Introduction to Intel® Distribution of OpenVINO™ toolkit for Computer Vision Applications

100: Beginner-level
Lesson 04

Introduction to Intel® Distribution of **OpenVINO™** toolkit for Computer Vision Application

OpenVINO 100 **– Course agenda**

Lesson 1: Introduction, why do we need Artificial Intelligence (AI).

Lesson 2: What is Video, what is computer vision, how do we accelerate it on modern computers.

Lesson 3: How to accelerate Video processing

Lesson 4: How to accelerate Neural Network for vision applications

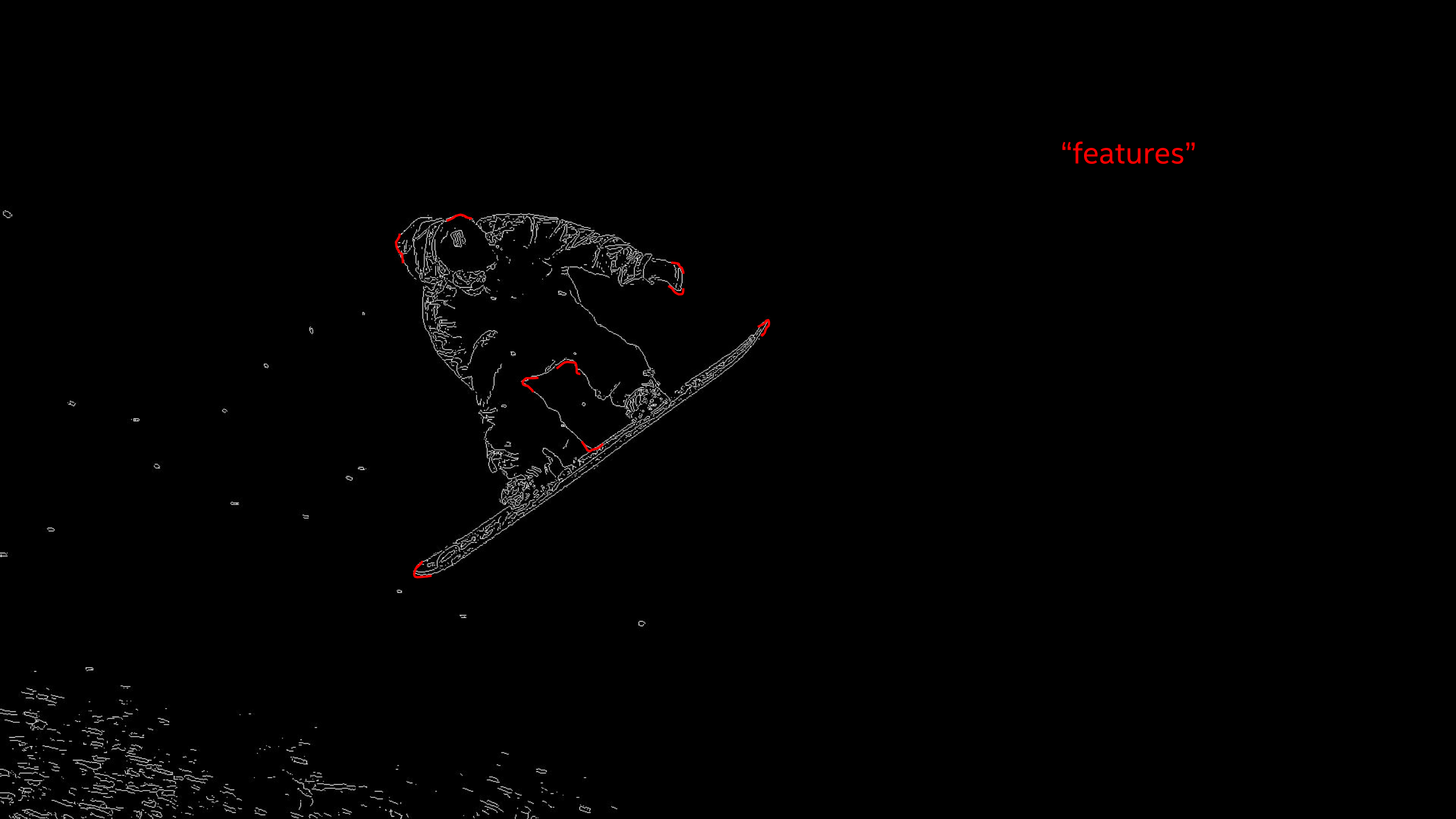
Lesson 5: Video Analytics pipeline

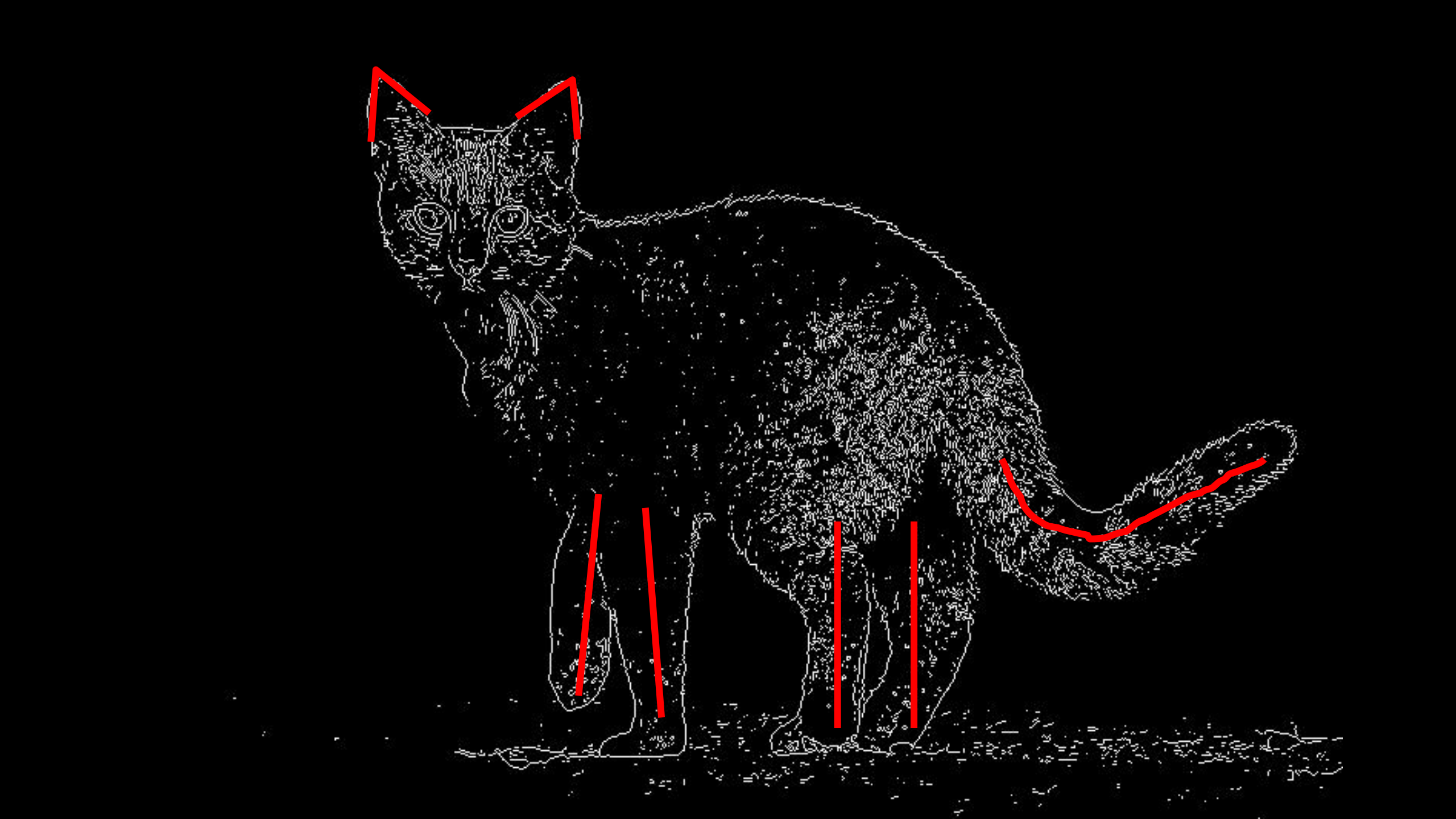
Lesson 6: Demos, OpenVINO at work

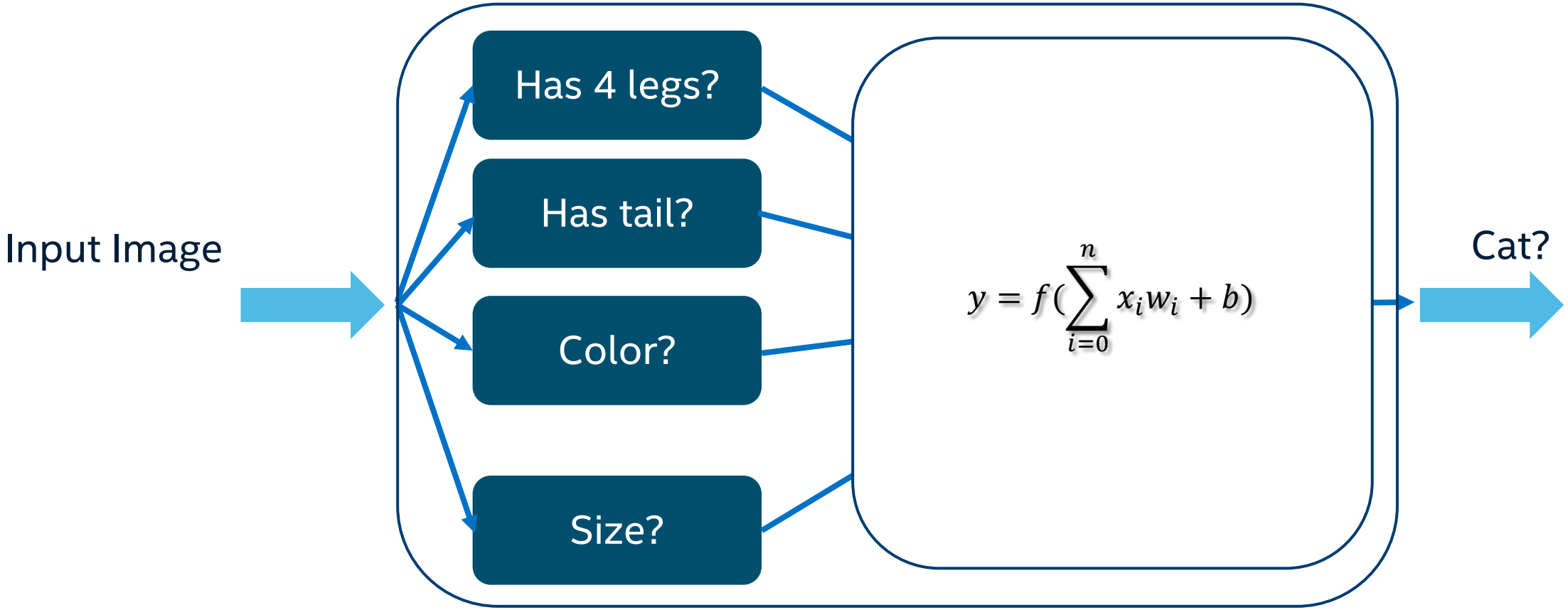
Lesson 7: The full flow, from Data to a product using Intel tools-Part 1.

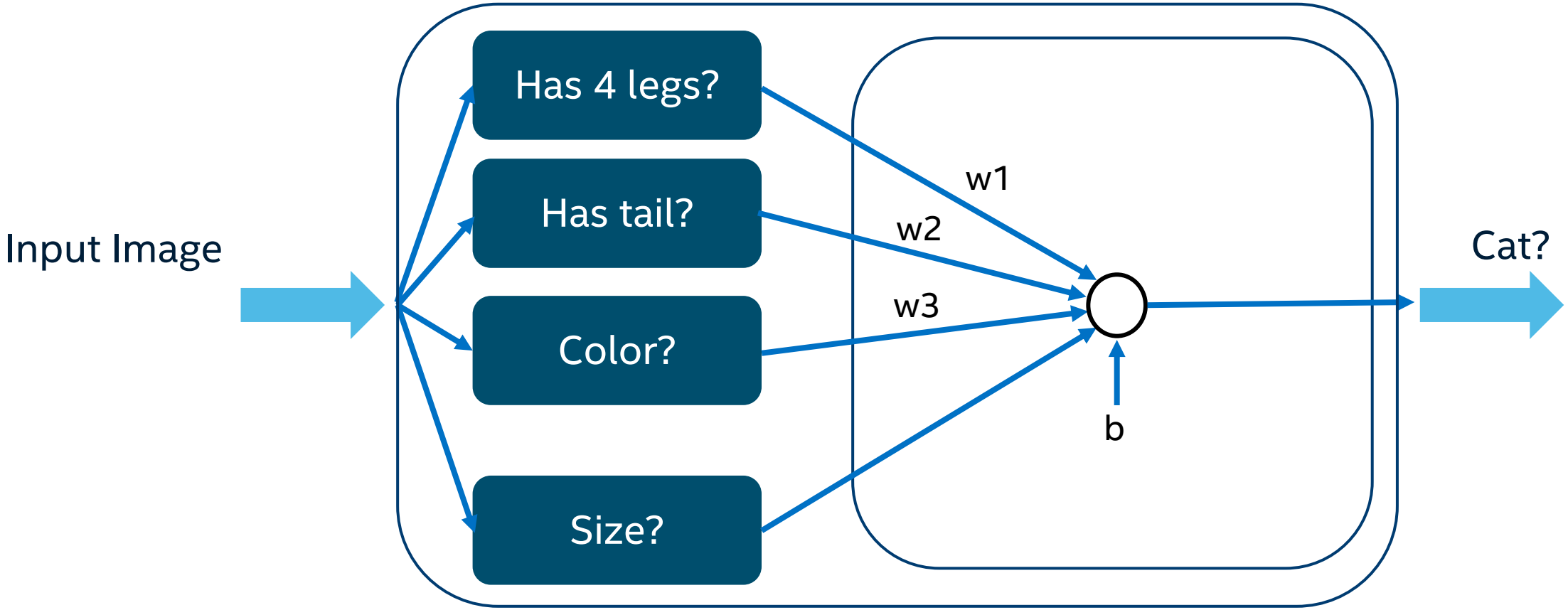
Lesson 8: The full flow, from Data to a product using Intel tools-Part 2.

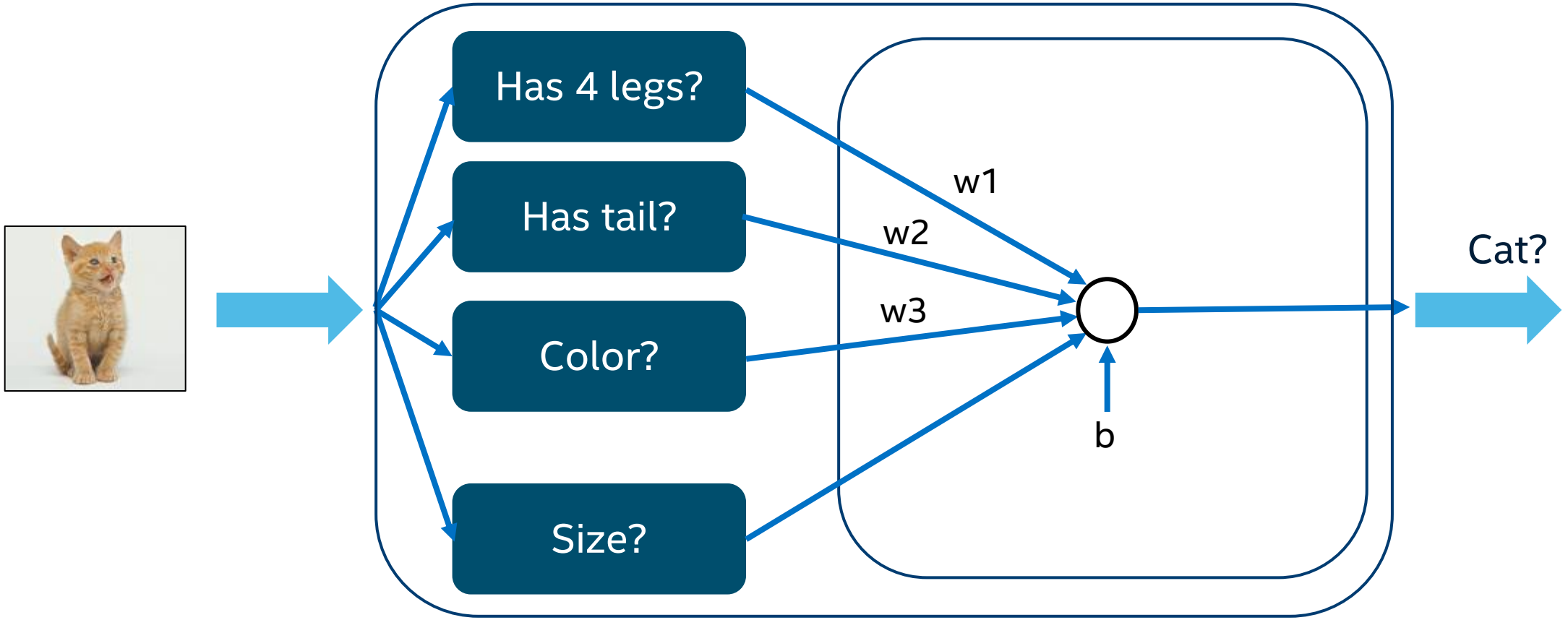
Lesson 9: Summary, intro to next course (200)

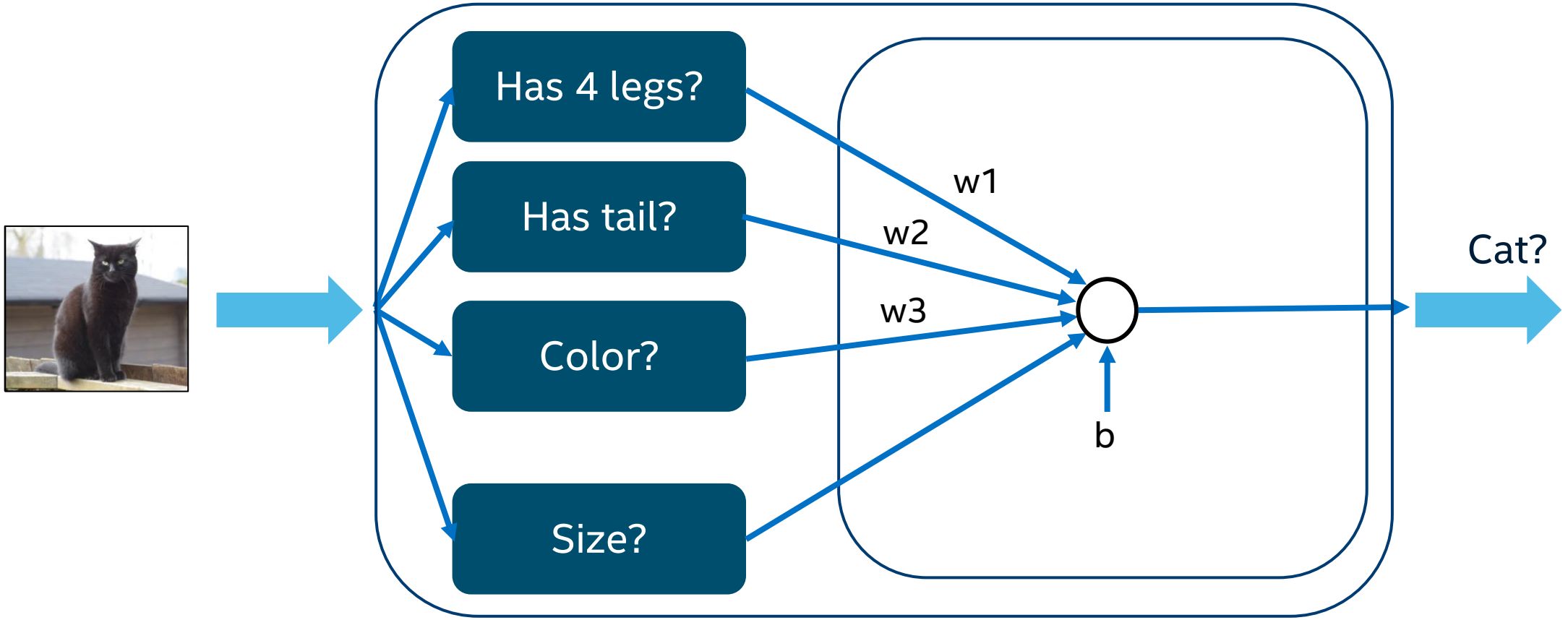

“features”

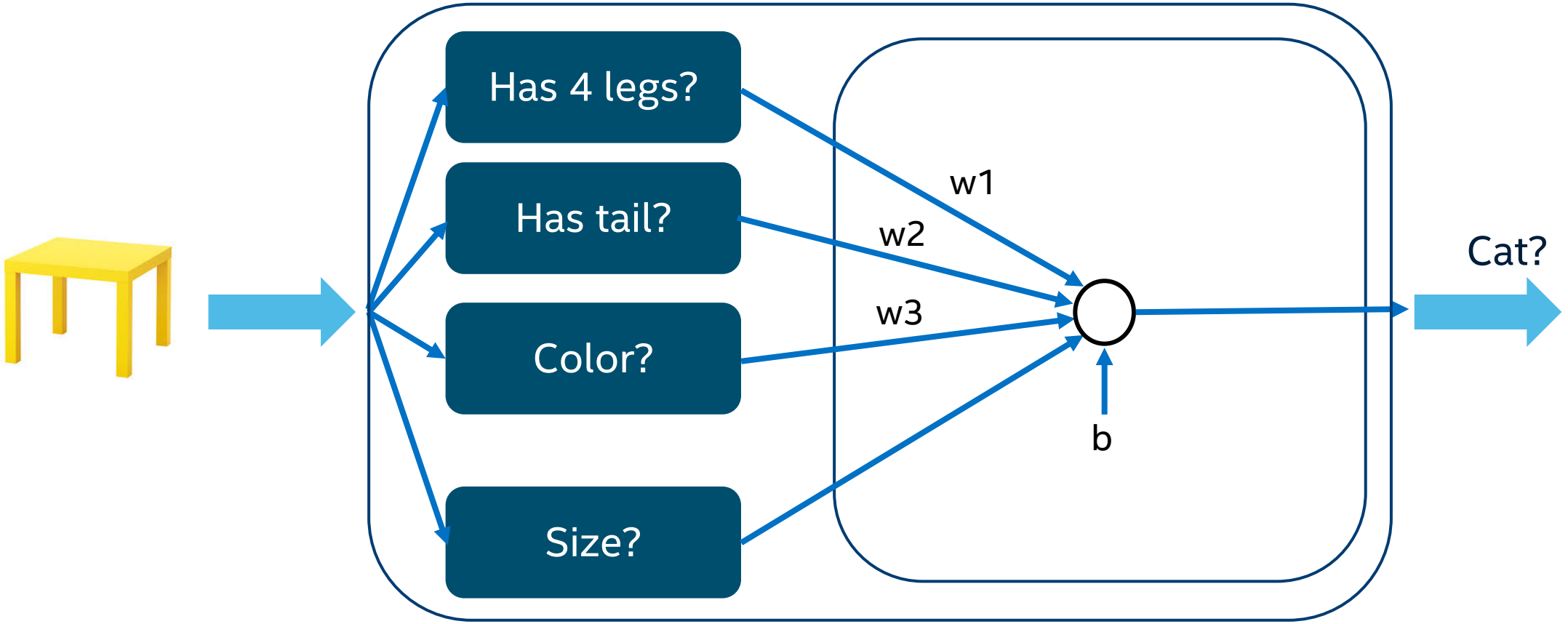


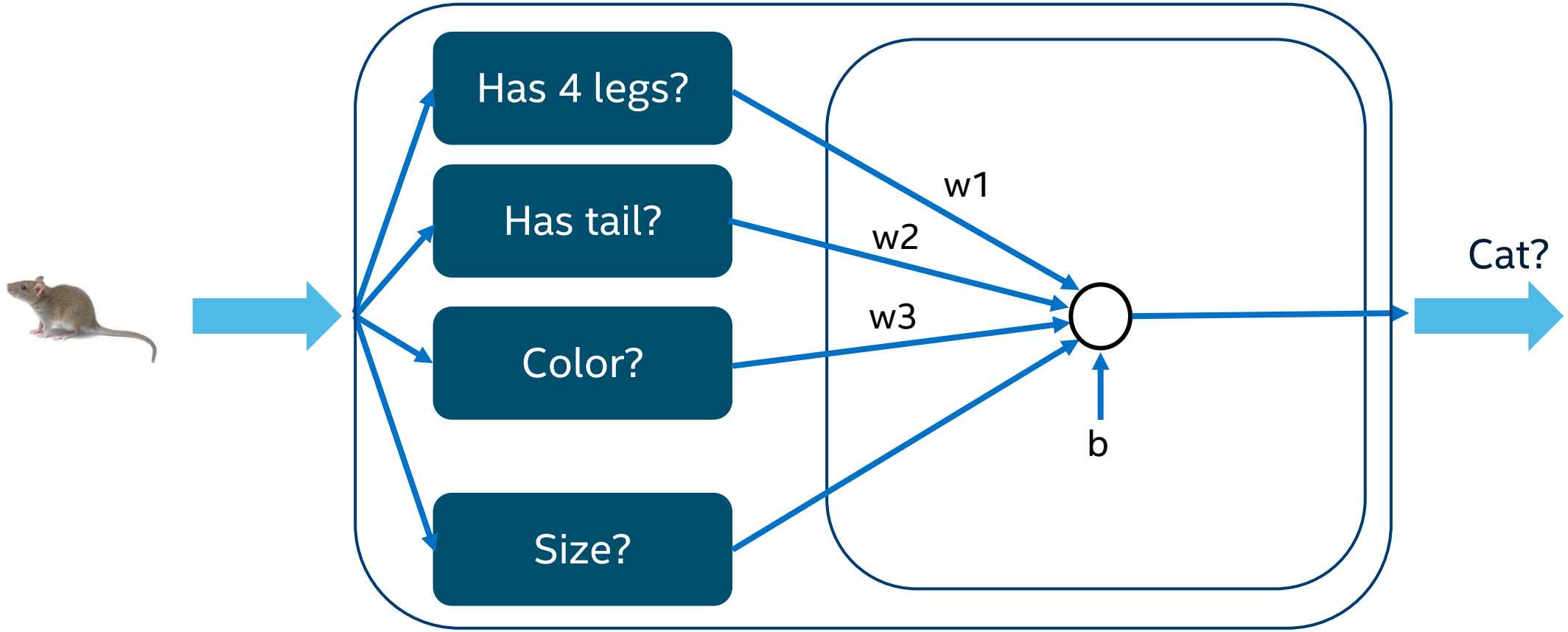

Input Image
Has 4 legs?
Has tail?
Color?
Size?
$y = f(\sum_{i=0}^{n} x_i w_i + b)$
Cat?


Input Image
Has 4 legs?
Has tail?
Color?
Size?
w1
w2
w3
b
Cat?

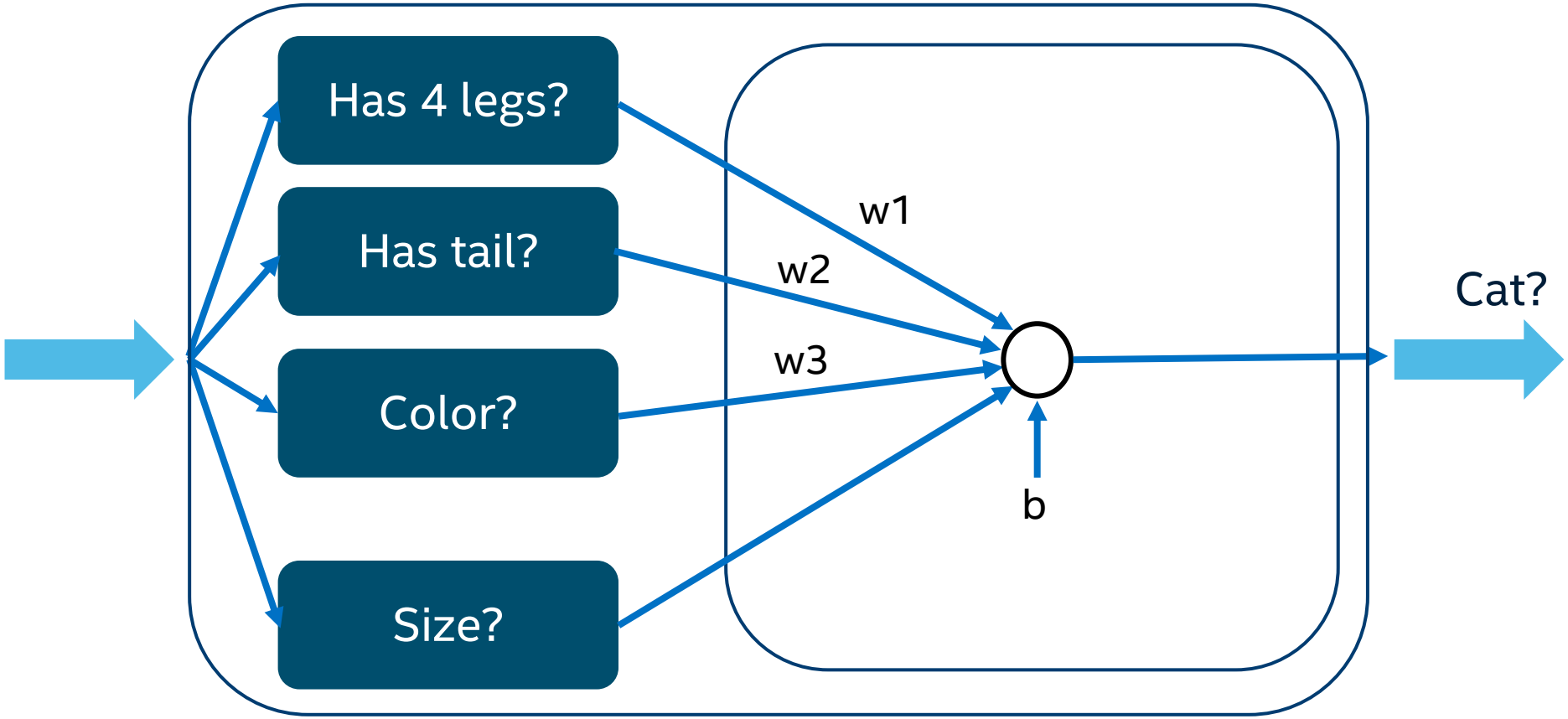

Has 4 legs?
Has tail?
Color?
Size?
w1
w2
w3
b
Cat?

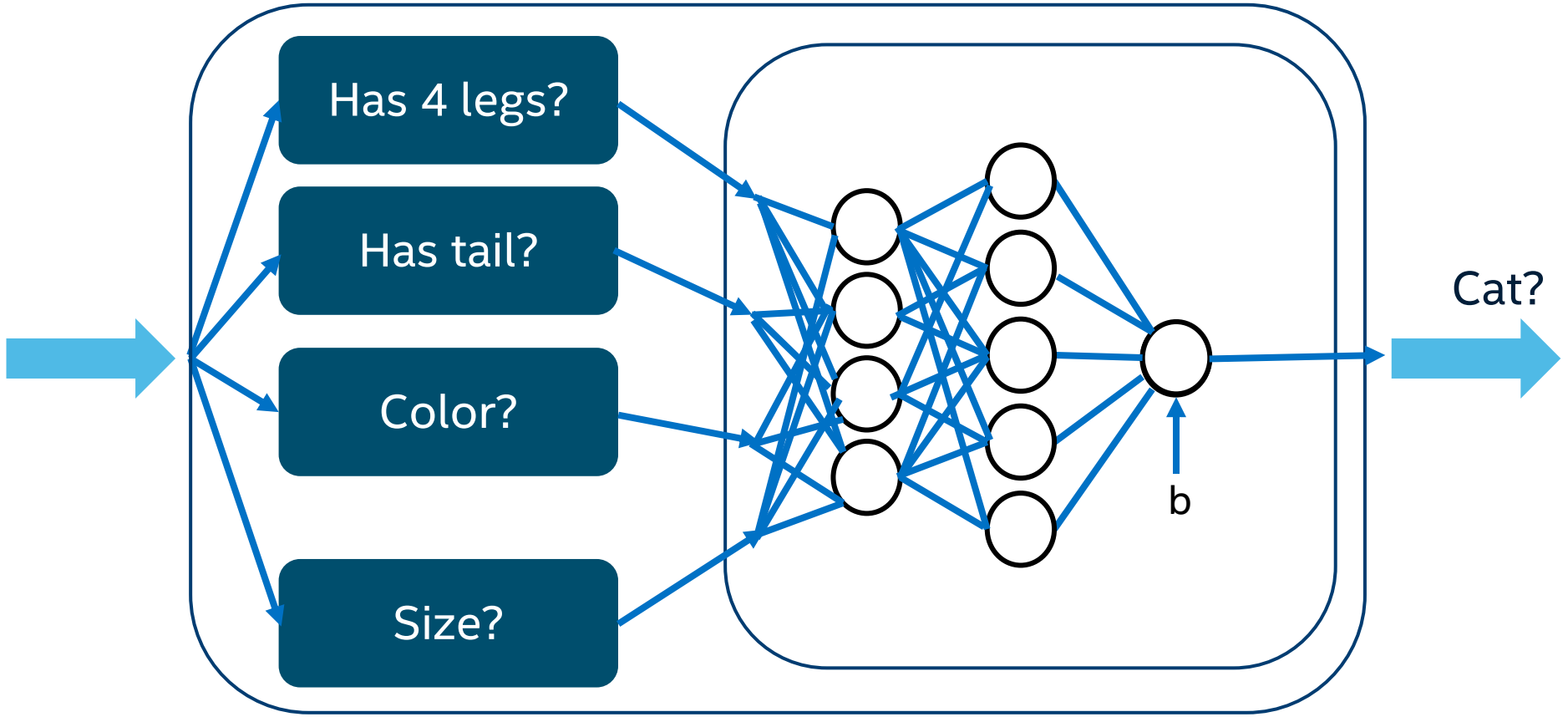

Has 4 legs?
Has tail?
Color?
Size?
w1
w2
w3
b
Cat?


Has 4 legs?
Has tail?
Color?
Size?
w1
w2
w3
b
Cat?

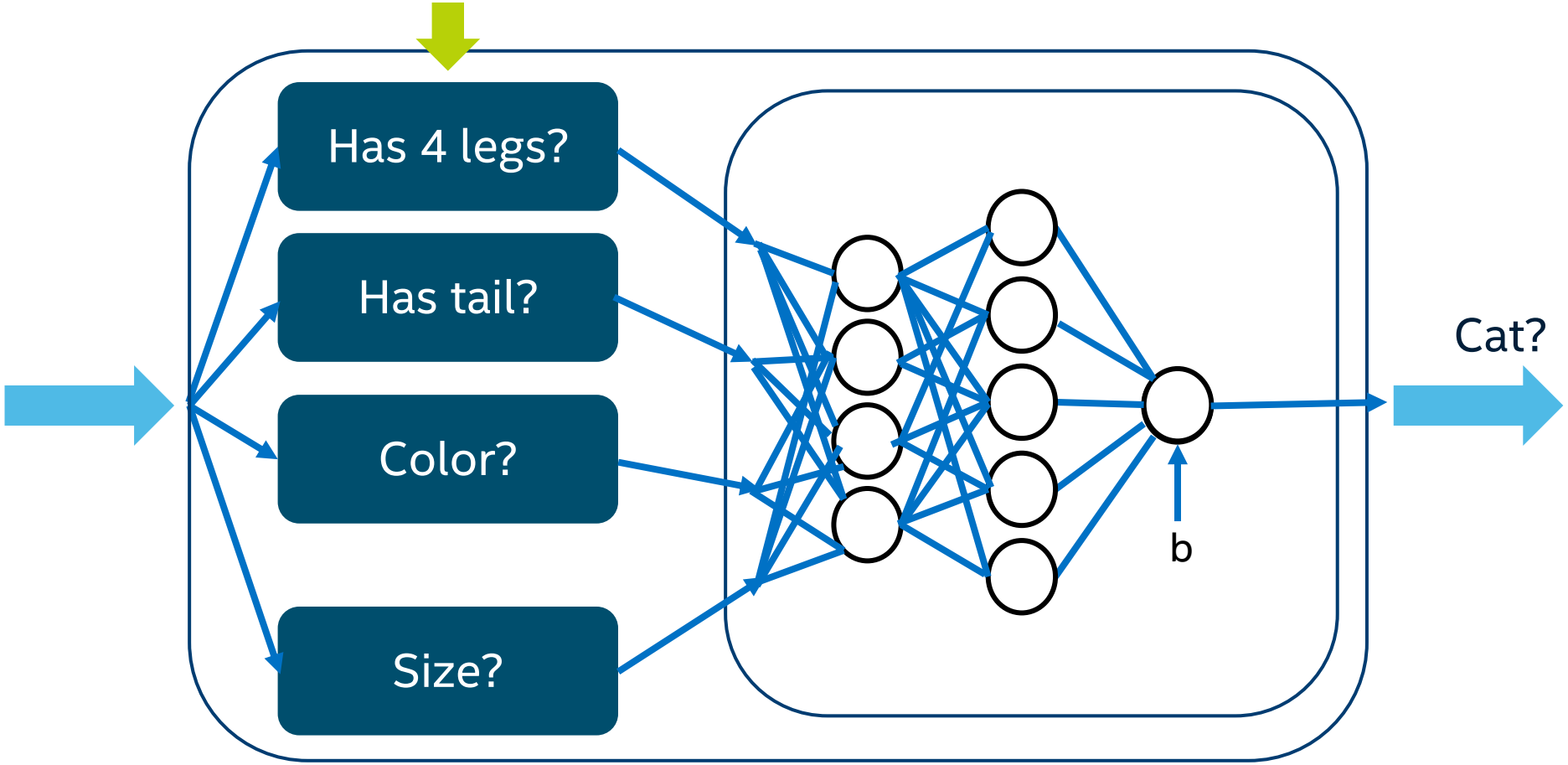

Has 4 legs?
Has tail?
Color?
Size?
w1
w2
w3
b
Cat?

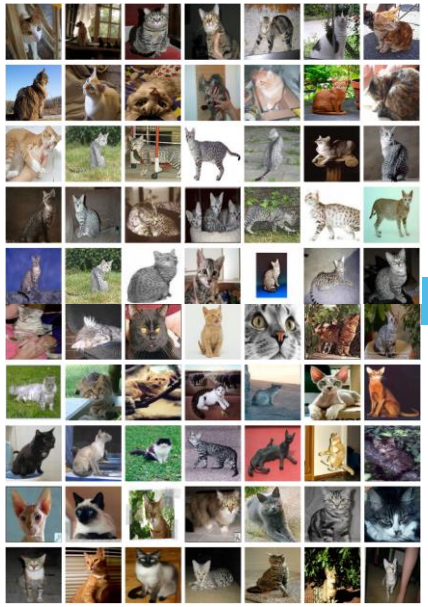
Need more layers,
Need non-linear functions

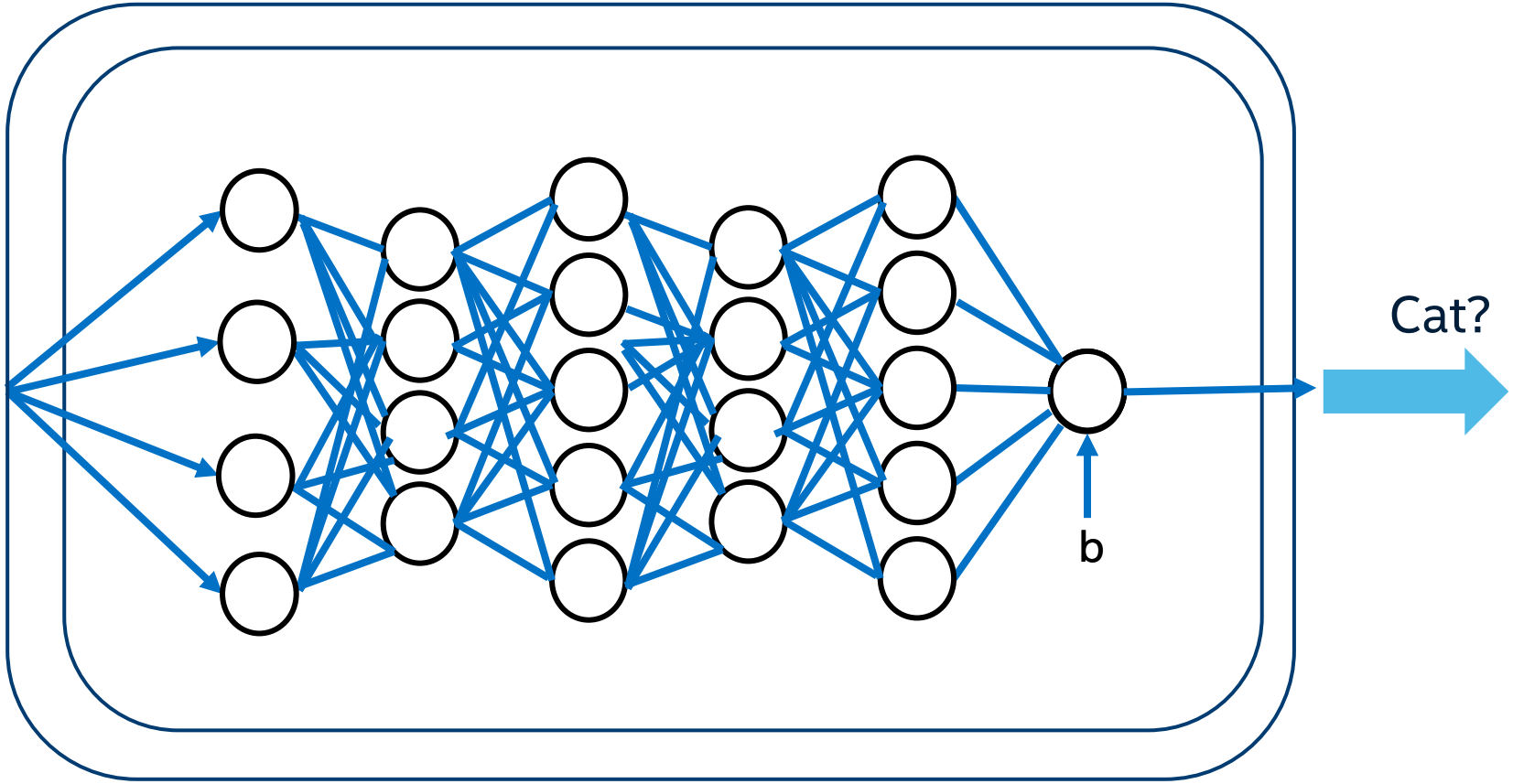



Has 4 legs?
Has tail?
Color?
Size?
b
Cat?

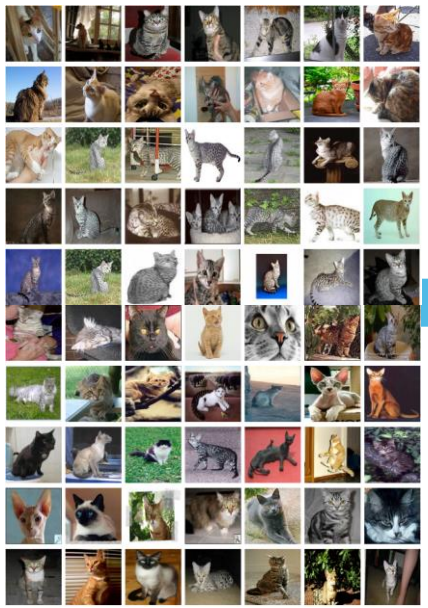
There feature extractors
are also just mathematical functions

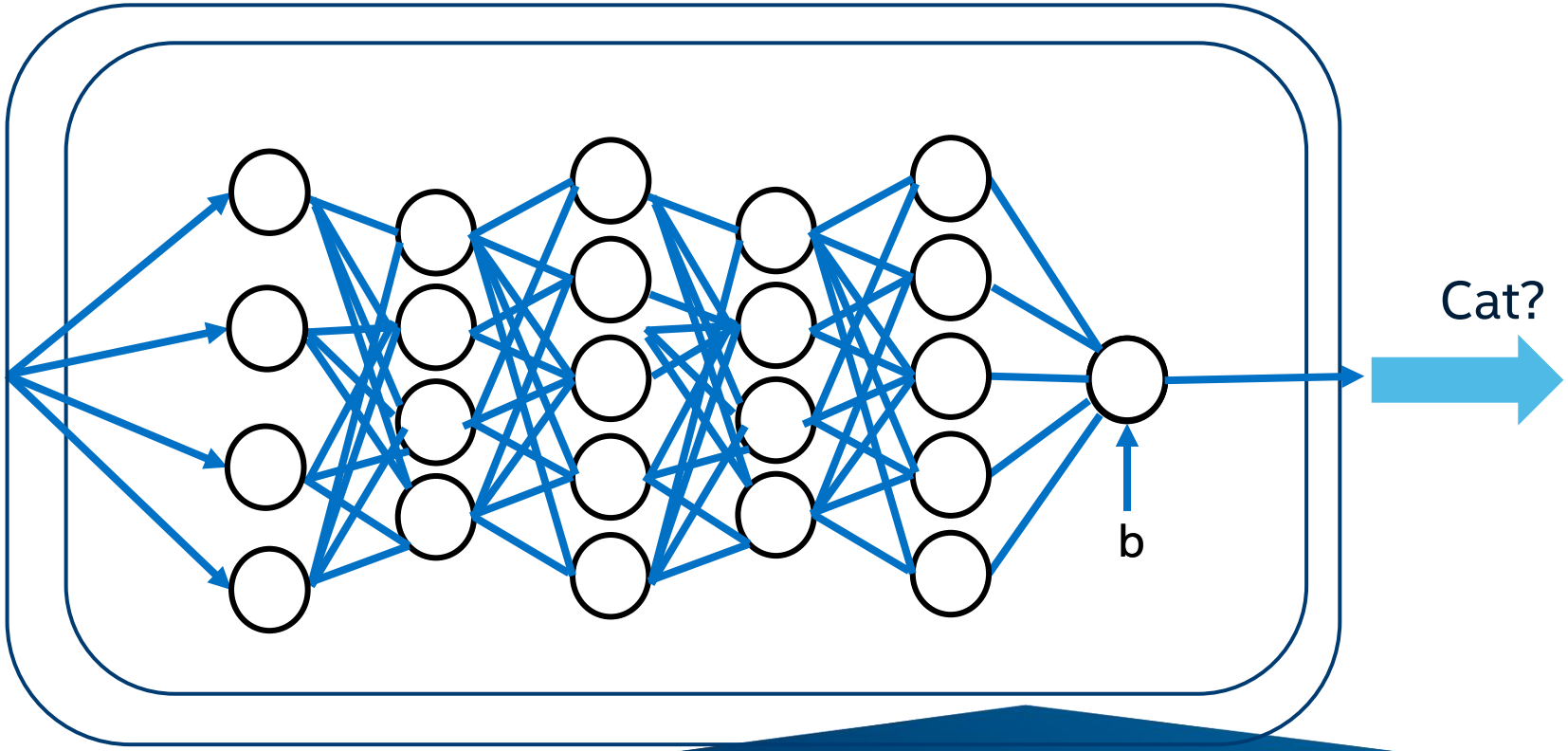





b
Cat?

Forward. Classify/Infer →Define Error

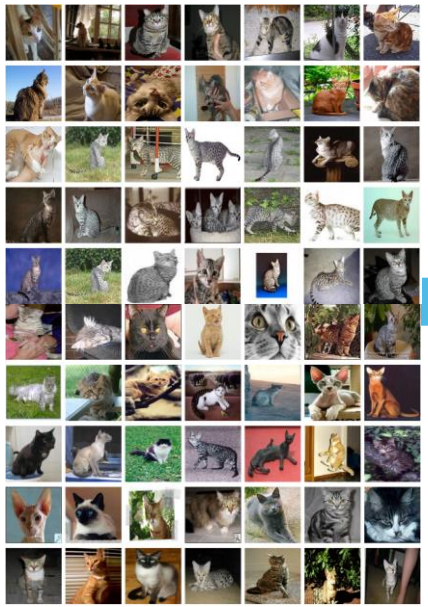


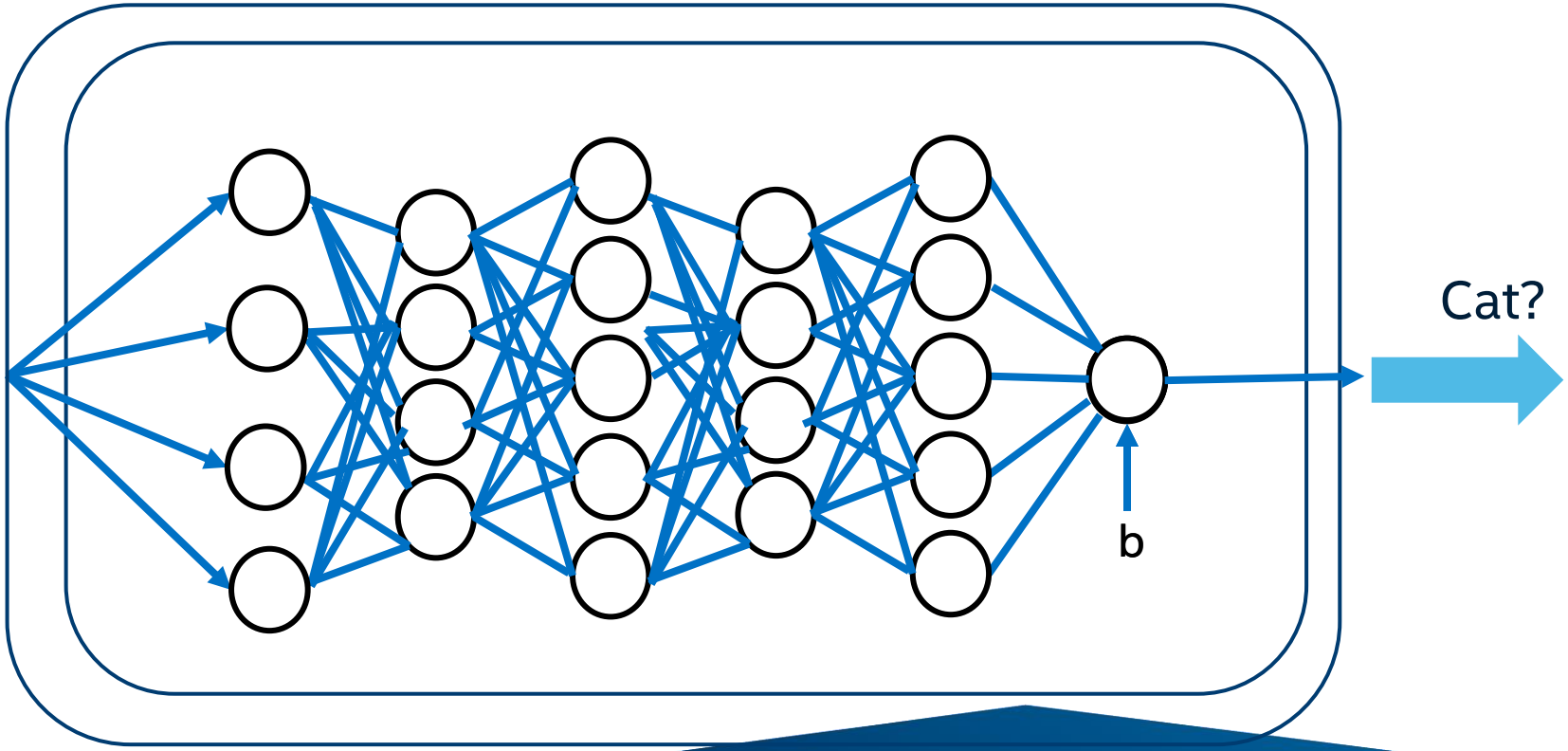


TRAINING

Forward. Classify/Infer →Define Error

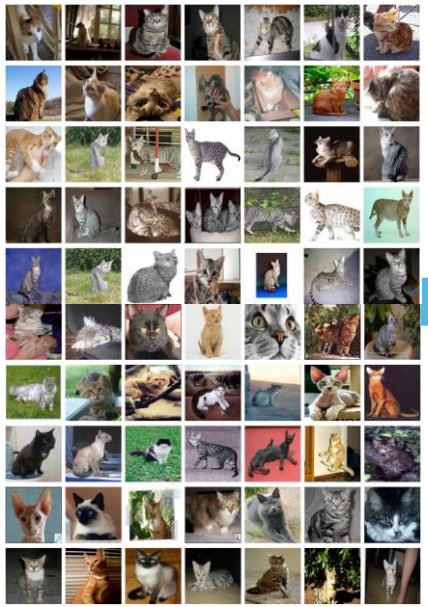
Backward.. Update weights (Backpropagation)





Forward. Classify/Infer →Define Error

INFERENCE



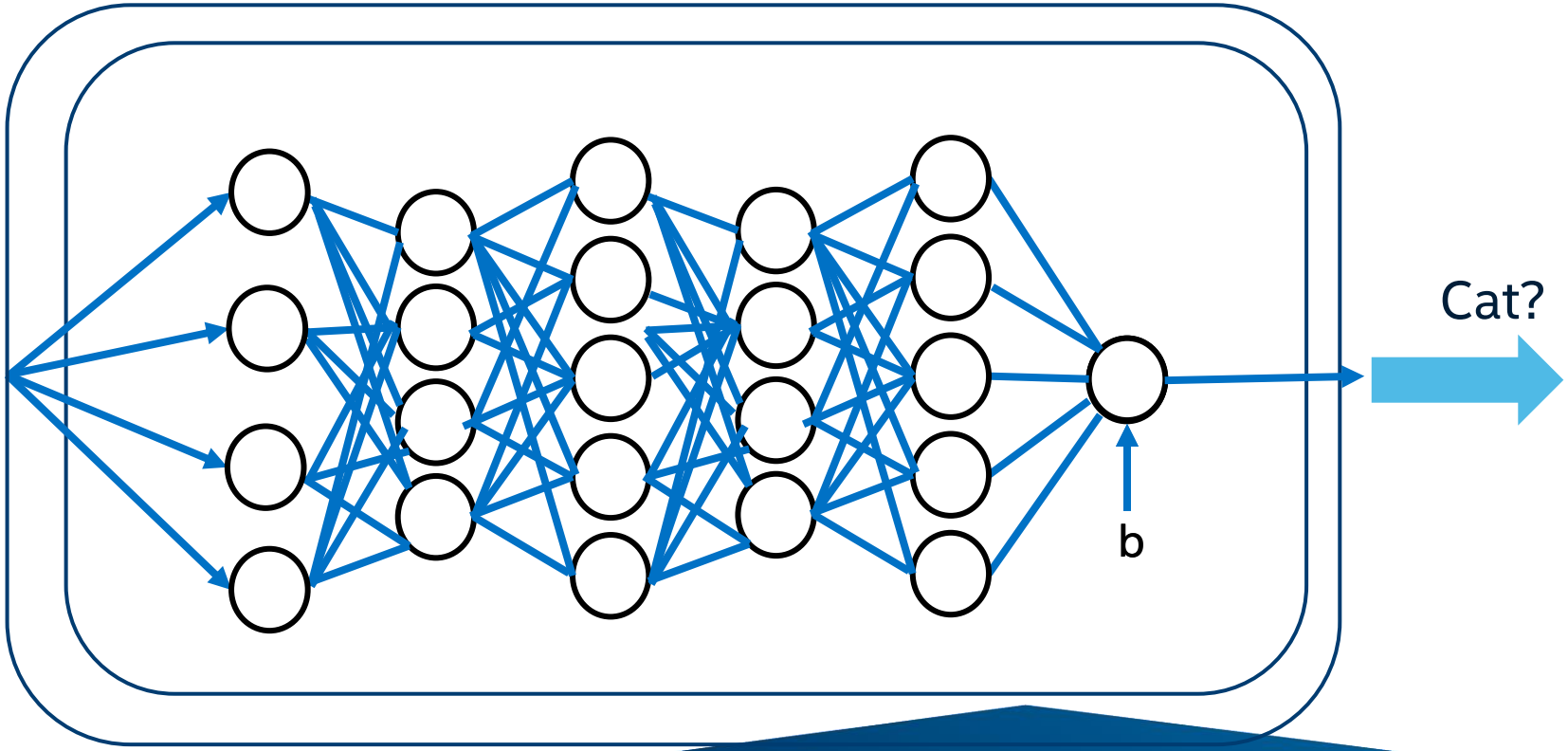
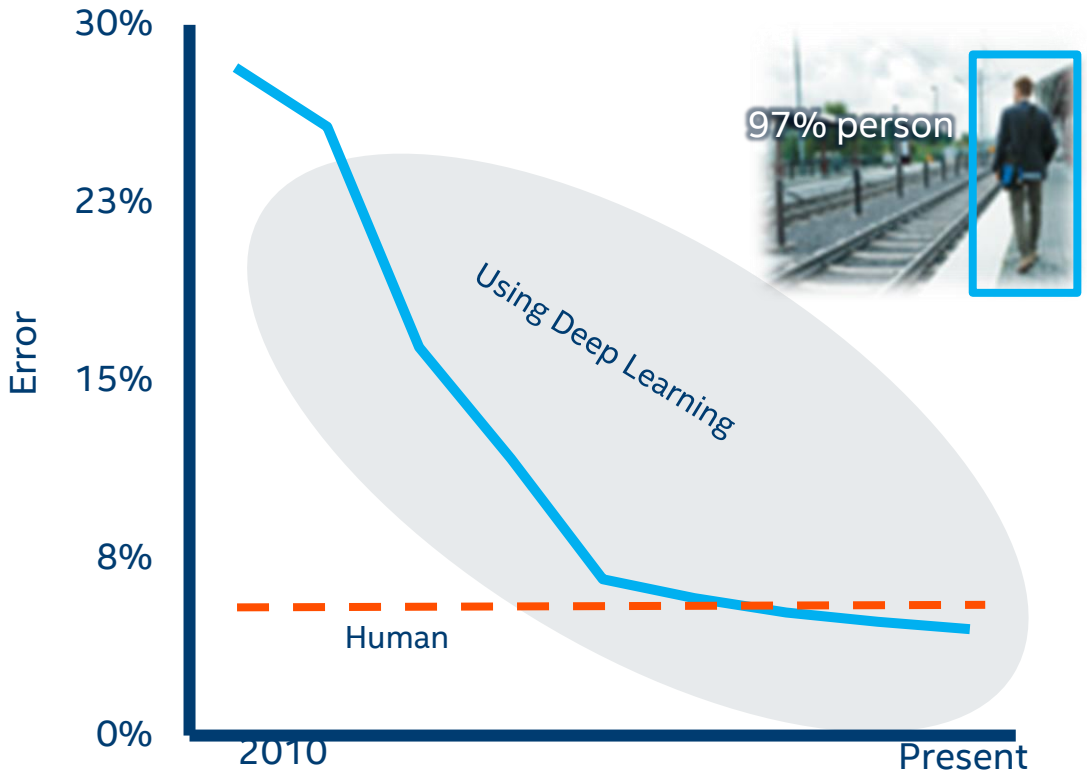
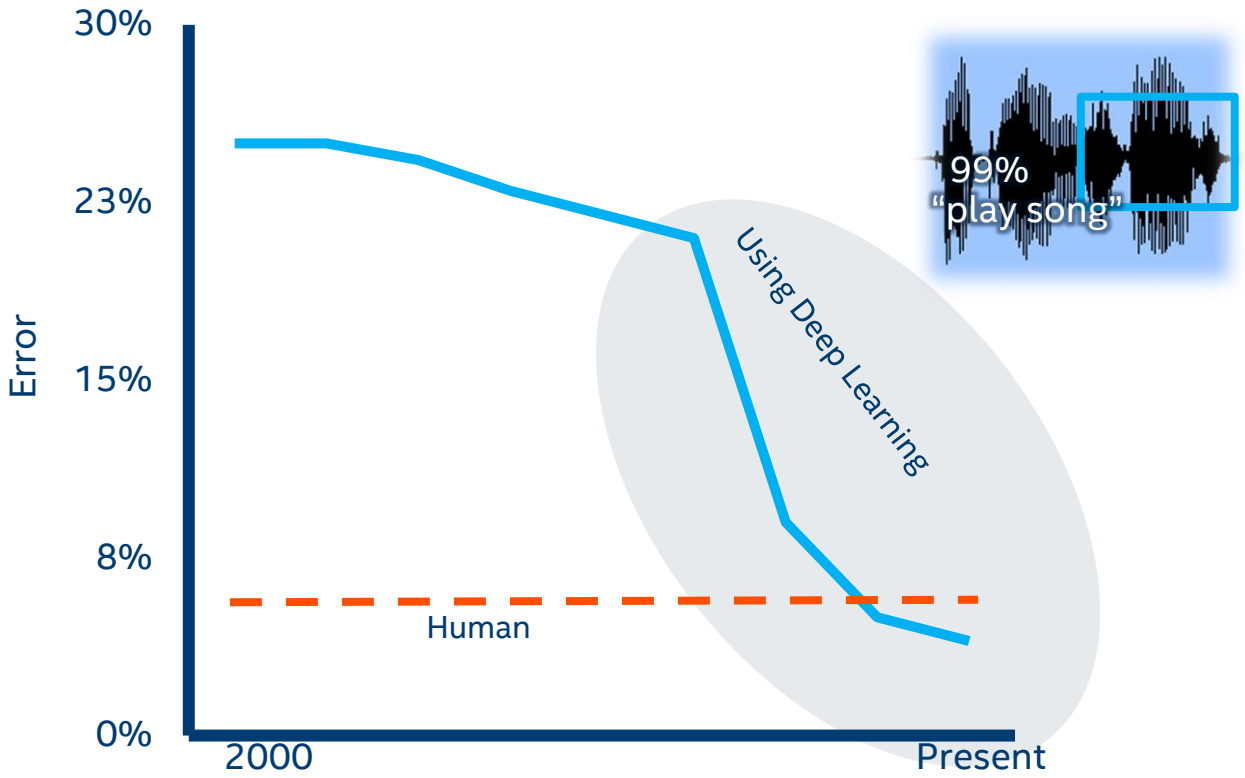


IMAGE RECOGNITION

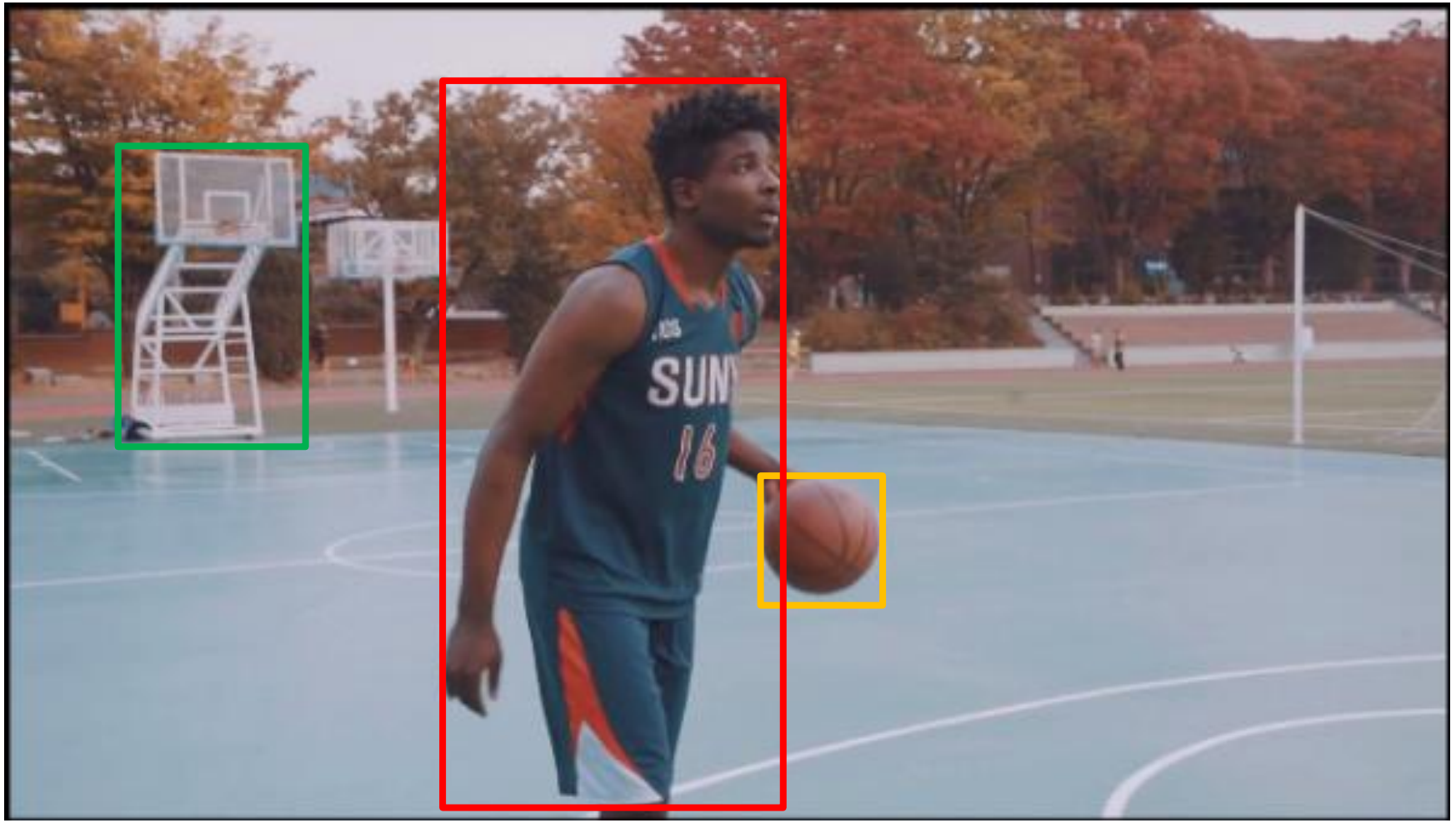


SPEECH RECOGNITION



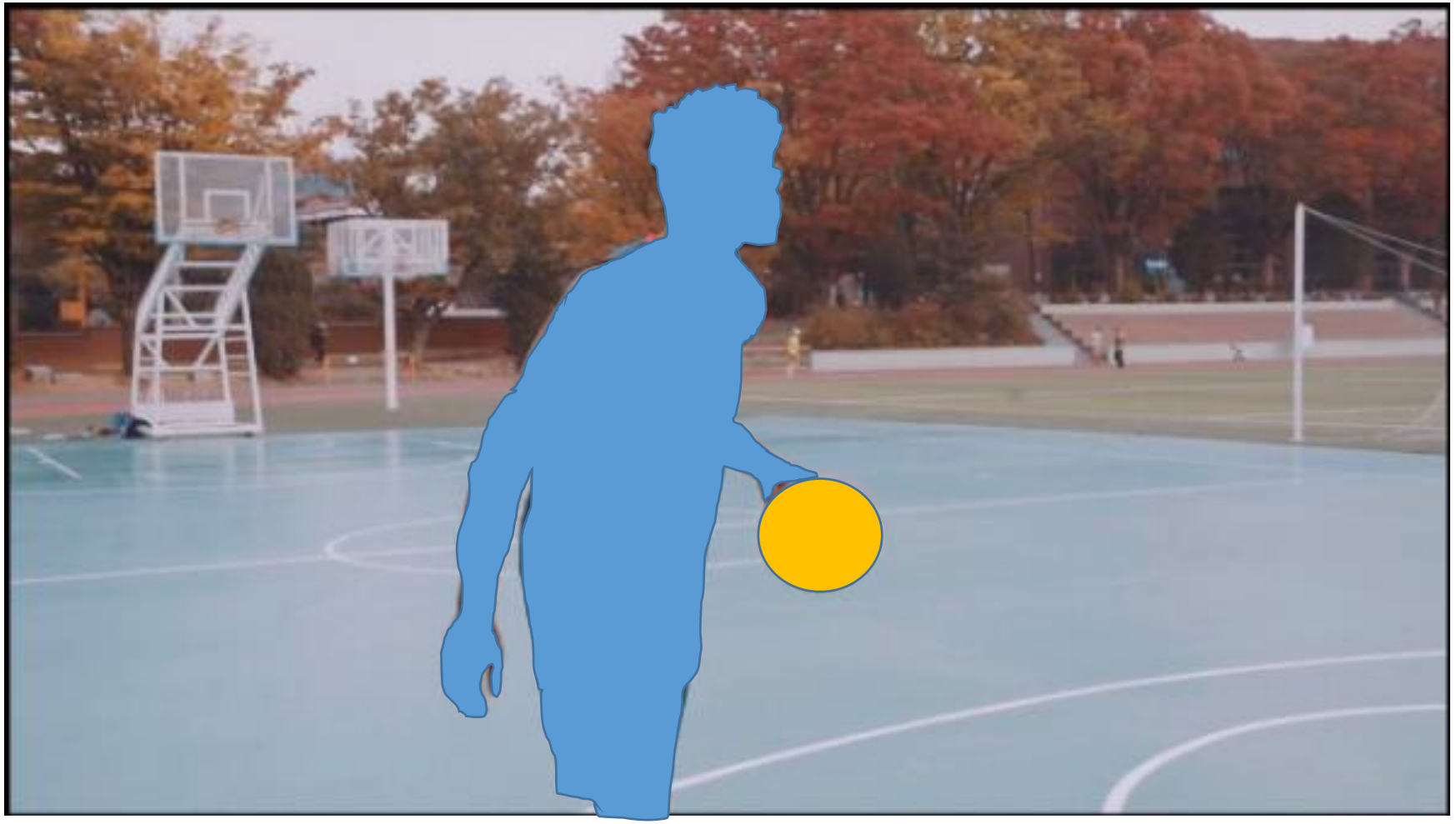
Classification

Detection



Recognition

This is Joe



Segmentation

Accelerating Neural Networks

Huge amount of operations and memory required

e.g. Resnet50 :

- 50 layers, 26M weights, 16M activations, 168MB of memory (FP32)
- 7 GFLOPs (7 Billions Floating point operations)
- Data itself is sometimes big (large images)

Accelerating Neural Networks

Huge amount of operations and memory required

Intel CPUs, iGPU, FPGA and VPUs can do it

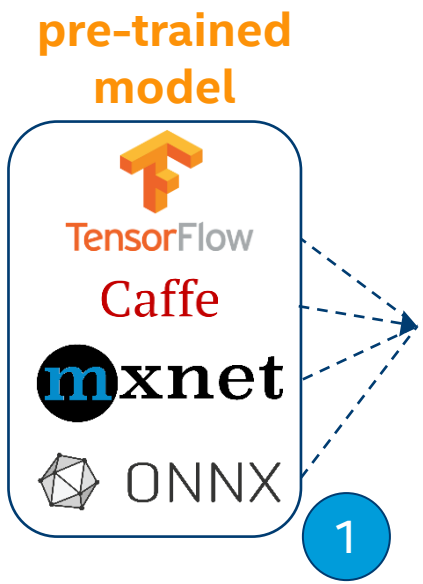
- All our devices can execute neural networks
- Dedicated HW (VPU, FPGA sometimes) are usually more efficient

Accelerating Neural Networks

DLDT - Deep Learning Deployment Toolkit

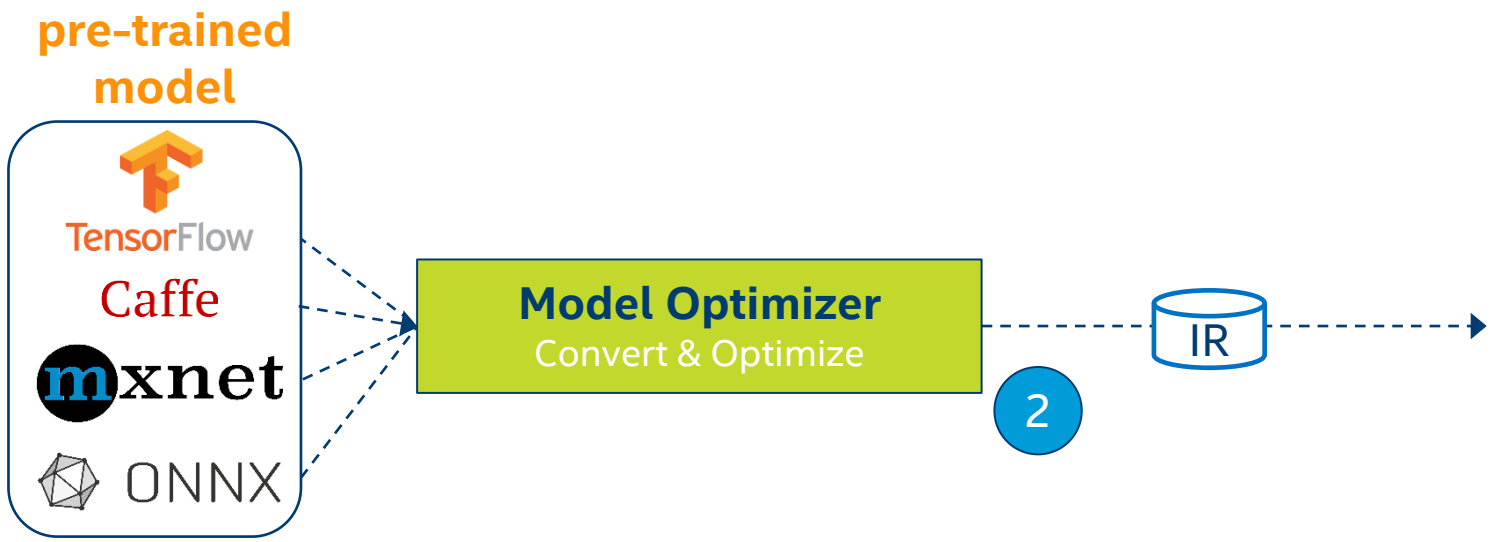
Accelerating Neural Networks

DLDT - Deep Learning Deployment Toolkit



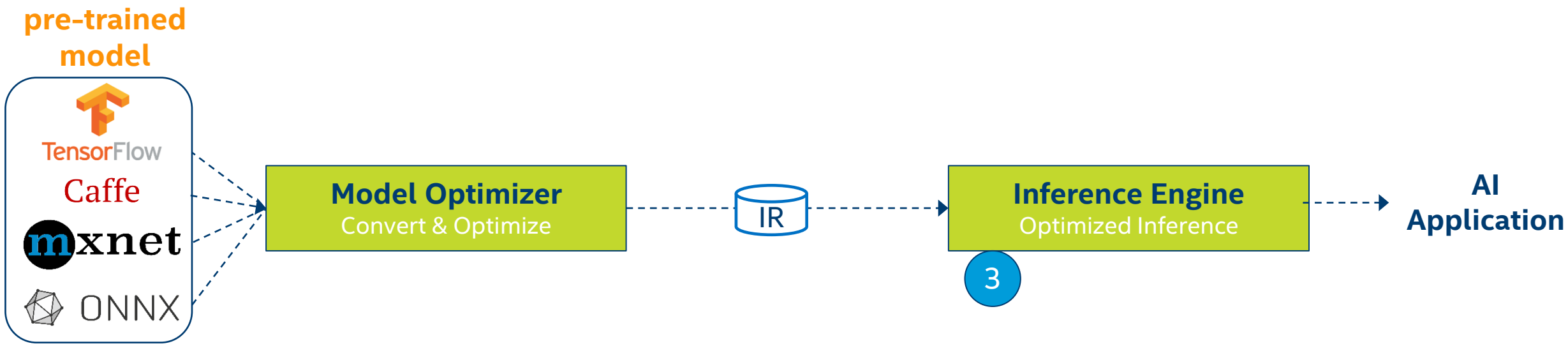
Accelerating Neural Networks

DLDT - Deep Learning Deployment Toolkit



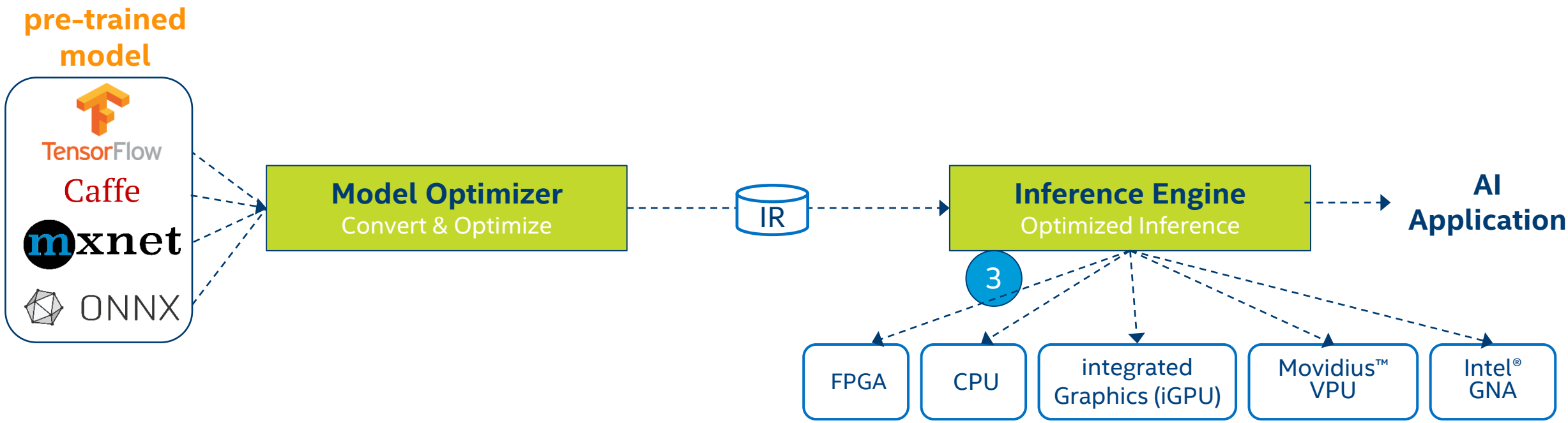
Accelerating Neural Networks

DLDT - Deep Learning Deployment Toolkit



Accelerating Neural Networks

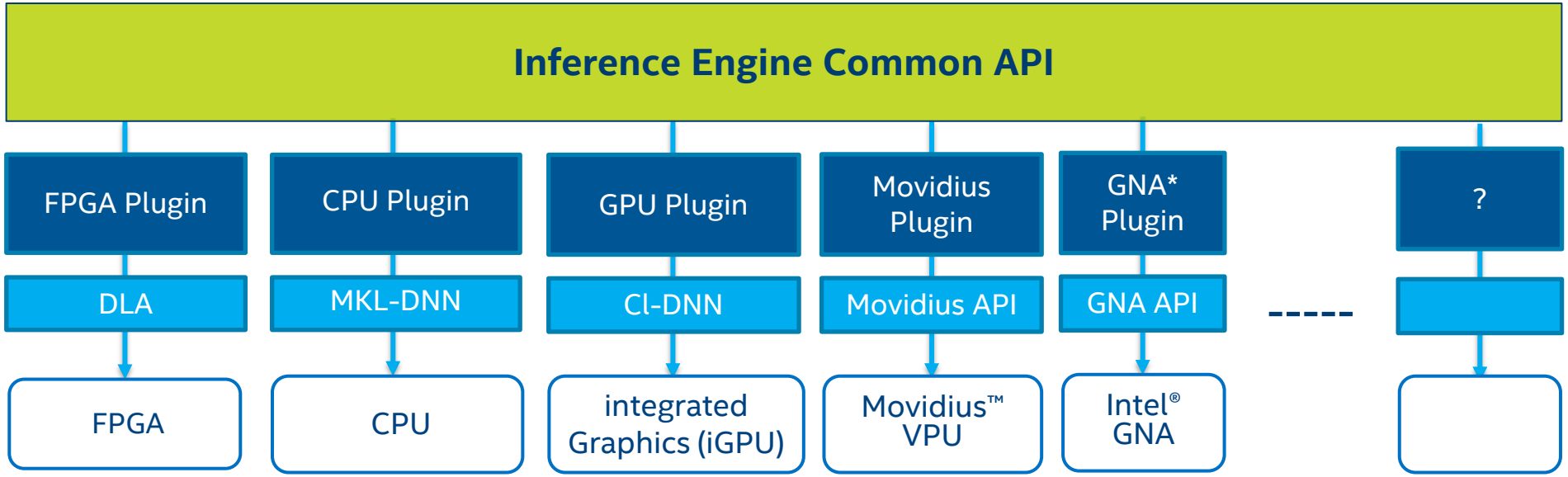
DLDT - Deep Learning Deployment Toolkit



Accelerating Neural Networks

DLDT - Deep Learning Deployment Toolkit

- Optimized libraries per device
- Excellent performance across multiple platforms



Accelerating Neural Networks

DLDT - Deep Learning Deployment Toolkit

- Optimized libraries per device
- Excellent performance across multiple platforms

Resnet50 (Inferences/second)

System	Caffe	TensorFlow	OpenVINO
Apollo-Lake (Atom)	2	2	**11**
Coffee-Lake (Core, i7)	15	31	**134**
Coffee-Lake GPU			**67**

Intel® Distribution of OpenVINO™ toolkit

Deep Learning

Intel® Deep Learning Deployment Toolkit

Model Optimizer
Convert & Optimize

IR

Inference Engine
Optimized Inference

IR = Intermediate Representation file

Open Model Zoo

40+ Pretrained Models

Samples

Model Downloader

Calibration Tool | Benchmark app | Model Analyzer | Accuracy Checker | Aux. Capabilities

Traditional Computer Vision

Optimized Libraries & Code Samples

OpenCV*

OpenVX*

Samples

For Intel® CPU & GPU/Intel® Processor Graphics

Tools & Libraries

Increase Media/Video/Graphics Performance

Intel® Media SDK
Open Source version

OpenCL™ Drivers & Runtimes

For GPU/Intel® Processor Graphics

Optimize Intel® FPGA (Linux* only)

FPGA RunTime Environment
(from Intel® FPGA SDK for OpenCL™)

Bitstreams

Intel® Architecture-Based Platforms Support

Intel® Vision Accelerator Design Products & AI in Production/ Developer Kits

OS Support: CentOS* 7.4 (64 bit), Ubuntu* 16.04.3 LTS (64 bit), Microsoft Windows* 10 (64 bit), Yocto Project* version Poky Jethro v2.0.3 (64 bit), macOS* 10.13 & 10.14 (64 bit)

Summary

- Deep Learning based **Neural Networks** are a very robust way to perform all kinds of vision related tasks.
- In order to get a working neural network model, a lot of data is being processed in a process called **“Training”**
- **Inference** is the forward path of the network, it provides the network result.
- DL models can be used for classification, detection, segmentation and many other tasks
- Intel platforms can execute neural networks, with very good performance
- The DLDT (Deep Learning Deployment Toolkit) is Intel SW to accelerate deep learning inference and is part of OpenVINO