



embedded **VISION** SUMMIT 2018

The Journey & Sunrise Processors: Leading-edge Performance for Embedded AI



Dr. Kai Yu

Founder & CEO, Horizon Robotics

Horizon at a Glance —— Pioneer of Embedded AI Solutions in China

embedded
VISION
SUMMIT
2018



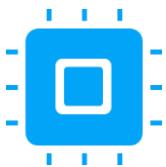
2015
founded



500+
employees



112
patents



1st AI ASIC

The screenshot shows a news article from the Chinese Government website (www.gov.cn) dated December 21, 2017. The headline reads "我国首款嵌入式人工智能视觉芯片发布" (The launch of China's first embedded AI visual chip). The article discusses the development of a chip by a team from the Chinese Academy of Sciences' Institute of Computing Technology, which can process 200 visual targets simultaneously. It highlights the chip's application in intelligent driving and smart city development.



**Technology
Wor**Economy**
Forum 2017**

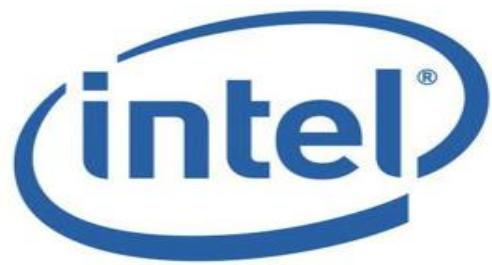
**The only Chinese
company ever recognized
as a Technology Pioneer
by the World Economic
Forum**

**Previous recipients
include Google, Palantir,
Twitter and Airbnb**

Supported by a group of renowned investors



+more



地平线
Horizon Robotics

A+ round -- more than 100 million USD, led by Intel Capital

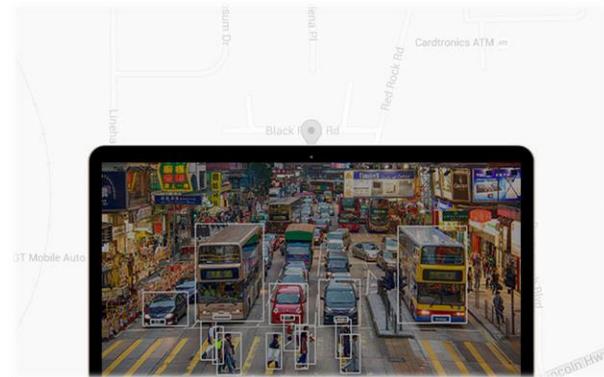
Focus on Edge AI Applications

Smart Mobility, Smart City, Smart Retail

embedded
VISION
SUMMIT
2018



Smart Mobility



Smart City



Smart Retail

The paradigm shift of AI computing



CPU

2011 Google DistBelief

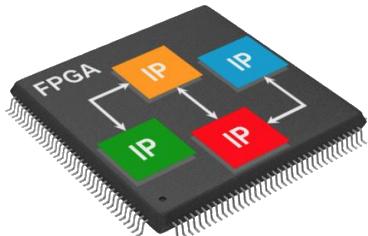


GPU

2012 Baidu

Good for training

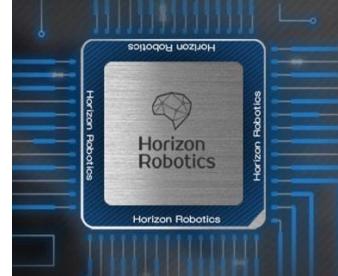
Good for inference



FPGA

2014
Baidu,
Micorsoft

Since 2012, Kai Yu and colleagues have been pushing the paradigm shift of AI computing, from CPU to GPU and FPGA, and then BPU



TPU & BPU

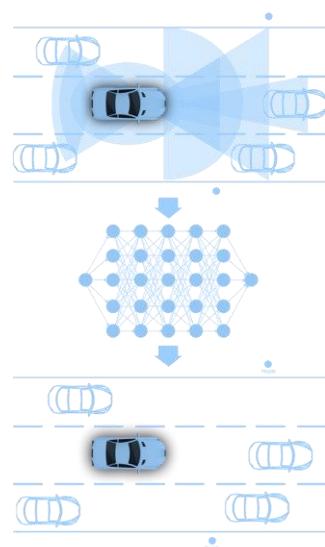
In 2015 BPU (Brain Processing Unit) architecture was proposed by Horizon Robotics for inference acceleration;

In 2016 TPU (Tensor Processing Unit) was announced by Google

The roadmap of our software platform

Hugo

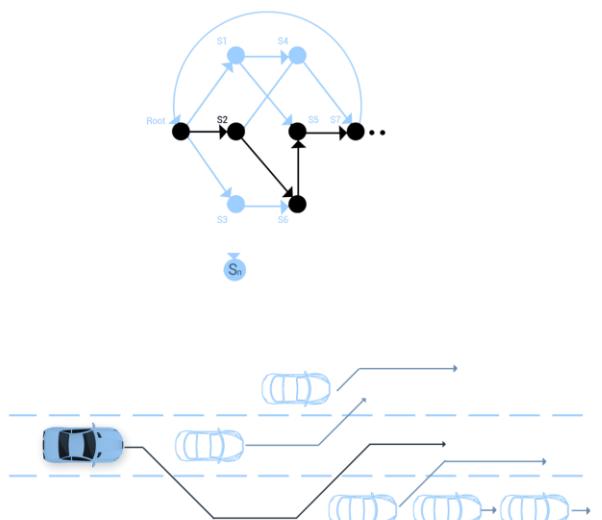
2017 : Perception



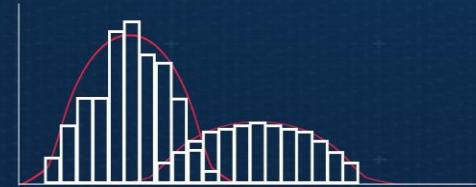
2018 : Modeling



2019 : Decision Making



The Roadmap of BPU Architecture for both FPGA and ASICs



Gauss Architecture

Spec: (L1-L2)

1080p@30fps object detection, recognition

Detect & recognize 250 objects per frame

Detect up to 8 categories of objects

17Q1

17Q2

17Q3

17Q4

18Q1

18Q2

18Q3

18Q4



Bernoulli Architecture

Spec: (L3-L4)

Sparse & fixed-point neural networks

1080p@30fps x 6 cameras

Detection, prediction, scene parsing

Multi-sensor fusion

$$\mathcal{P}(\mathcal{A}/\mathcal{B}) = \frac{\mathcal{P}(\mathcal{B}/\mathcal{A}) * \mathcal{P}(\mathcal{A})}{\mathcal{P}(\mathcal{B})}$$

Bayes Architecture

Spec: (L4-L5)

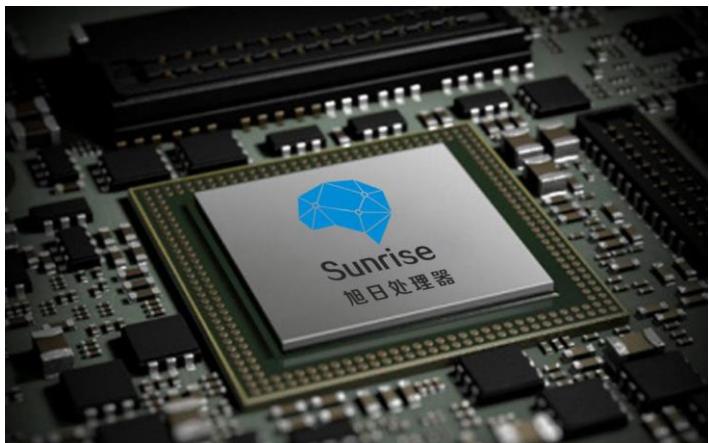
CNN, RNN, Bayes network inference

4K@30fps x 12 cameras

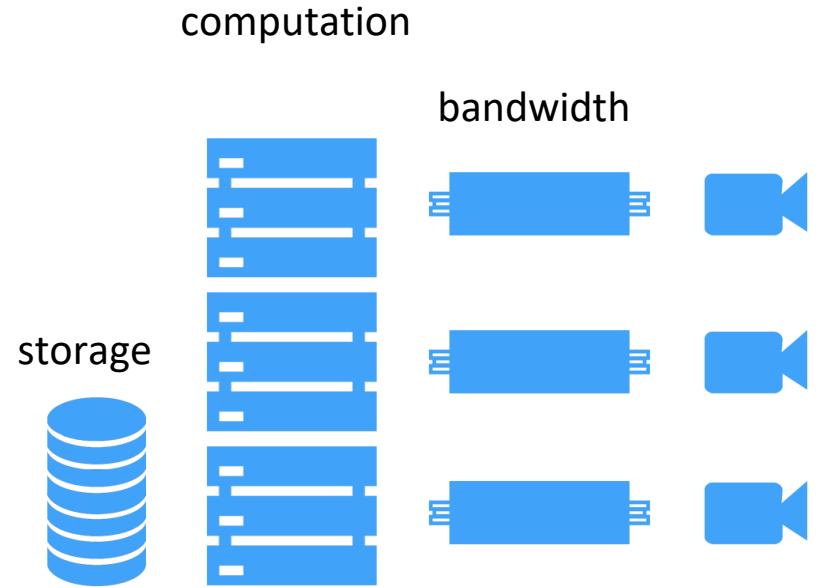
Semantic 3D scene model

Trajectory planning

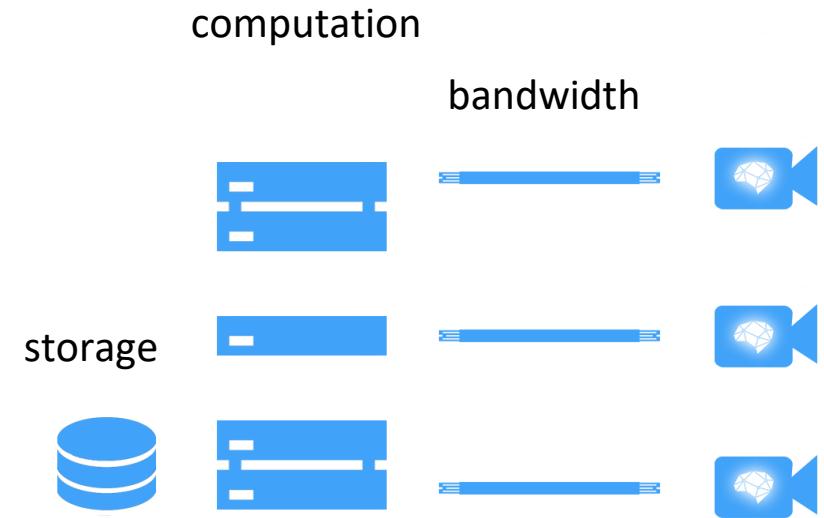
Sunrise 1.0 ASIC processor for smart IP cameras



Video analytics for smart city: reduce the cost by 10x



traditional approach:
AI computation on servers



new approach:
AI computation on cameras

Sunrise Edge AI Processor – High Performance, Low Power Cost

Sunrise 1.0 – Redefining the Smart Camera



20 Watts



1.5 Watts



30 Objects
Simultaneously



200+ Objects
Simultaneously



12 Frames
/ Second



30 Frames
/ Second



Note:

1. Brain Processing Unit is our proprietary and high-performance AI processor architecture that supports ARM/GPU/FPGA/ASICs



- Using Nvidia TX1, power consumption 20w, 12 frames per second, detecting 30 objects simultaneously
- Using Horizon Sunrise 1.0 processor, power consumption **1.5w, 30 frames per second**, detecting **200 objects** simultaneously



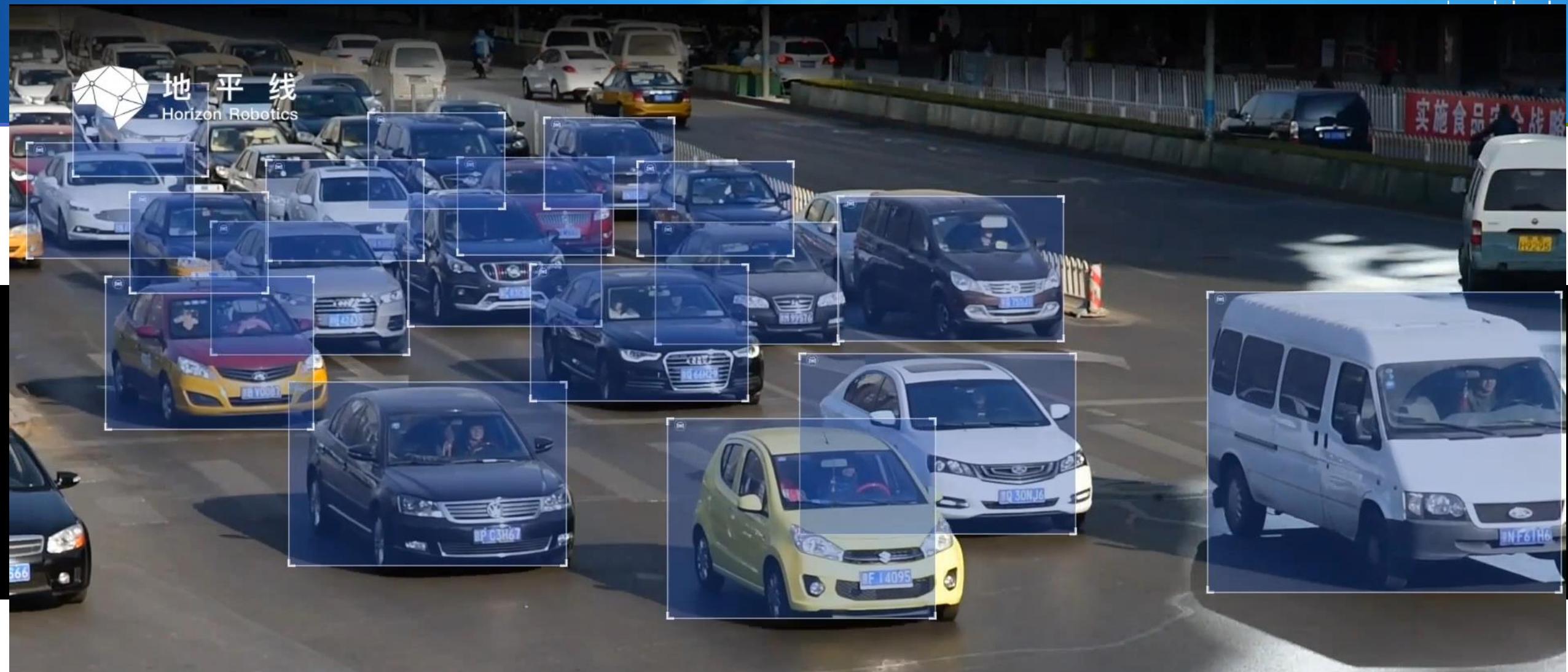
地平线
Horizon Robotics



 女性 青年	 男性 青年	 女性 青年	 男 青						
--------------	--------------	--------------	--------------	--------------	--------------	--------------	--------------	--------------	------------



地平线
Horizon Robotics



机动车总量 **901** 辆

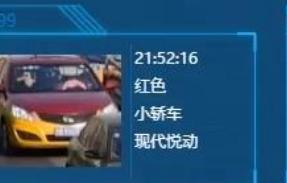
非机动车总量 **331** 辆

行人

机动车

非机动车

行人总量 **379** 人





113



111



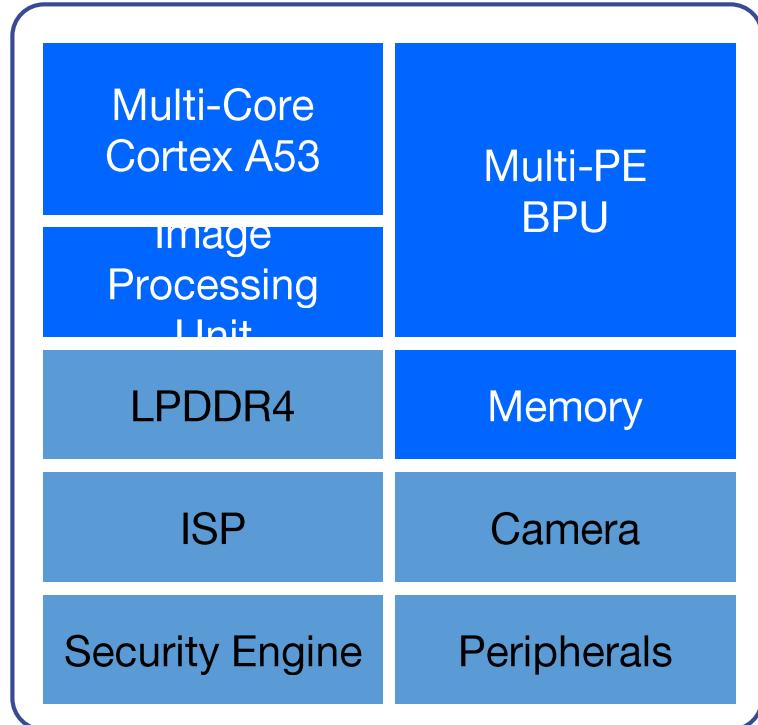
114



平铺显示

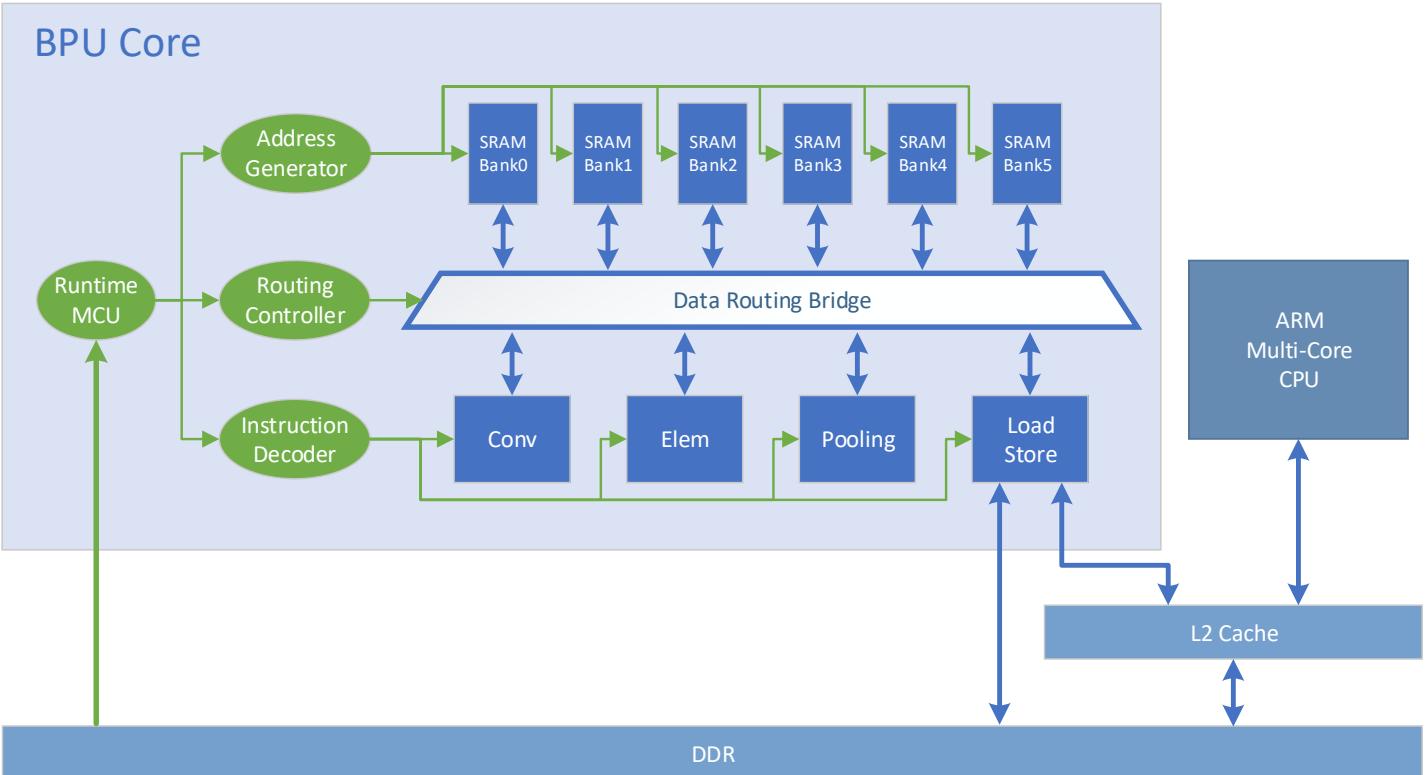


Journey 2.0 AI Processor



- Multi-PE BPU delivering over 4 TOPs/s
- Multi-Core Cortex A53 up to 1.5GHz
- Advanced ISP and IPU(Image processing unit)
- 4K single or 720p stereo depth camera input
- On-chip memory, up to 640GB/s internal bandwidth
- 32bit LPDDR4 @ 3200Mbps
- Typical Power: 1Watt
- TSMC 28nm HPC+
- Package : FC-BGA14mm×14mm

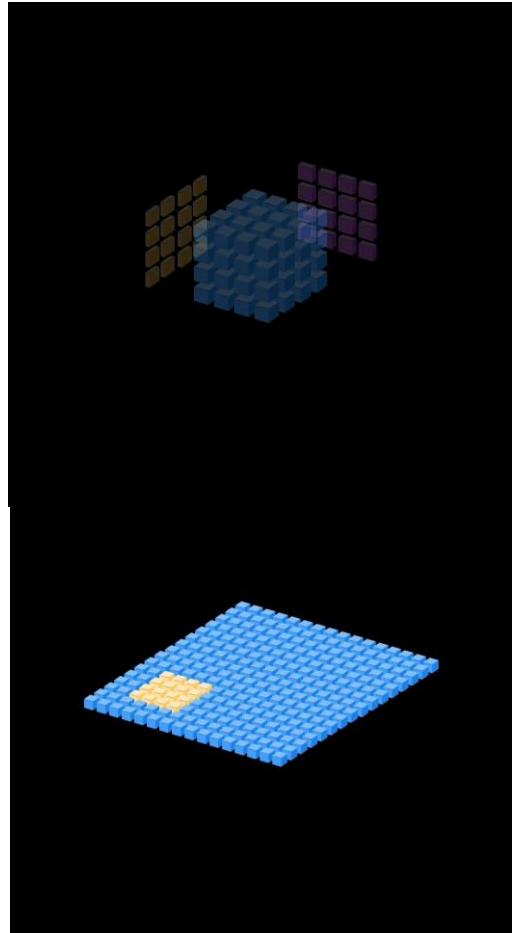
BPU Architectural Highlights



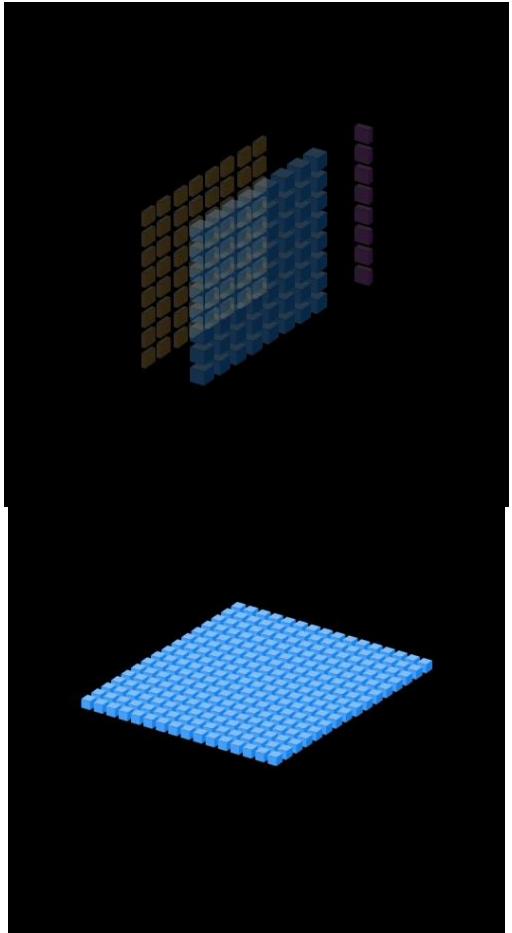
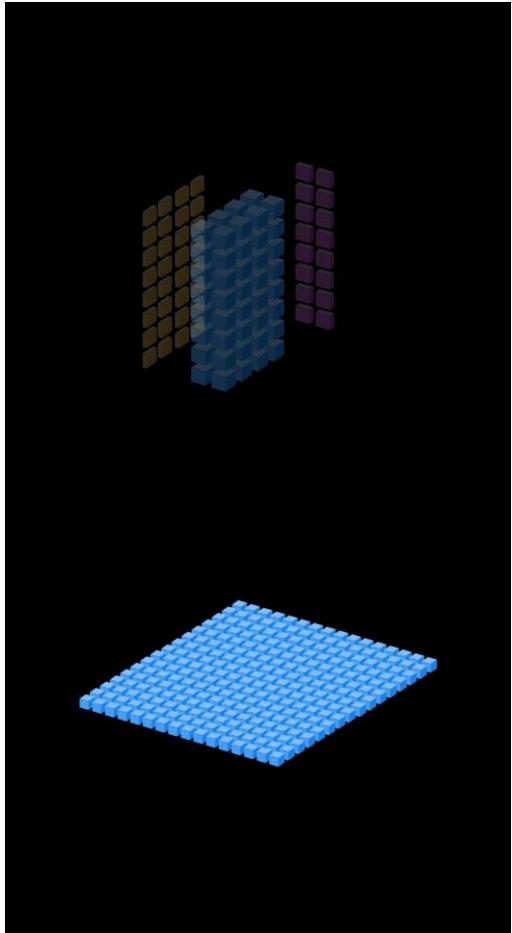
- 64bit MIMD (Multi-Instruction and Multi-Data) processor
- 24 stages pipeline
- High MAC utilization: >90%
- Highly optimized DDR bandwidth: <50MB/s @1080p image

Elastic tensor core – optimize the NN computation

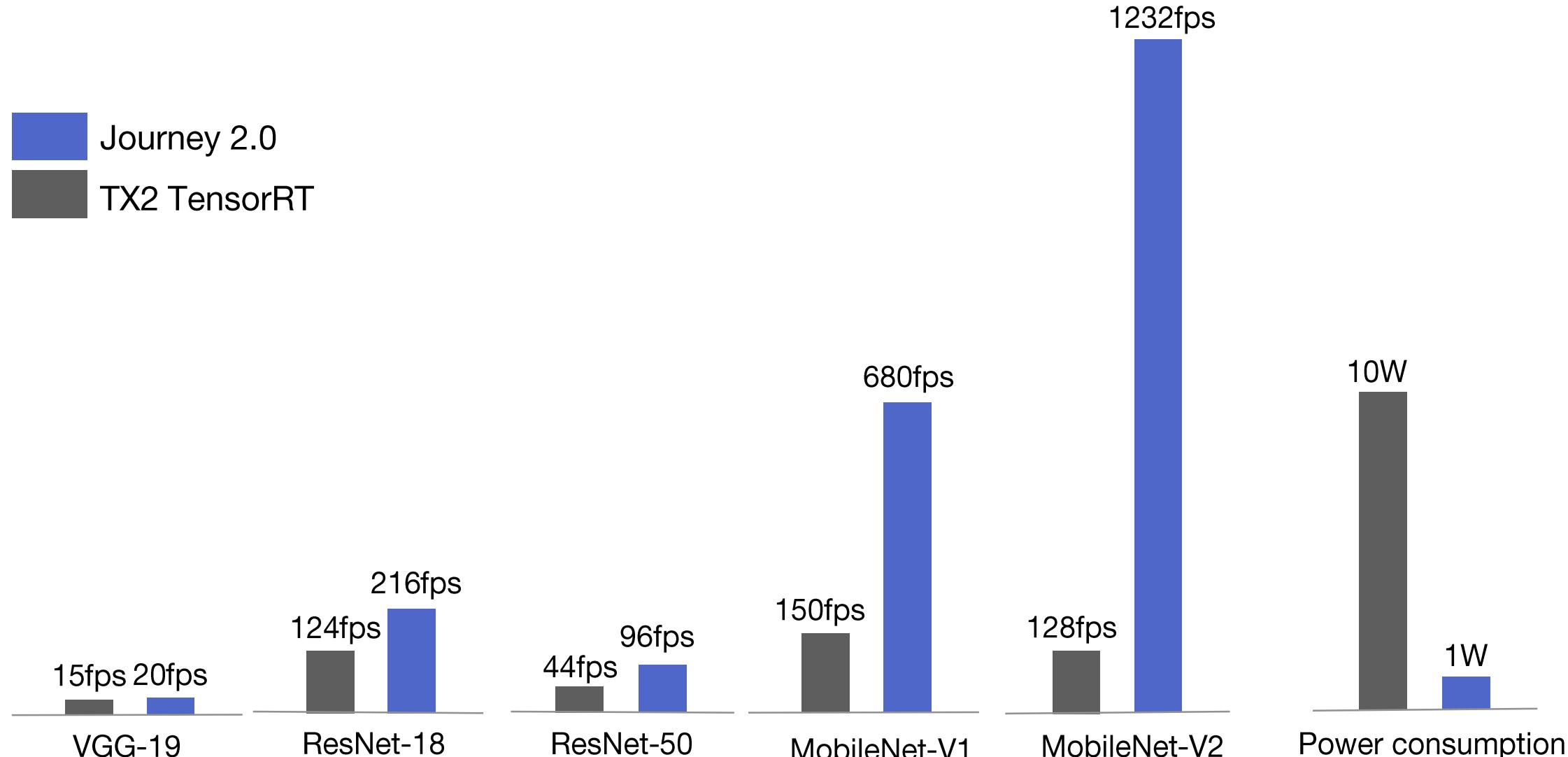
Elastic Tensor Core



Memory Fabric

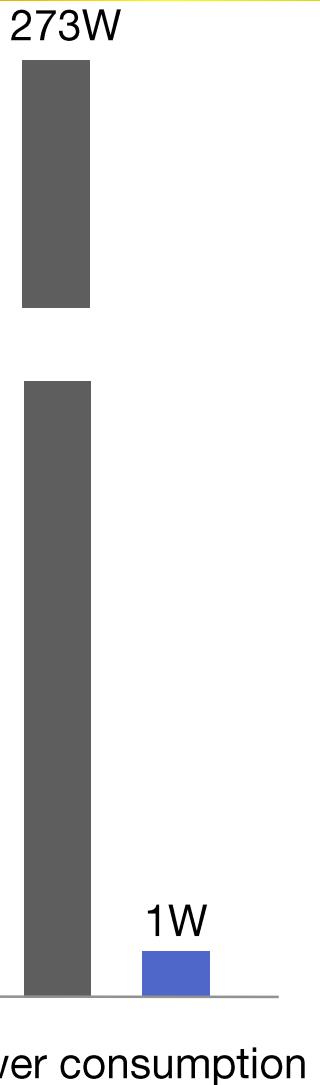
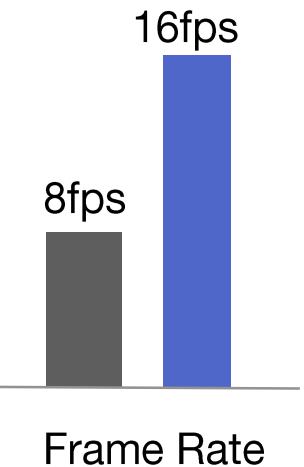
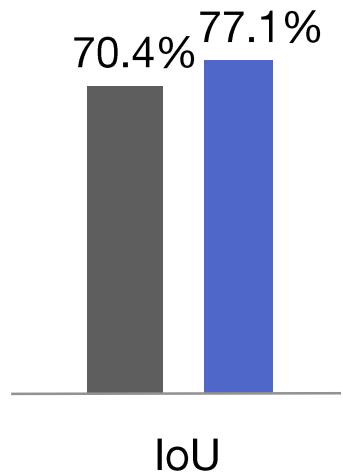


Classification Performance on ImageNet

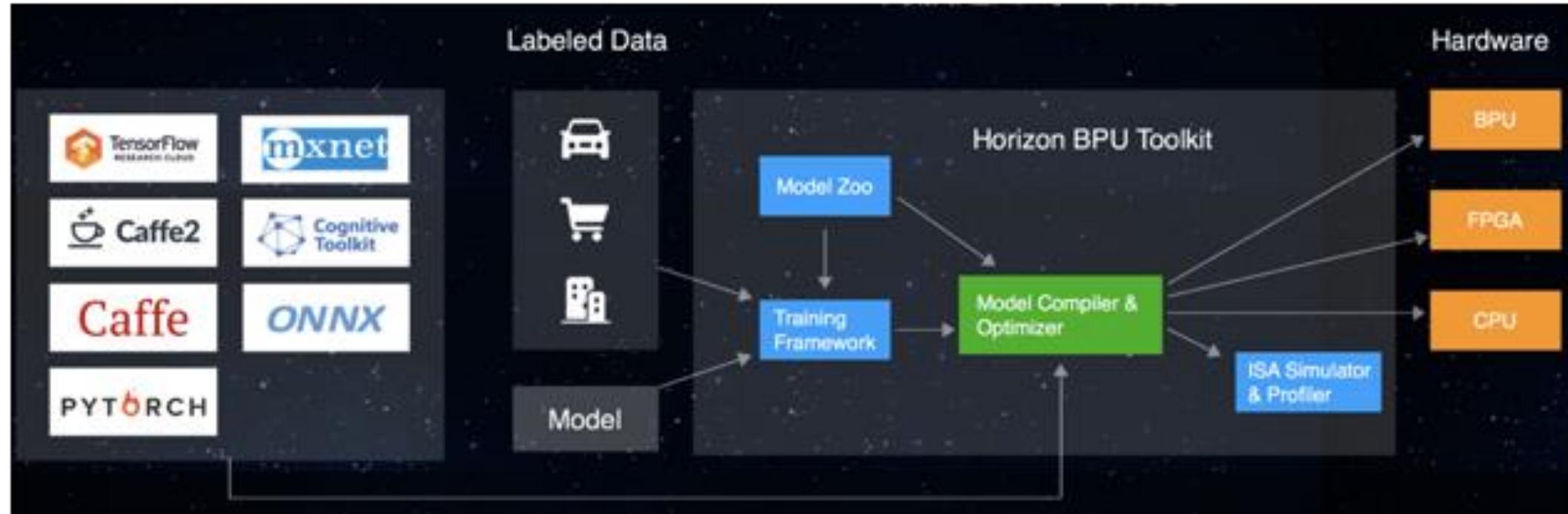


Segmentation Performance on Cityscapes

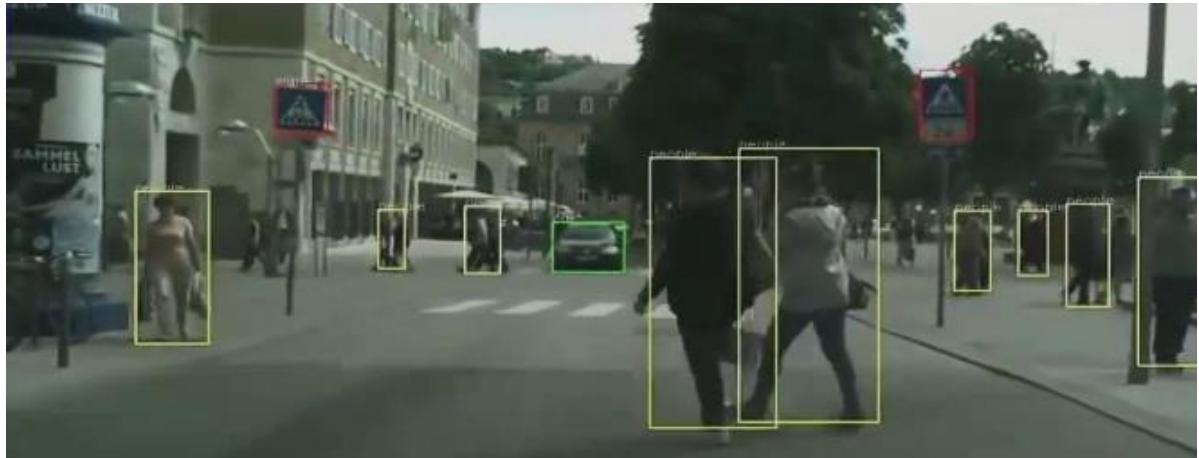
Journey 2.0
Titan X



Horizon BPU Toolkit



Journey 2.0 architecture



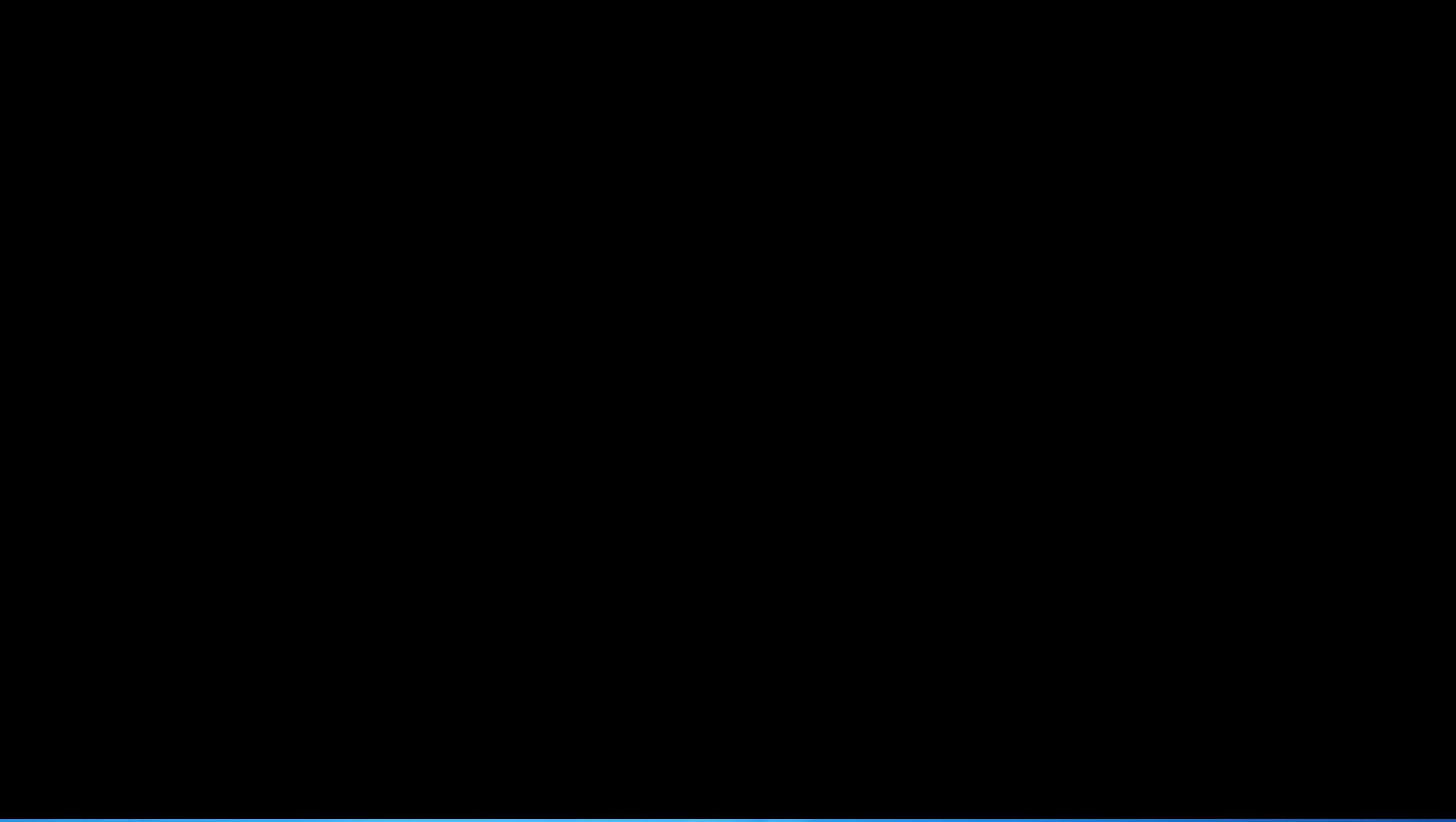
Boundingbox based detection

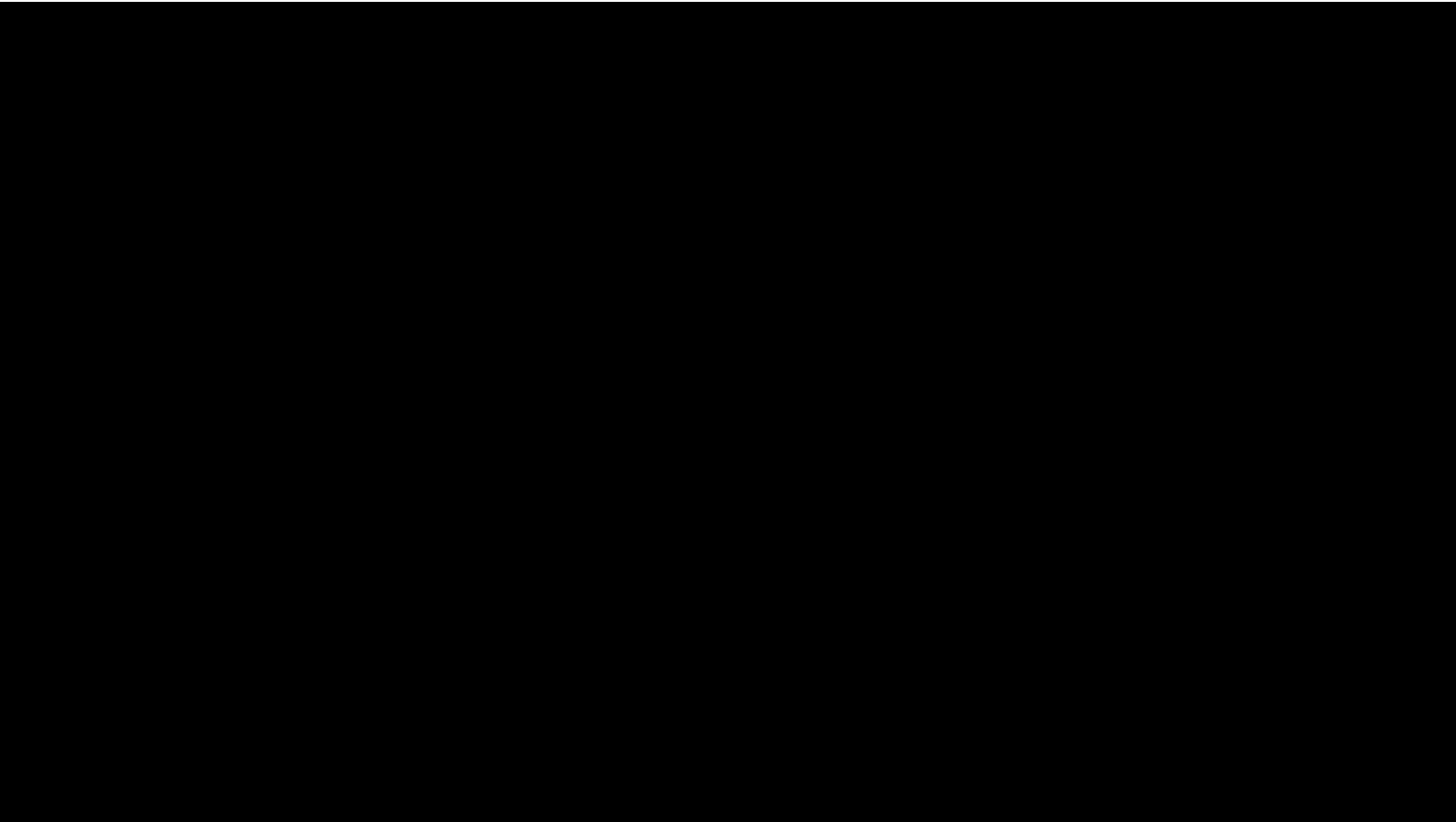
vs. Pixel-wise segmentation

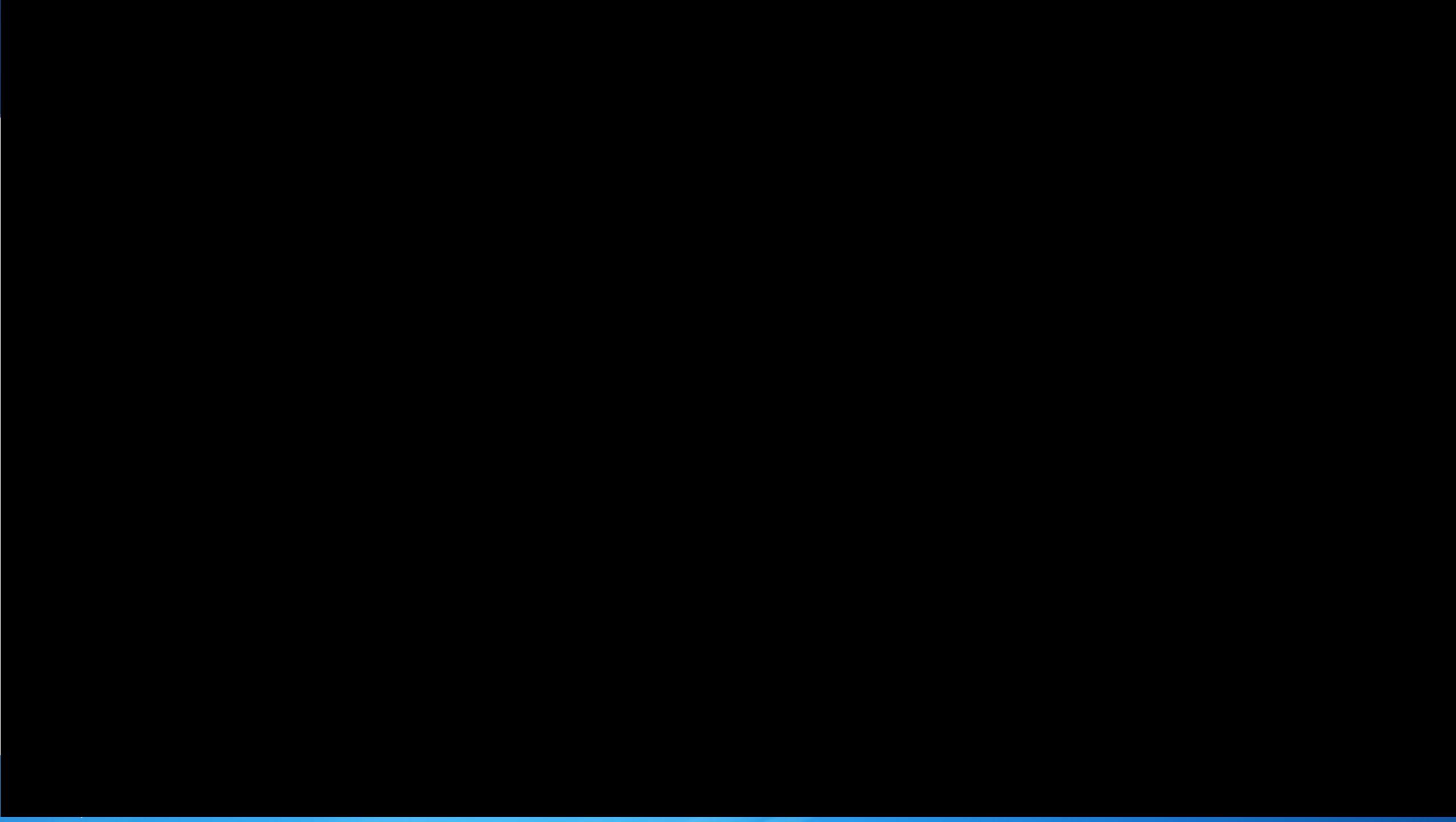
Matrix: autonomous driving computing platform based on Journey 2.0

embedded
VISION
SUMMIT
2018









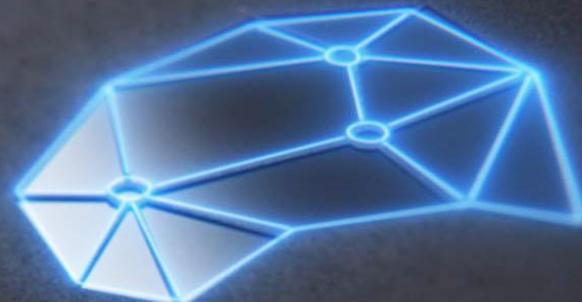


地平线
horizon robotics

麦行

庆卓机械

Sunrise Now, Journey to the Future



We hire engineers at
Beijing, Shanghai, Nanjing, Silicon Valley
dream@hobot.cc