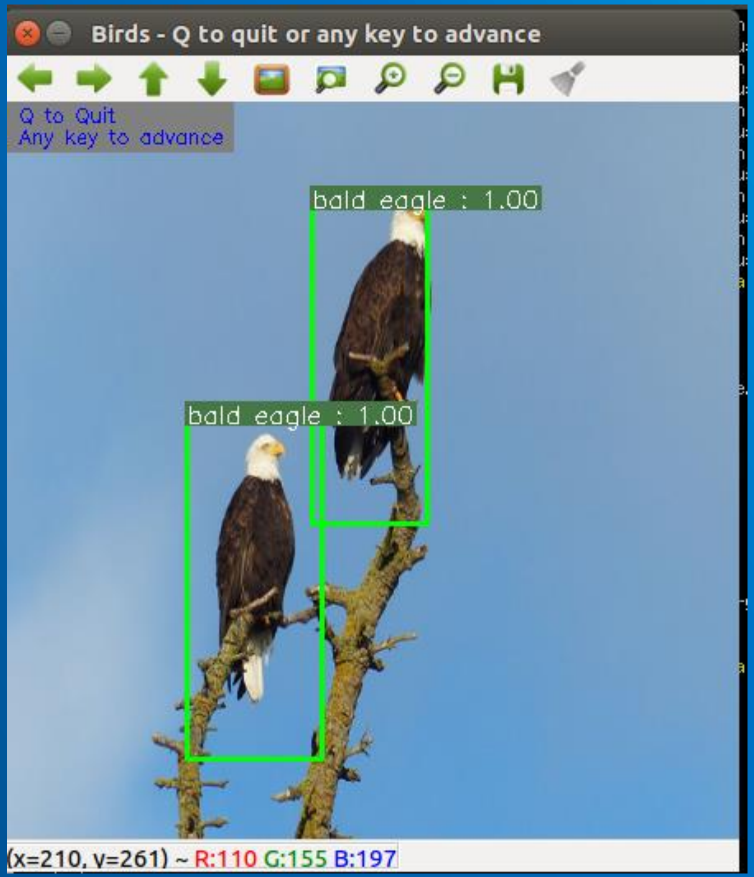
## Redefining the AI developer kit


- Neural network accelerator in USB stick form factor
- TensorFlow™ and Caffe frameworks supported, along with many popular networks
- Source is available for the SDK, which allows you to compile for other platforms
- Features the same Intel Movidius vision processing unit (Intel Movidius VPU) used in drones, surveillance cameras, VR headsets, and other low-power intelligent and autonomous products

# Demo: Scaling inference performance with multiple sticks

# Object Detection and Classification



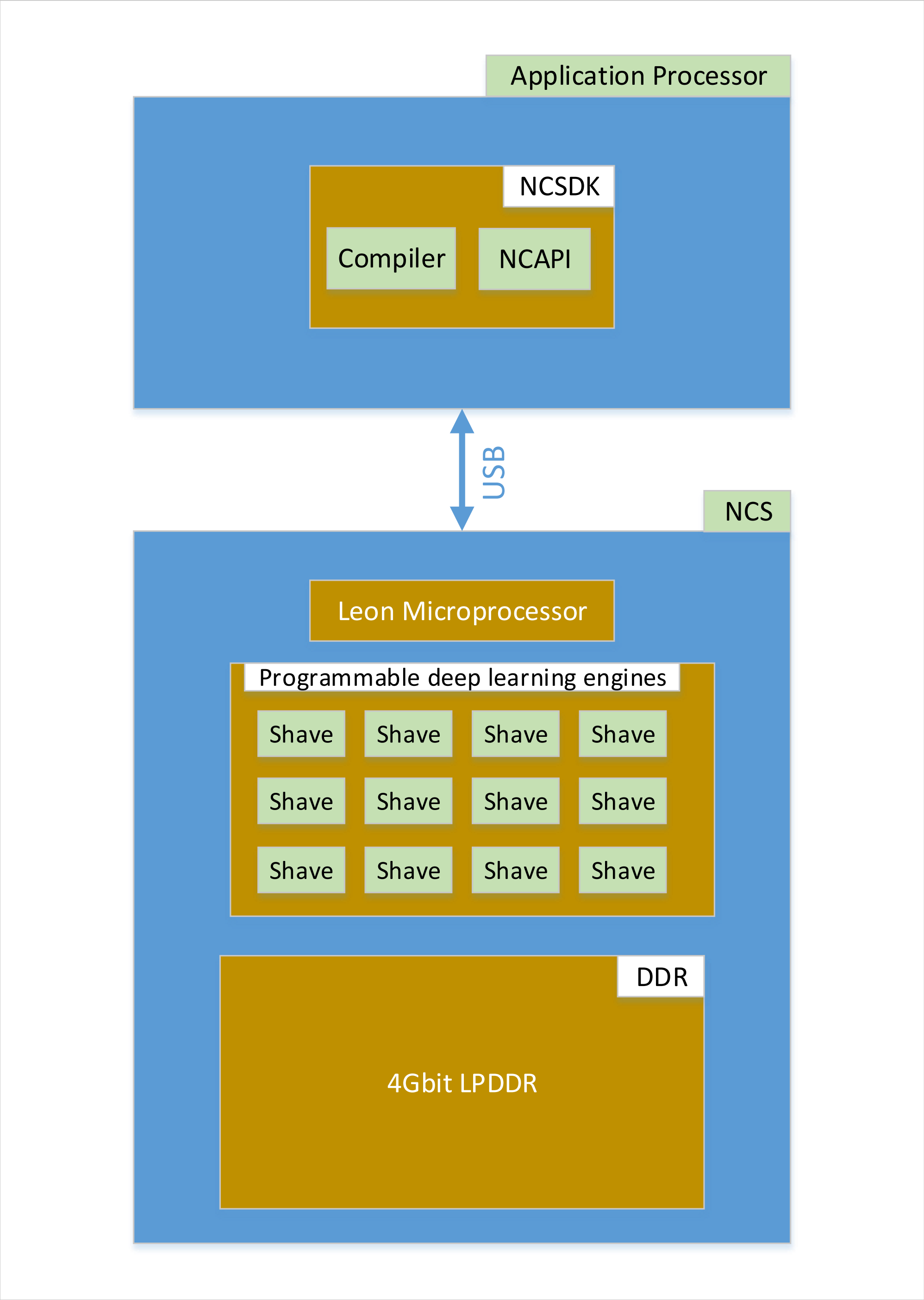

https://github.com/movidius/ncappzoo/tree/master/apps/birds

# Architecture

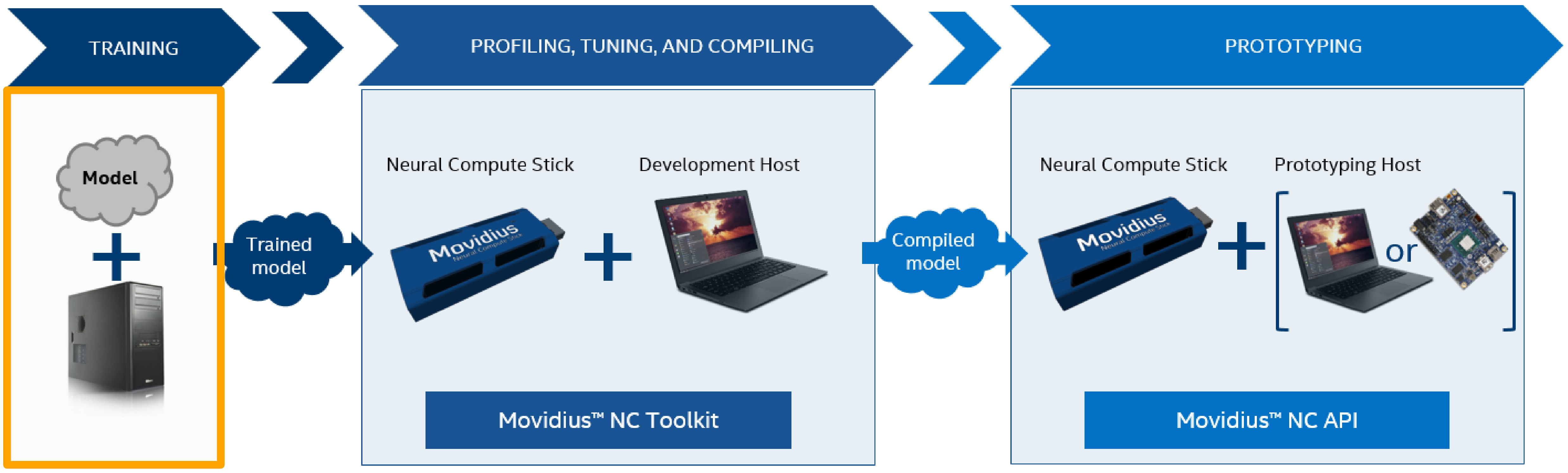
Intel Movidius  NCS contains the Intel® Movidius™ Myriad™ 2 vision processing unit, including 4 Gbit of LPDDR.

Intel Movidius NCS is connected to an application processor (AP), such as a Raspberry Pi or UP Squared board.

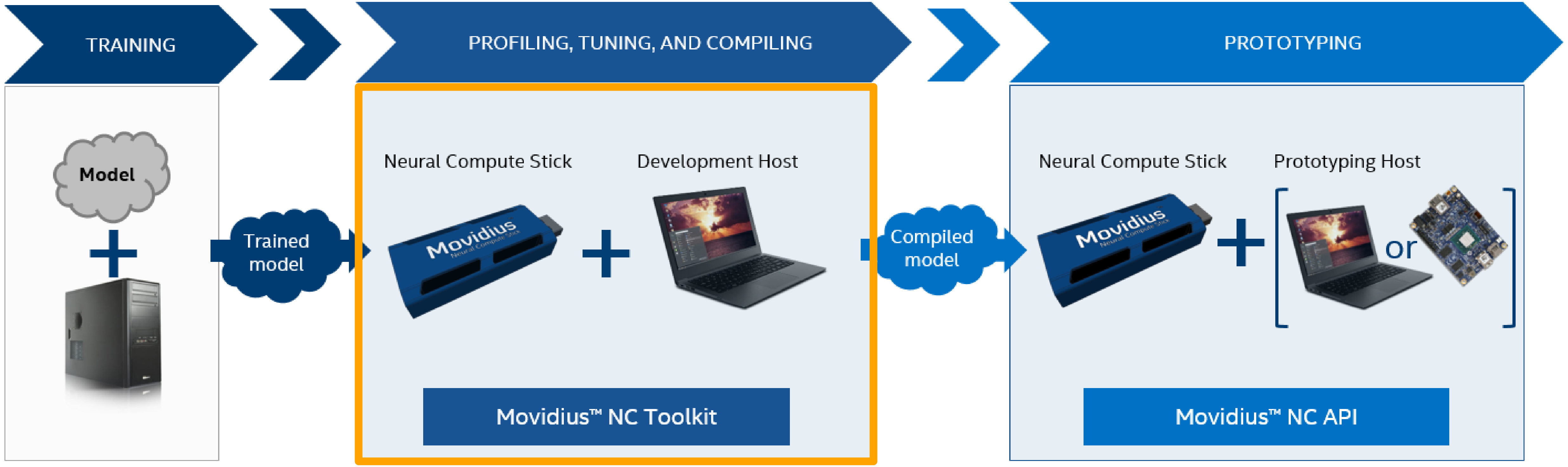
Execution is controlled by the LEON microprocessor, and the calculations are done on the SHAVE processors.



# NC SDK Workflow

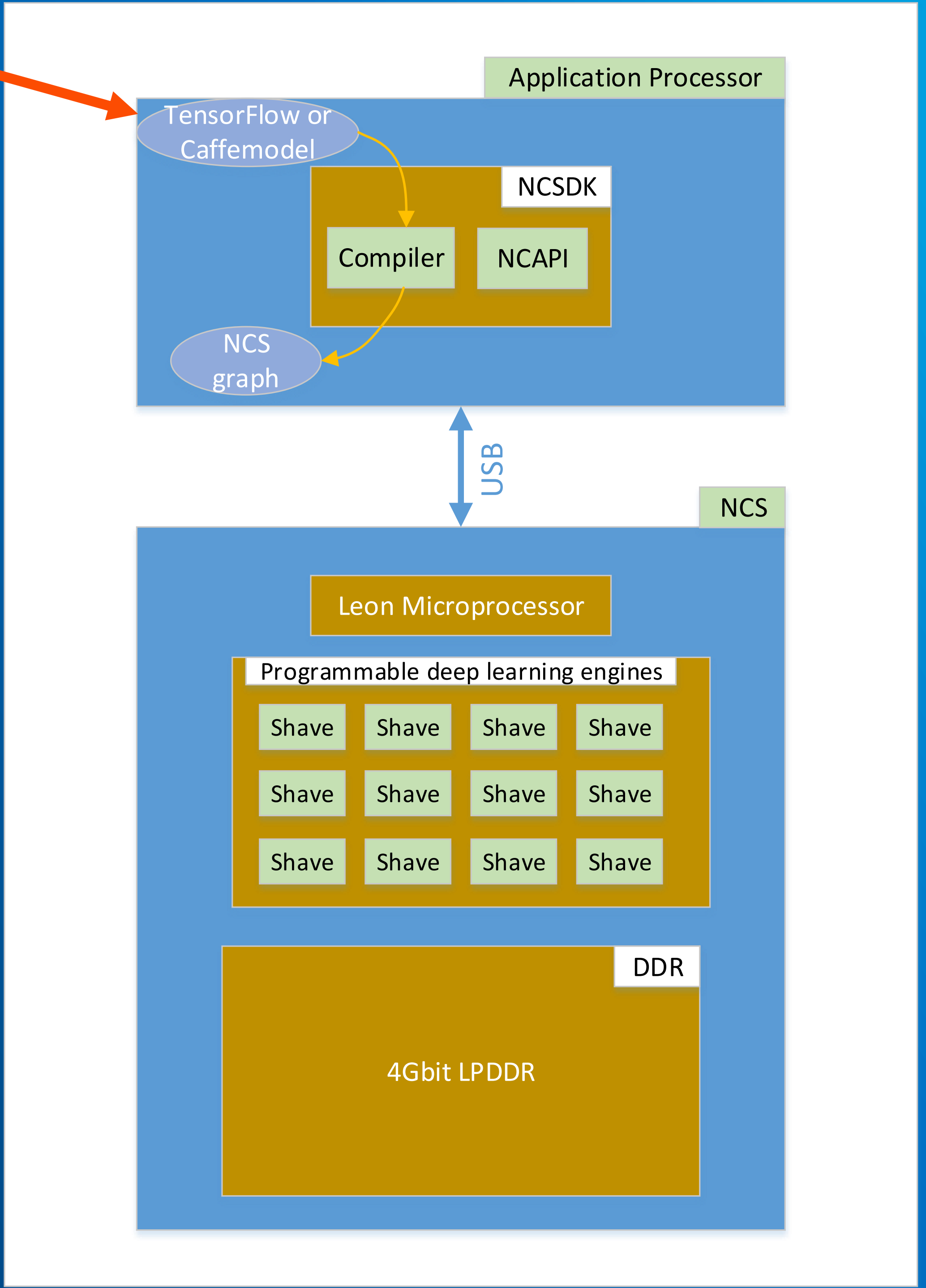


# NC SDK Workflow

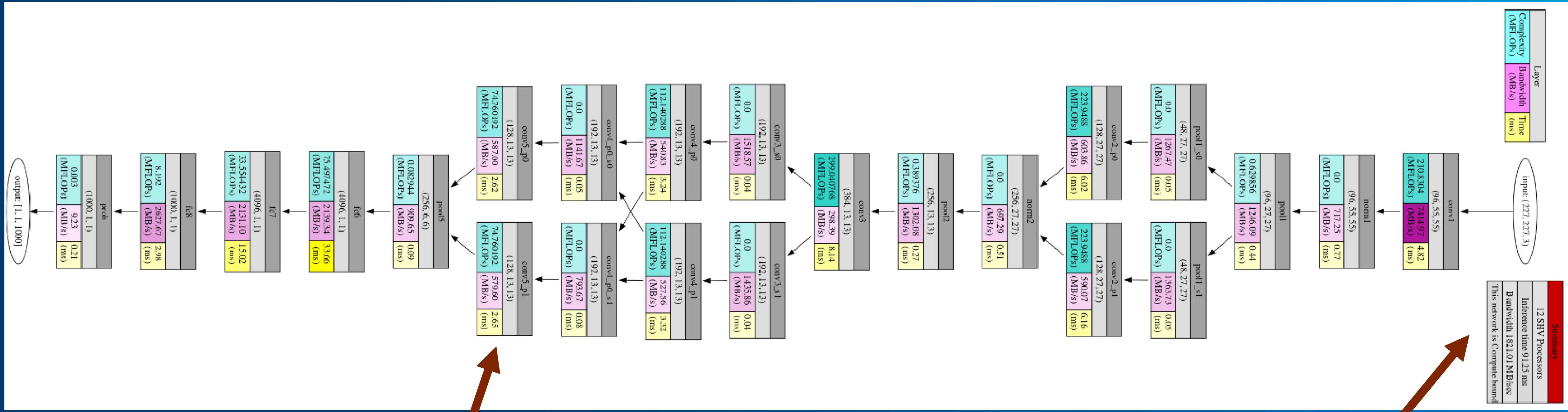


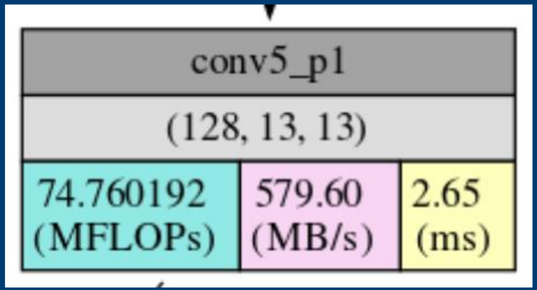
# Development Flow

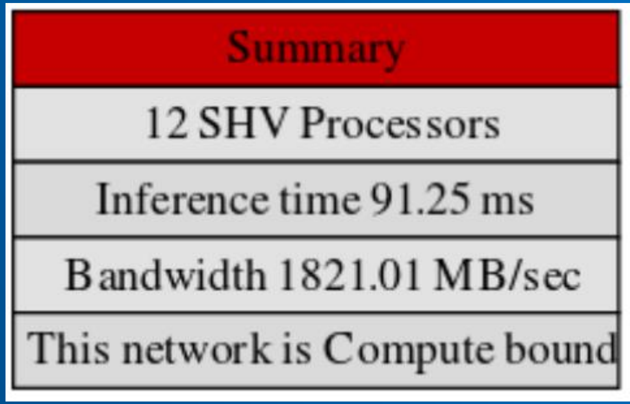
## Step 1: Convert the model



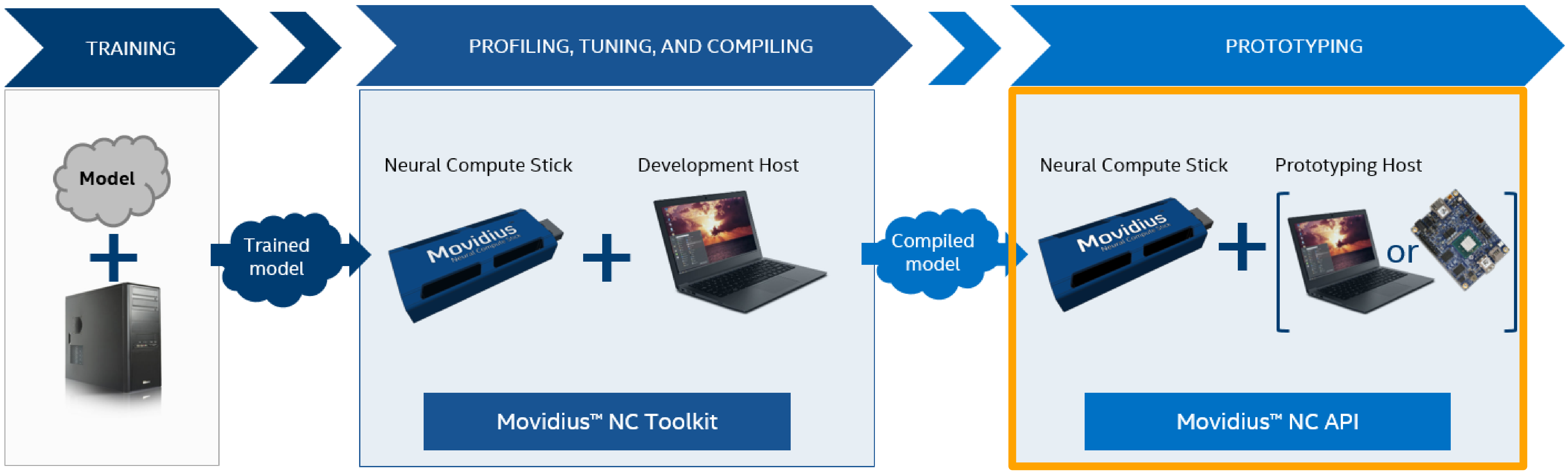
# Trained model can be profiled on NCS for performance:





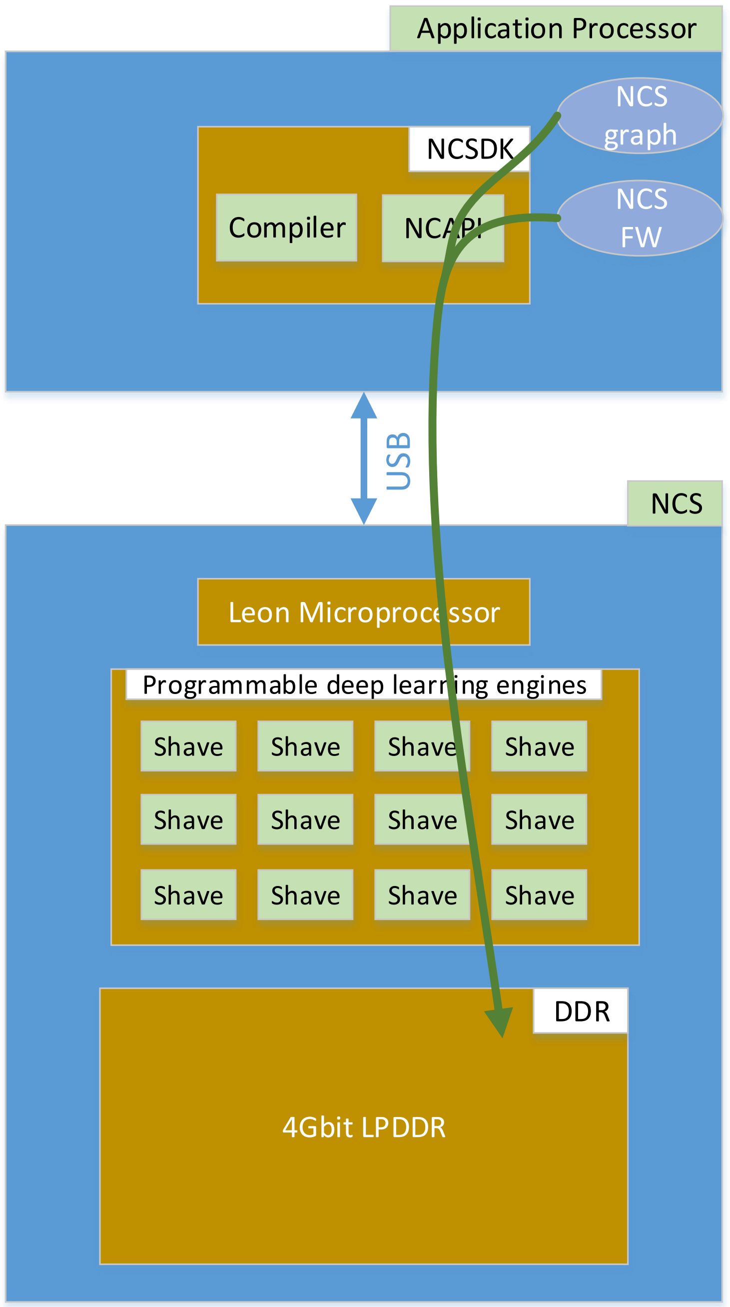



# NC SDK Workflow



# Development Flow

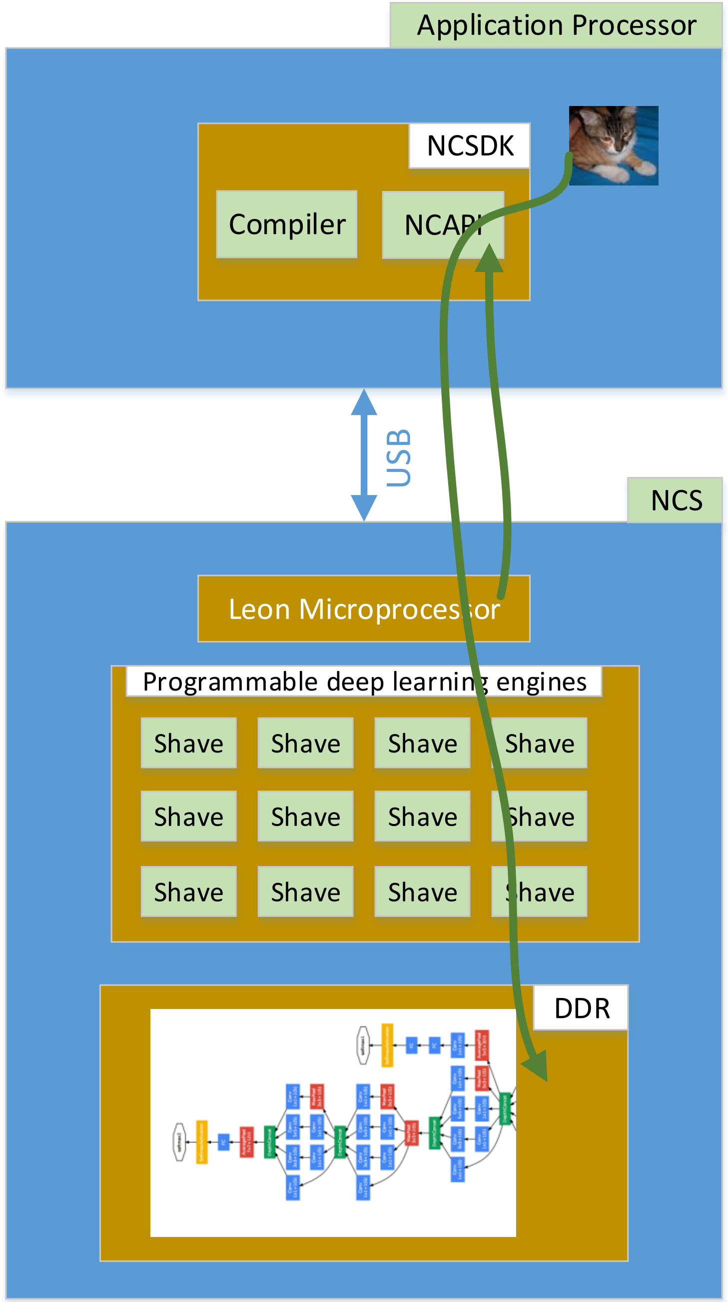
## Step 2: Load the model and the FW



# Development Flow

Step 3: Perform inference

- Load the image
- Run the model
- Return the results



# Demo

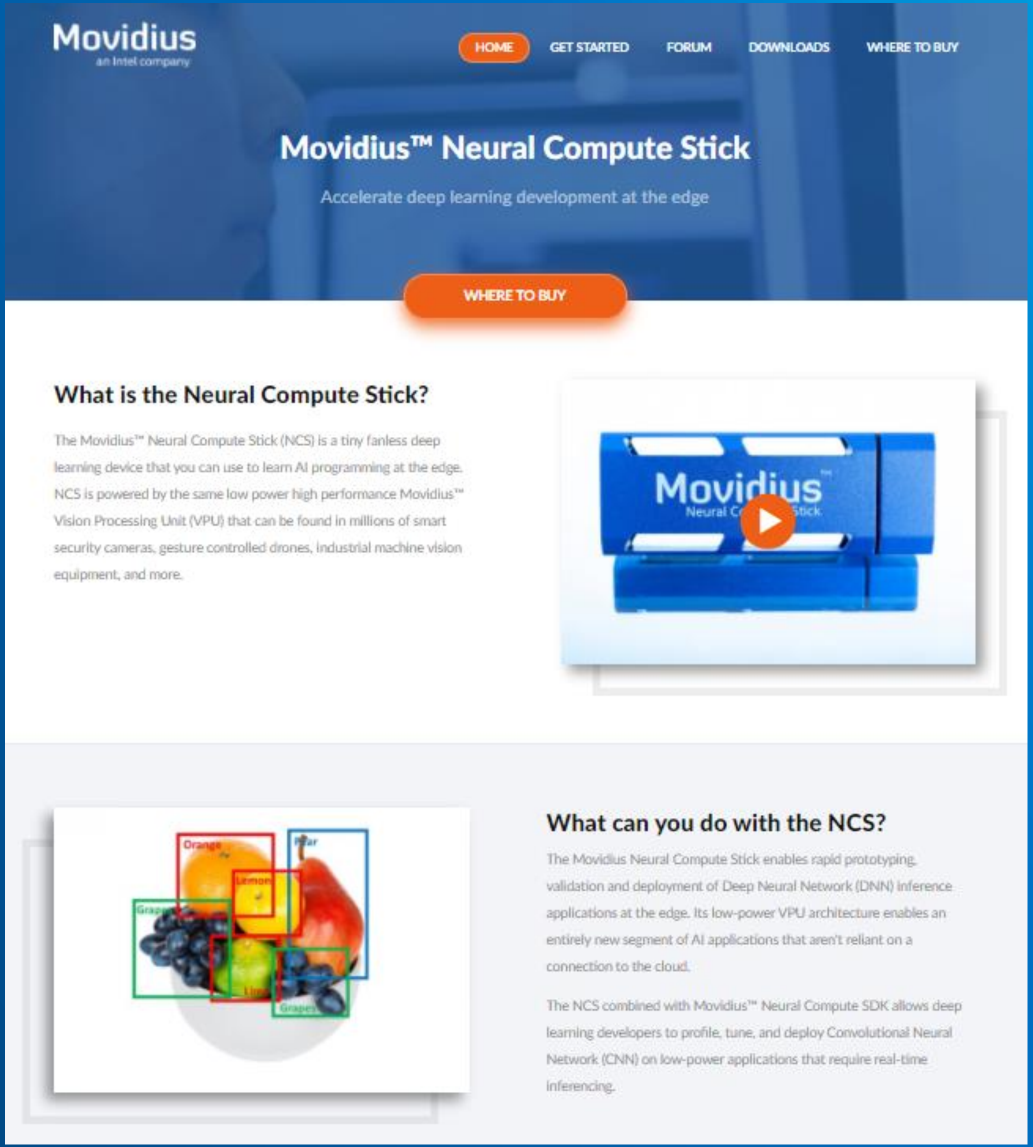
# Let's Review

- Moving AI to the edge is important for a number of reasons (lots of data, latency, and privacy)
- Deep Learning has progressed in many areas where these use cases can be run on the edge (object detection, classification, etc.)
- Intel Movidius NCS is an easy-to-use prototyping vehicle for developing your edge devices

# Explore developer.movidius.com

Try out the following pages:

- Main page
- Getting started
- Downloads
- Docs
- Forums
- Where to buy



# Questions?

intel®

experience what's inside™