



基于 Simulink 和 NVIDIA Jetson 开展机器学习

雷昊, NVIDIA



袁航, MathWorks



MATLAB EXPO

内容概要

- NVIDIA 和 MathWorks 的合作
- 什么是 Jetson AGX Orin
- workflow: Simulink to Jetson AGX Orin
- 示例: Jetson AGX Orin 车道和车辆检测算法部署

NVIDIA 和 MATHWORKS 的合作

从嵌入式到企业系统 - NVIDIA GPU 和 MATLAB 集成

通用科学和工程计算

支持使用 NVIDIA GPU 计算
超过1000个函数:

- MATLAB
- Deep Learning Toolbox
- Image Processing Toolbox
- Signal Processing Toolbox
- Plus others...

无需 CUDA 编程



本地部署和云计算

MATLAB 和 NVIDIA GPU CLOUD

Instantly access on-premises and
cloud GPUs with MATLAB and NGC



嵌入式系统

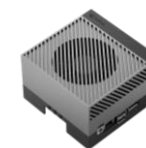
由MATLAB代码自动生成
优化的CUDA代码

MATLAB &
Simulink

GPU Coder

CUDA

cuBLAS, cuSolver,
cuDNN, TensorRT



实现AI应用嵌入式部署需要解决的问题

- 为AI应用场景选择合适的硬件
- 仿真并测试AI应用
- 将AI应用部署到嵌入式设备



POLL # 1

How long have you been working with AI?

- a. < 1 year
- b. < 2 years
- c. 2+ years
- d. 5+ years

POLL # 2

What AI application focus:

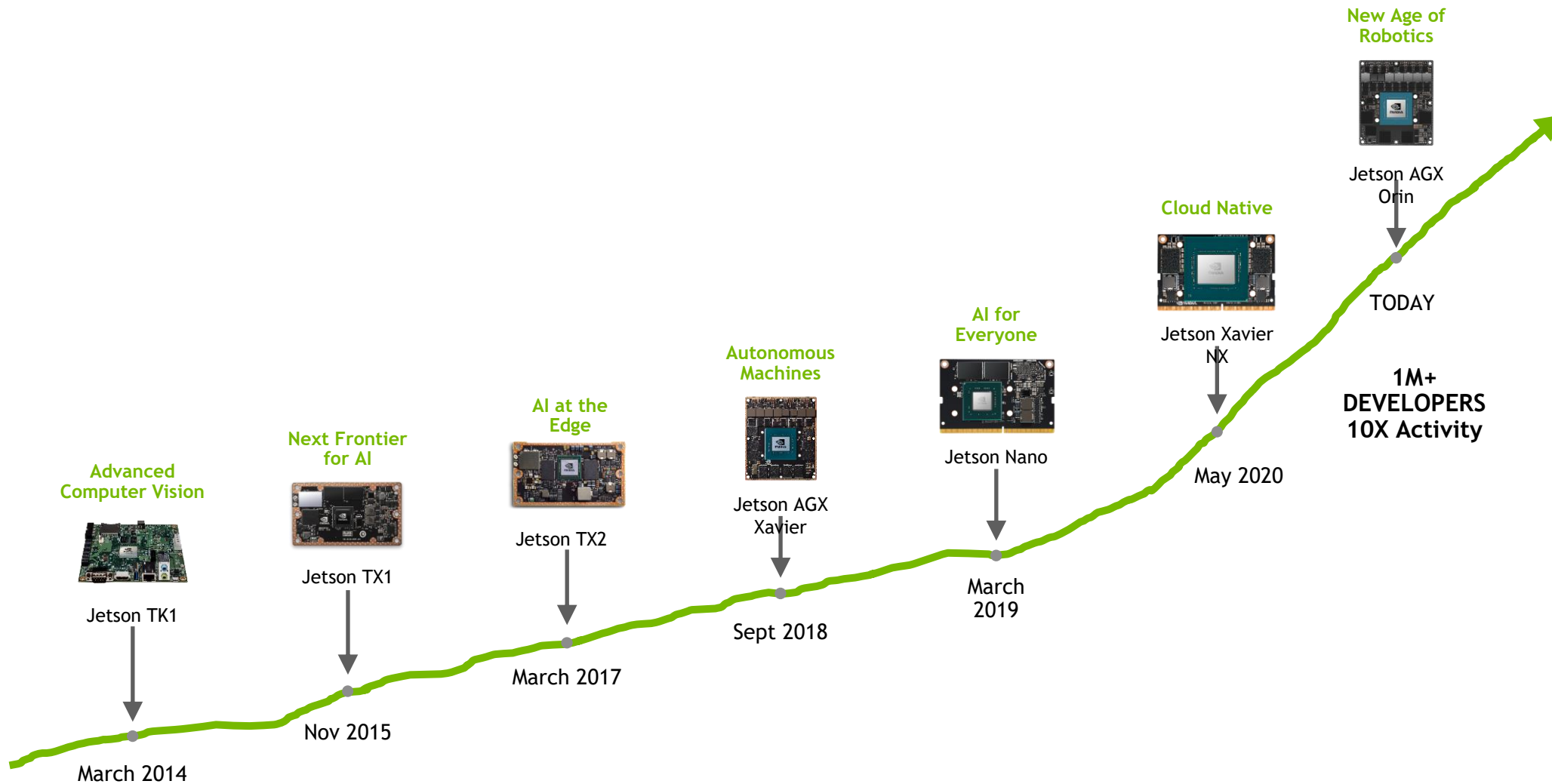
- a. NLP
- b. Computer vision
- c. Speech
- d. Recommender systems
- e. Other

实现AI应用嵌入式部署需要解决的问题

- 为AI应用场景选择合适的硬件
- 仿真并测试AI应用
- 将AI应用部署到嵌入式设备

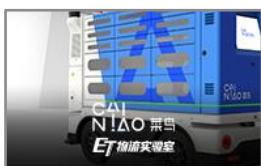
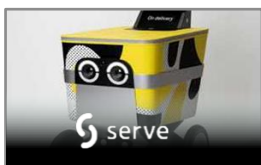


NVIDIA JETSON 赋能下一代智慧物联网以及机器人应用



JETSON 在边缘计算的应用

DELIVERY



RETAIL



AMR



AGRICULTURE



SERVICES



LOGISTICS



MANUFACTURING



HEALTHCARE



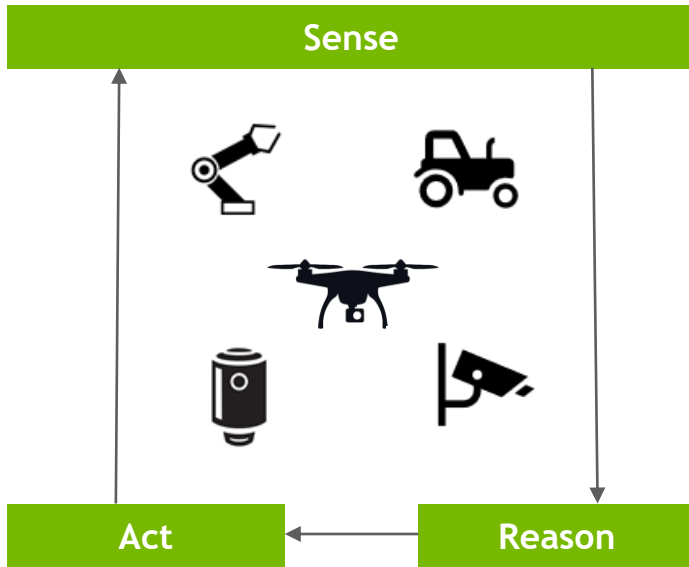
1,000,000+
Developers

6,000+
Customers

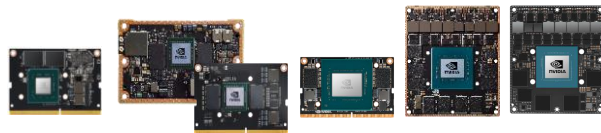
软件定义的AI平台

AI at the Edge

Sensor Fusion & Compute Performance

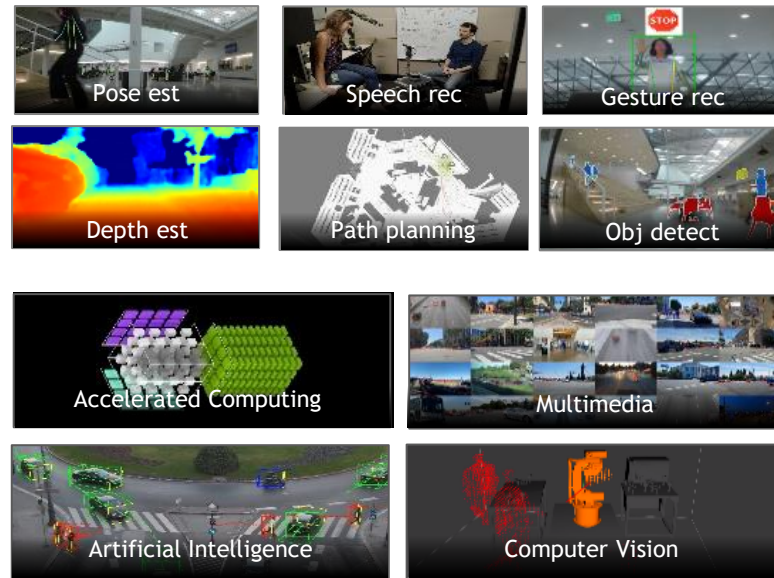


JETSON COMPUTER



SOFTWARE DEFINED

SDK, Design Tools, Libs, GEMs



Jetpack SDK · CUDA · TensorRT · TensorFlow · ONNX · ROS

ECOSYSTEM

Expertise, Time to Market



持续增长的生态合作伙伴 (130+)



DISTRIBUTION



SOFTWARE



HW AND SENSORS



NVIDIA JETSON 产品线

For AI at the Edge and Autonomous Machines

Next Gen
Jetson

JETSON Orin NX Series

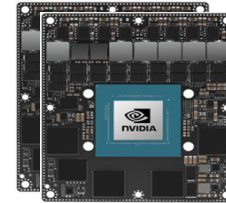
100 TOPs (INT8)



10 - 25W
8GB/16GB
45mm x 70mm

JETSON AGX Orin Series

275 TOPs (INT8)



15 - 60W
32GB/64GB
100mm x 87mm

ENTRY

MAINSTREAM

AUTONOMOUS MACHINES

JETSON NANO

0.5 TFLOPS (FP16)



5 - 10W
45mm x 70mm

JETSON TX2 series

1.33 TFLOPS (FP16)



7.5 - 15W*
45mm x 70mm
50mm x 87mm

JETSON Xavier NX series

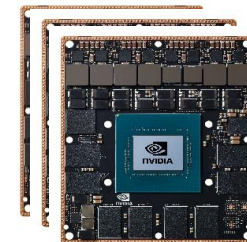
6 TFLOPS (FP16)
21 TOPs (INT8)



10 - 20W
8GB/16GB
45mm x 70mm

JETSON AGX Xavier Series

11 TFLOPS (FP16)



10 - 30W**
32GB/64GB
100mm x 87mm

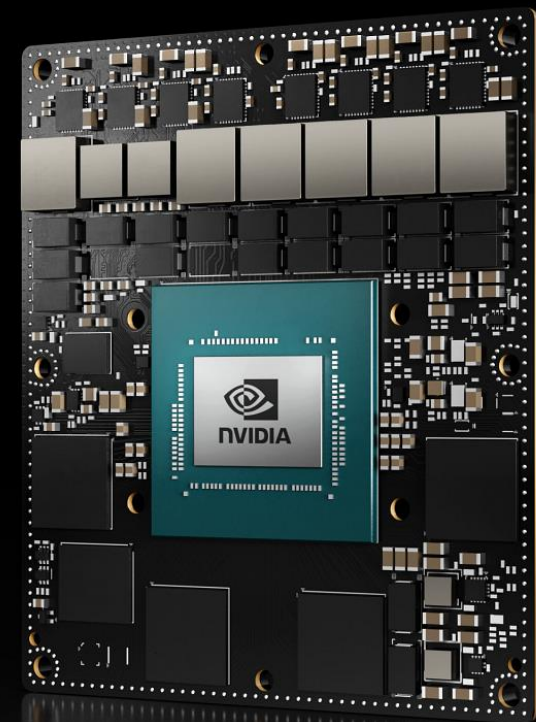
Full specs at developer.nvidia.com/jetson

*TX2i: 10-20W ** Jetson AGX Xavier Industrial 20W-40W

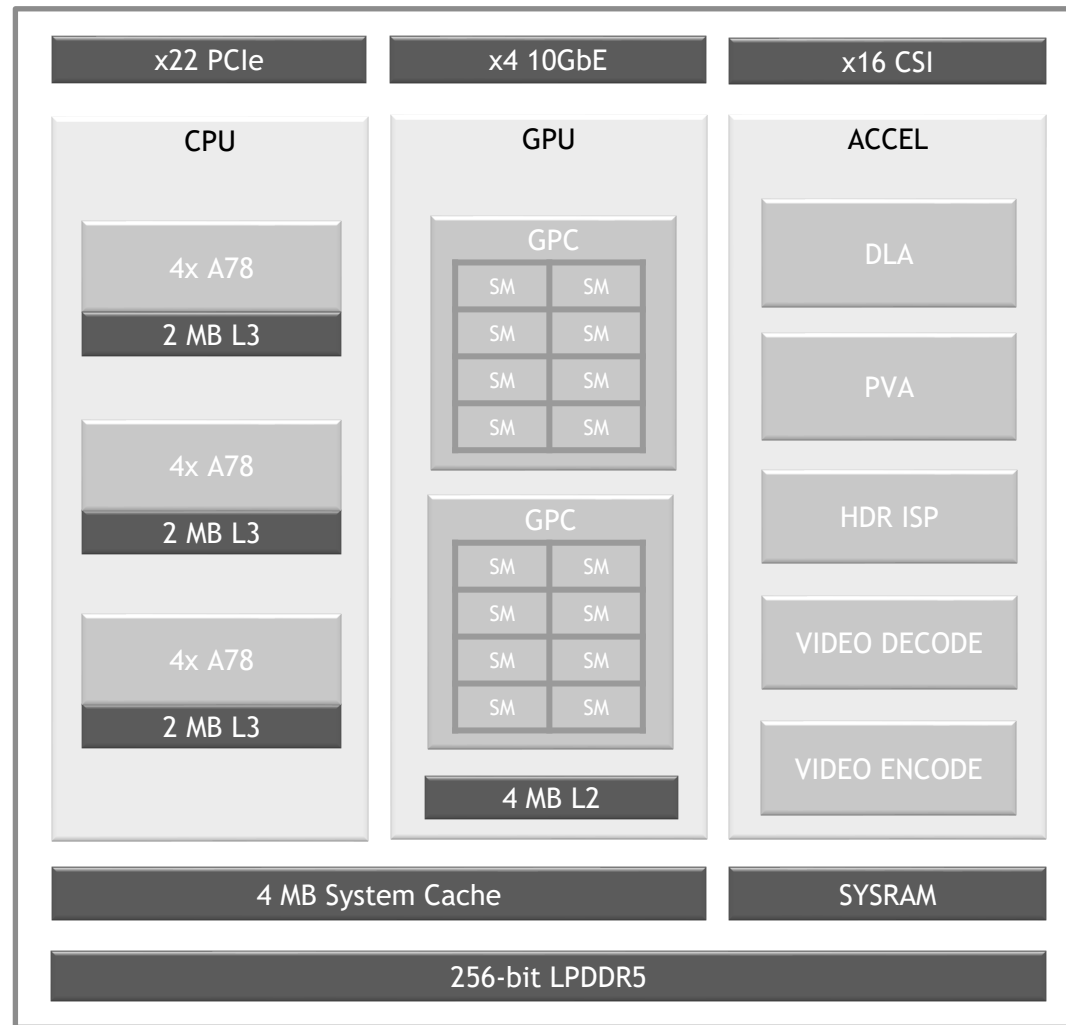
JETSON AGX ORIN 系列

Server Class AI Performance at the Edge

- Up to **275 INT8 TOPS** powered by Ampere GPU +DLA
- Up to 12x A78 ARM CPU
- Up to 64 GB memory, 204 GB/s
- Production Modules will be available as early as **July 2022**
- Get started developing for AGX Orin series with AGX Orin Devkit



JETSON ORIN 系统架构框图



Advanced CPU

- *12x Cortex A78 Cores
ARM Arch V8.2
- *259 SPECint_rate 2006

Next-gen GPU

- *2 GPC / 8 TPC / 16 SMs
- *5.3 FP32 CUDA TFLOPs
- *10.6 FP16 CUDA TFLOPs

Higher DRAM BW

- 256-bit LPDDR5
- 205 GB/s

Rich IO Connectivity

- x4 10 GbE
- x22 PCIe, x16 CSI

DL Performance

- *275 INT8 DL Sparse TOPs
- *138 INT8 DL Dense TOPs

Enhanced PVA

- 512 INT16 GMAC/s
- 2048 INT8 GMAC/s

JETSON AGX ORIN 模组概览

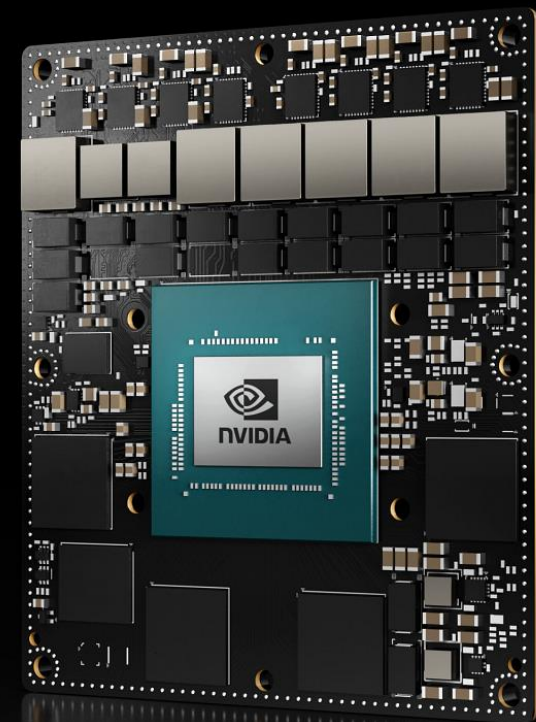
JETSON AGX ORIN 64GB

- 15W to 60W
- **275 INT8 TOPS**
- 12 CPU Cores, 2048 CUDA cores + 64 Tensor Cores, 2 DLAs
- 64 GB memory, 204 GB/s
- \$1599*

JETSON AGX ORIN 32GB

- 15W to 40W
- **200 INT8 TOPS**
- 8 CPU Cores, 1792 CUDA cores + 56 Tensor Cores, 2 DLAs
- 32 GB memory, 204 GB/s
- \$899*

**All prices are based on 1Ku+ Qty.*



JETSON AGX ORIN 模组

	Jetson AGX Xavier	Jetson AGX Xavier 64GB	Jetson AGX Orin 32GB	Jetson AGX Orin 64GB
AI Performance	32 TOPS (De)		200 TOPS (Sp) 100 TOPS (De)	275 TOPS (Sp) 138 TOPS (De)
GPU	512-core NVIDIA Volta GPU with 64 Tensor Cores		1792-core NVIDIA Ampere GPU with 56 Tensor Cores	2048-core NVIDIA Ampere GPU with 64 Tensor Cores
DL Accelerator	2x NVDLA		2x NVDLA v2	
Vision Accelerator	2x PVA v1		PVA v2	
CPU	8-core NVIDIA Carmel Arm®v8.2 64-bit CPU 8MB L2 + 4MB L3		8-core NVIDIA Arm® Cortex A78AE v8.2 64-bit CPU 2MB L2 + 4MB L3	12-core NVIDIA Arm® Cortex A78AE v8.2 64-bit CPU 3MB L2 + 6MB L3
Memory	32GB 256-bit LPDDR4x @ 2133MHz 137 GB/s	64GB 256-bit LPDDR4x @ 2133MHz 137 GB/s	32 GB 256-bit LPDDR5 @ 3200MHz 204.8 GB/s	64 GB 256-bit LPDDR5 @ 3200MHz 204.8 GB/s
Storage	32GB eMMC 5.1		64 GB eMMC 5.1	
Video Encode	4x 4K60 8x 4K30 16x 1080p60 32x 1080p30 (H.265) H.264, VP9		1x 4K60 3x 4K30 6x 1080p60 12x 1080p30 (H.265) H.264, AV1	2x 4K60 4x 4K30 8x 1080p60 16x 1080p30 (H.265) H.264, AV1
Video Decode	2x 8K30 6x 4K60 12x 4K30 26x 1080p60 52x 1080p30 (H.265) H.264, VP9		1x 8K30 2x 4K60 4x 4K30 9x 1080p60 18x 1080p30 (H.265) H.264, VP9, AV1	1x 8K30 3x 4K60 6x 4K30 12x 1080p60 24x 1080p30 (H.265) H.264, VP9, AV1
Camera	16 lanes MIPI CSI-2 (36 Virtual Channels) 8 lanes SLVS-EC D-PHY 40Gbps / C-PHY 62 Gbps		16 lanes MIPI CSI-2 (16 Virtual Channels*) D-PHY 2.1 40Gbps / C-PHY 2.0 164Gbps	
PCI Express	16 lanes PCIe Gen 4 1 x8 + 1 x4 + 1 x2 + 2 x1		22 lanes PCIe Gen 4 Up to 2 x8, 1 x4, 2 x1	
Ethernet	1 Gbe RGMII		1 Gbe RGMII 4x 10Gbe XFI	
Mechanical	100mm x 87mm 699 pin connector		100mm x 87mm 699 pin connector	
Power	10W to 30W		15W to 40W	15W to 60W



*Virtual Channel related camera information for Jetson AGX Orin is not final and subject to change. Video Decode will be finalized by Mid Match

JETSON 模组横向比较



JETSON AGX XAVIER 64GB

- 32 Dense INT8 TOPS
- 10W to 30W
- \$1299*

JETSON AGX XAVIER (32GB)

- 32 Dense INT8 TOPS
- 10W to 30W
- \$899*

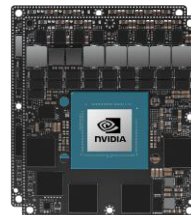


JETSON XAVIER NX 16GB

- 21 Dense INT8 TOPS
- 10W to 20W
- \$499*

JETSON XAVIER NX (8GB)

- 21 Dense INT8 TOPS
- 10W to 20W
- \$399*

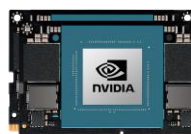


JETSON AGX ORIN 64GB

- 275 SPARSE | 138 DENSE, INT8 TOPS
- 15W to 60W
- \$1599*
- Available: **End Oct 2022**

JETSON AGX ORIN 32GB

- 200 SPARSE | 100 DENSE, INT8 TOPS
- 15W to 40W
- \$899*
- Available: **Mid July 2022**



JETSON ORIN NX 16GB

- 100 SPARSE | 50 DENSE, INT8 TOPS
- 10W to 25W
- \$599*
- Available: **Mid Aug 2022**

JETSON ORIN NX 8GB

- 70 SPARSE | 35 DENSE, INT8 TOPS
- 10W to 20W
- \$399*
- Available: **Mid Nov 2022**

JETSON AGX ORIN 开发套件

Get Started With the AGX Orin Series and the Orin NX Series

EMULATION MODE

- The AGX Orin developer kit can emulate the performance of the AGX Orin Series and the Orin NX Series
- Developers can get started with their development of all four modules today!

DEVELOPER KIT CONTENTS

- Jetson AGX Orin Module and Heatsink
- Reference carrier board
- 802.11ac/abgn Wireless Network Interface Controller
- Power adapter and USB Type-C cord
- USB Type-C to USB Type-A cord
- Quick Start and Support Guide



JETSON AGX ORIN DEVELOPER KIT*

15W - 60W

275 TOPS (INT8)

\$1999

Available Now

110mm x 110mm x 71.65mm

[Jetson AGX Orin for Next-Gen Robotics | NVIDIA](https://developer.nvidia.com/embedded/jetson-agx-orin-developer-kit)
<https://developer.nvidia.com/embedded/jetson-agx-orin-developer-kit>

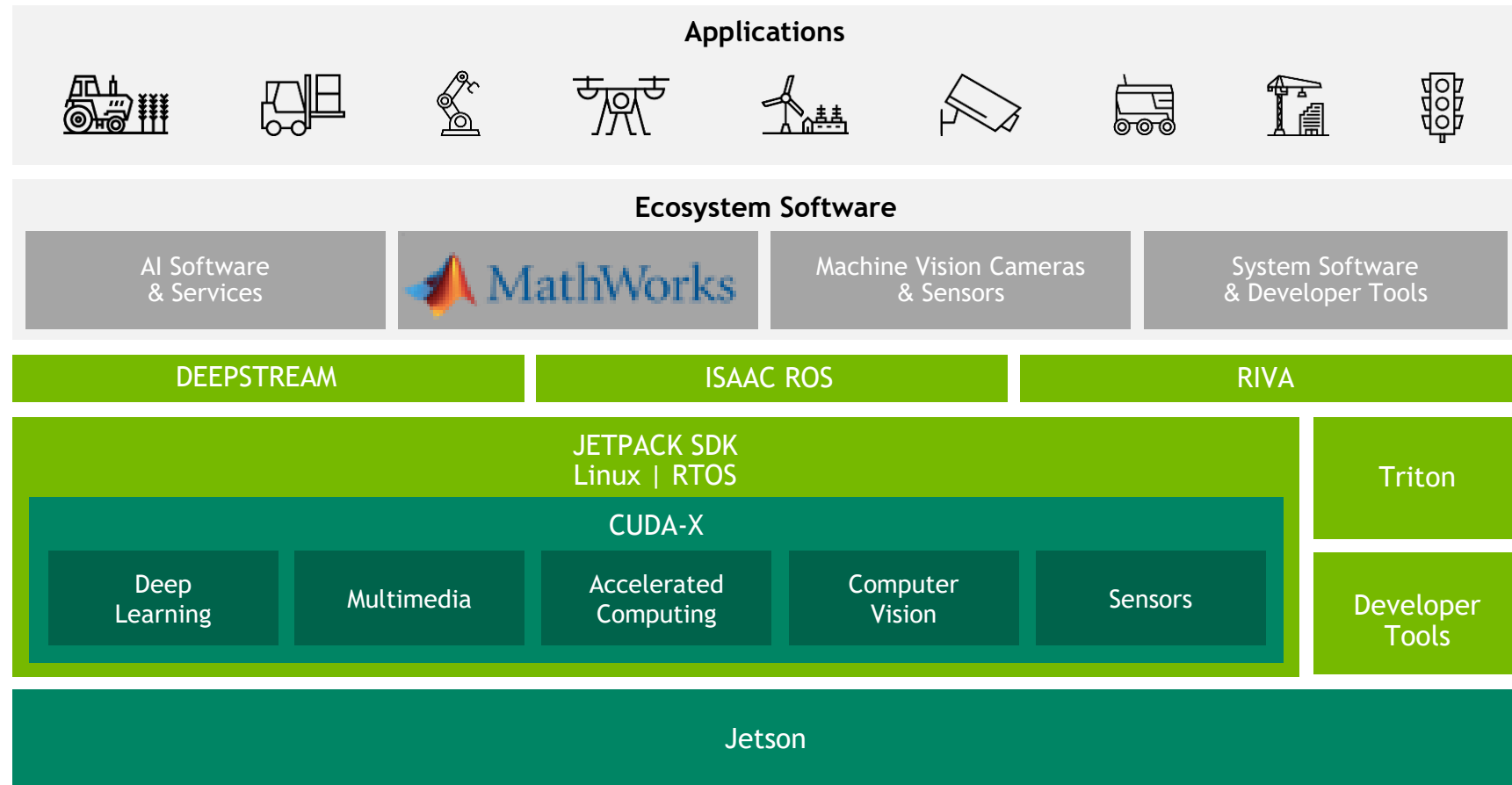
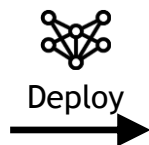
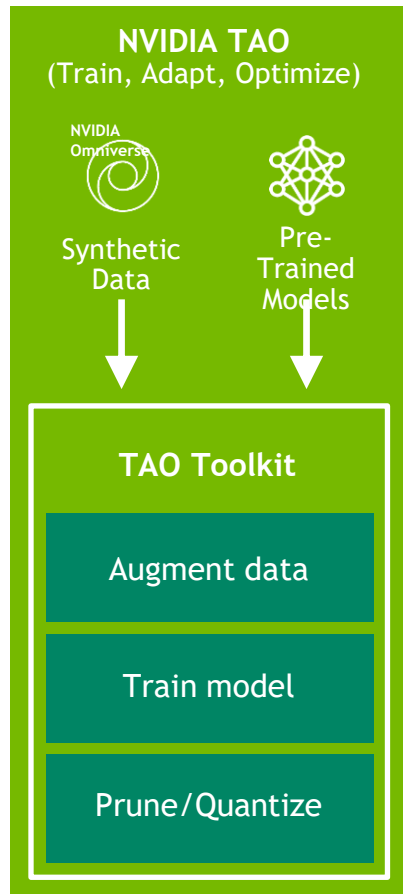
* The Jetson AGX Orin Developer Kit will be available with 32GB of memory.

NVIDIA recommends to use Developer Kits for only development purposes. For production, customers should leverage our production modules. Our production modules have been tested for production environments.



基于 JETSON 的AI软件平台

Accelerate AI Applications and Time-to-Market



POLL # 3

Are you using Jetson today, if so, which is your ideal module?

- a. Jetson AGX Orin
- b. Jetson AGX Xavier
- c. Jetson Orin NX
- d. Jetson Xavier NX
- e. Jetson Nano

实现AI应用嵌入式部署需要解决的问题

- 为AI应用场景选择合适的硬件
- 仿真并测试AI应用
- 将AI应用部署到嵌入式设备



AI驱动系统设计

数据准备

 数据清洗和准备

 领域知识和经验

 仿真生成数据

AI 建模

 模型设计和调试

 硬件加速训练

 框架互操作

仿真和测试

 复杂系统集成

 系统仿真

 系统测试验证

部署

 嵌入式硬件

 企业系统

 边缘、云端、桌面

AI驱动系统设计

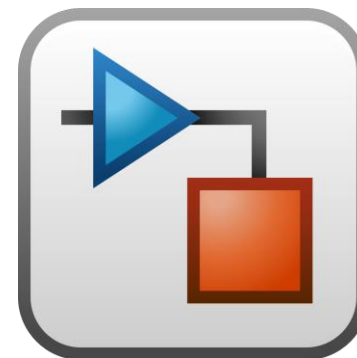
数据准备



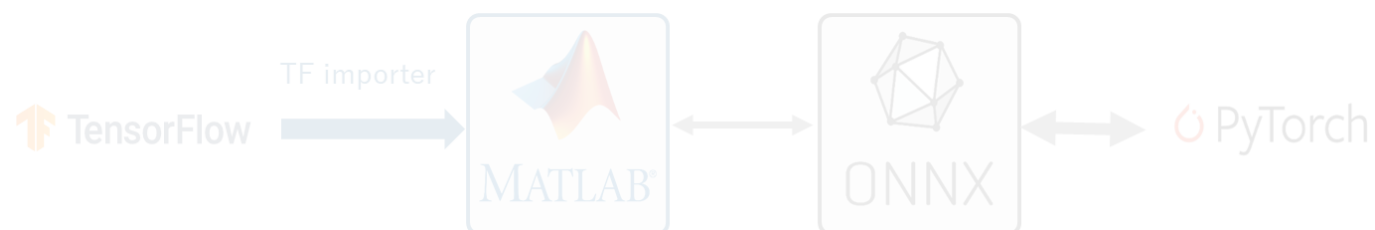
AI 建模



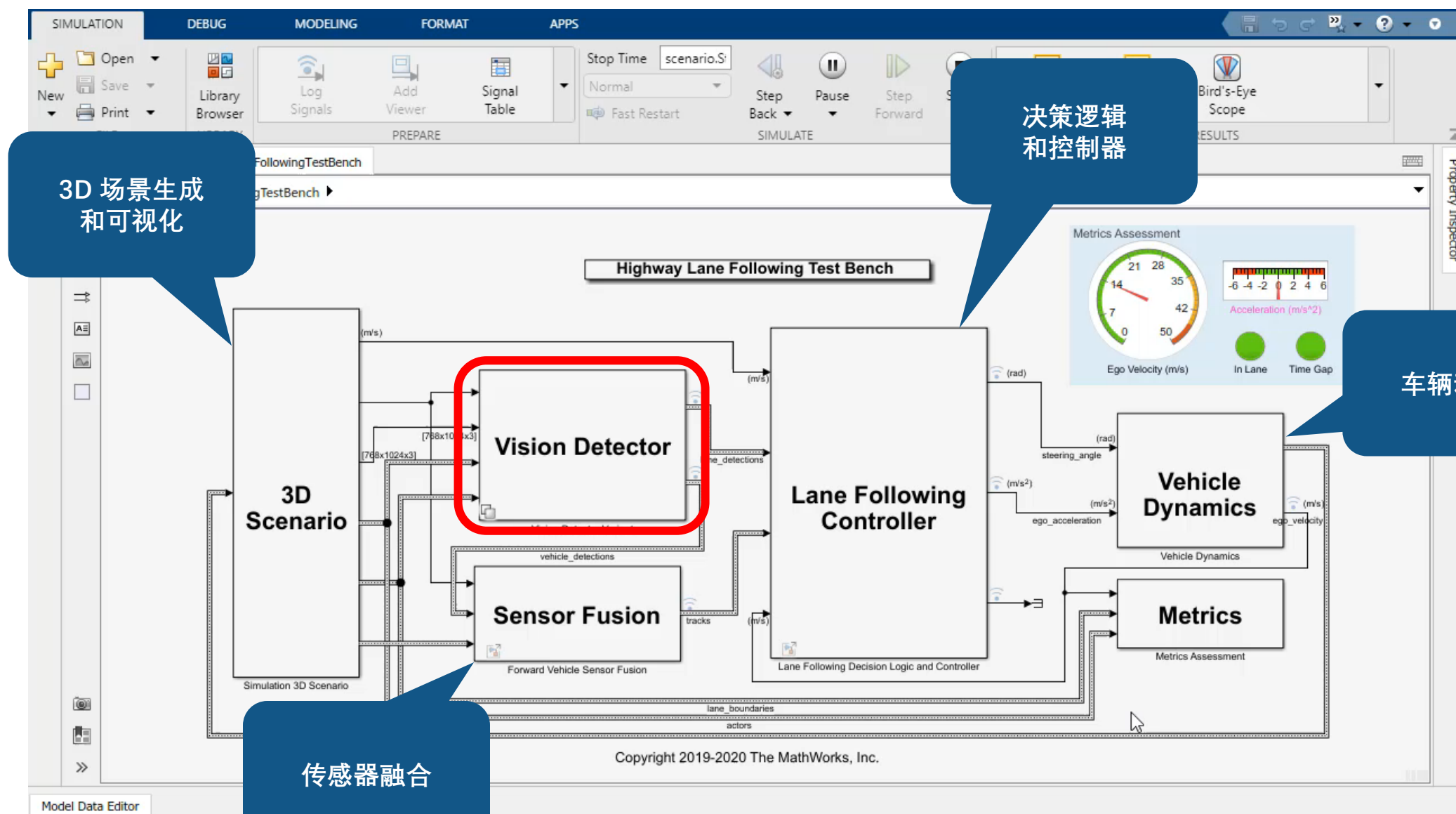
仿真和测试



部署

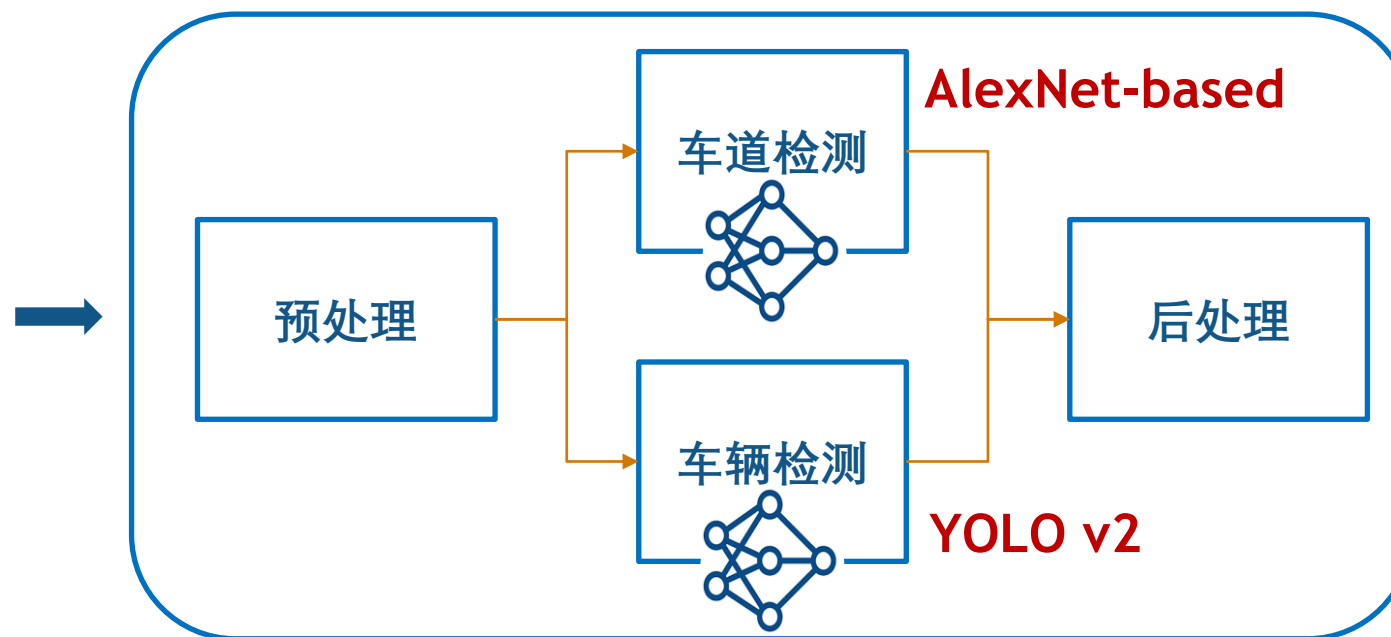


高速公路车道跟随SIMULINK仿真



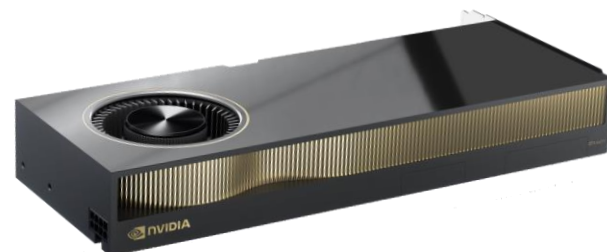
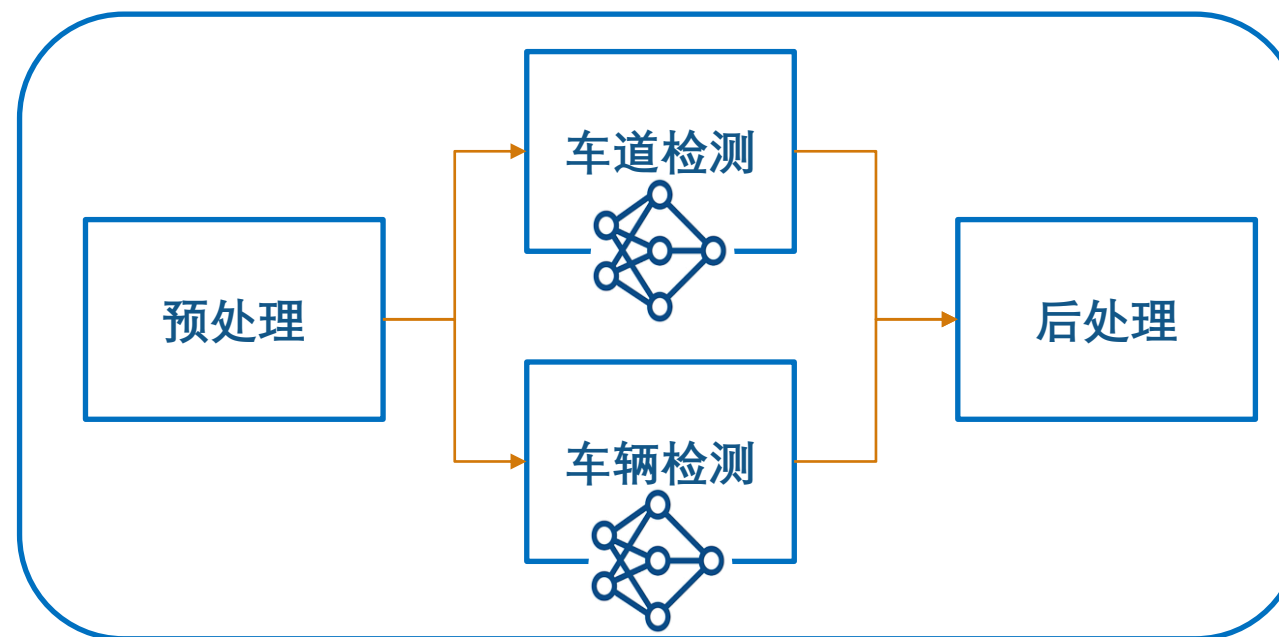


车道和车辆检测

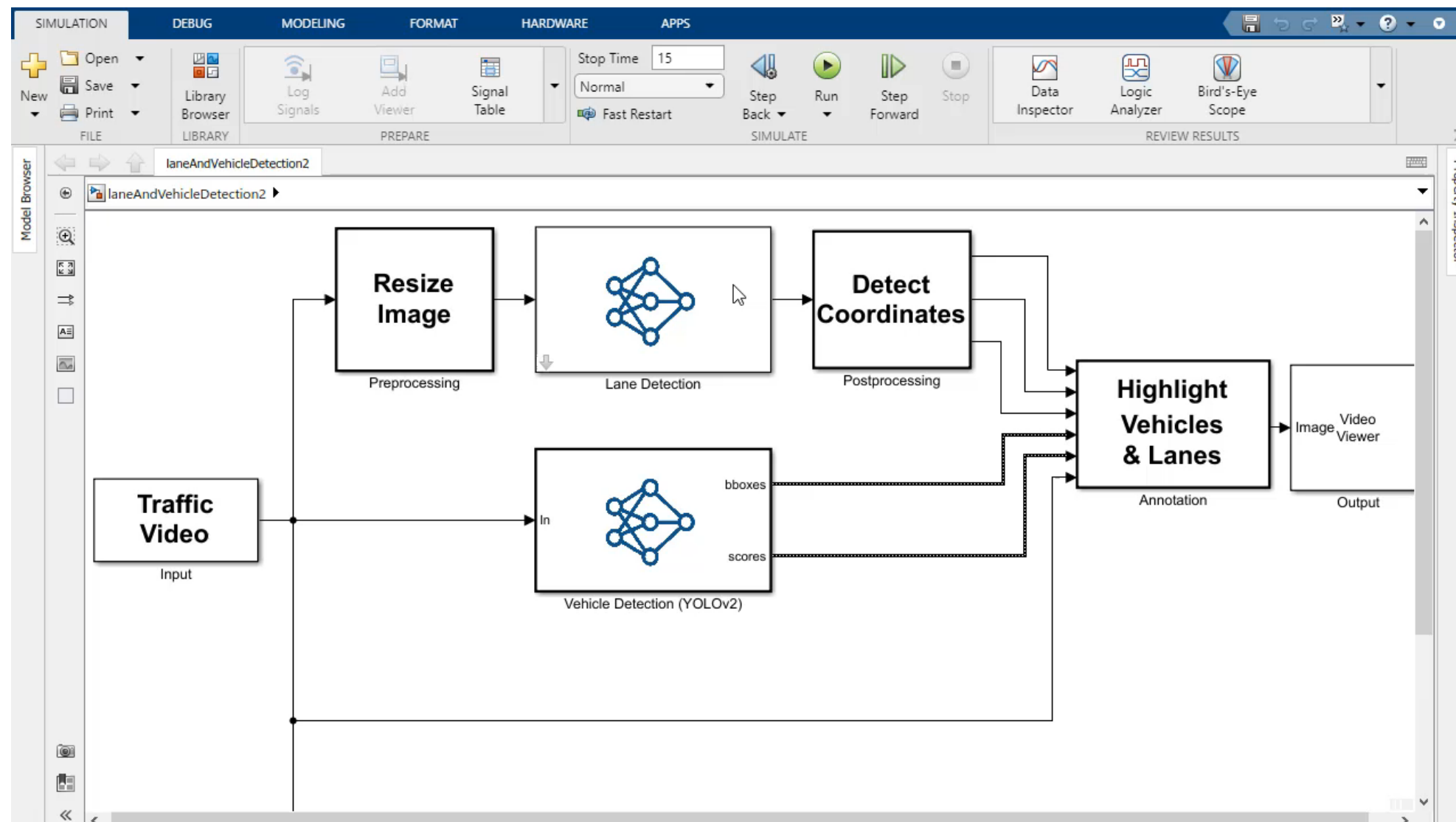




使用桌面GPU运行仿真



使用桌面GPU运行仿真



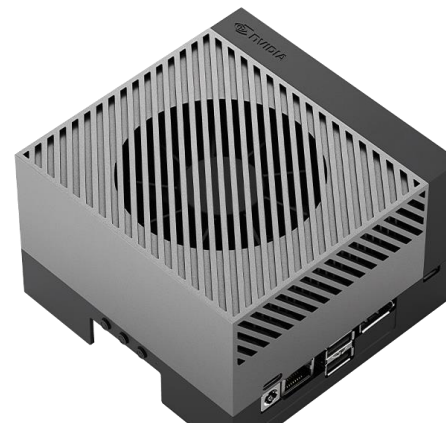
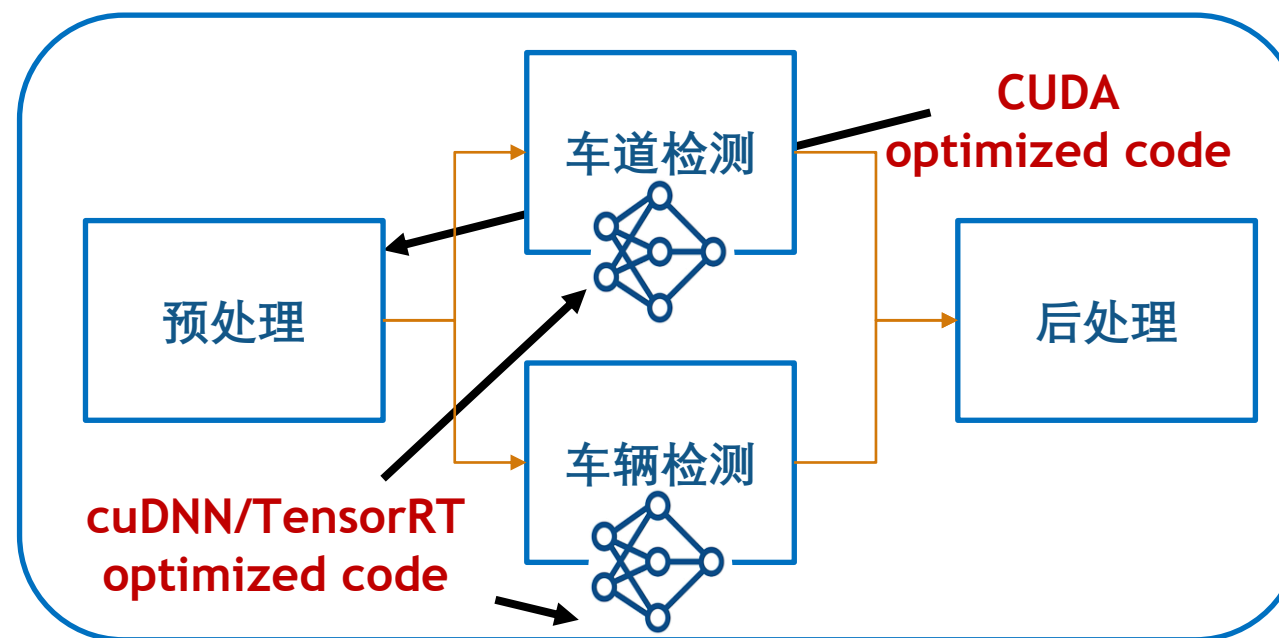
实现AI应用嵌入式部署需要解决的问题

- 为AI应用场景选择合适的硬件
- 仿真并测试AI应用
- 将AI应用部署到嵌入式设备



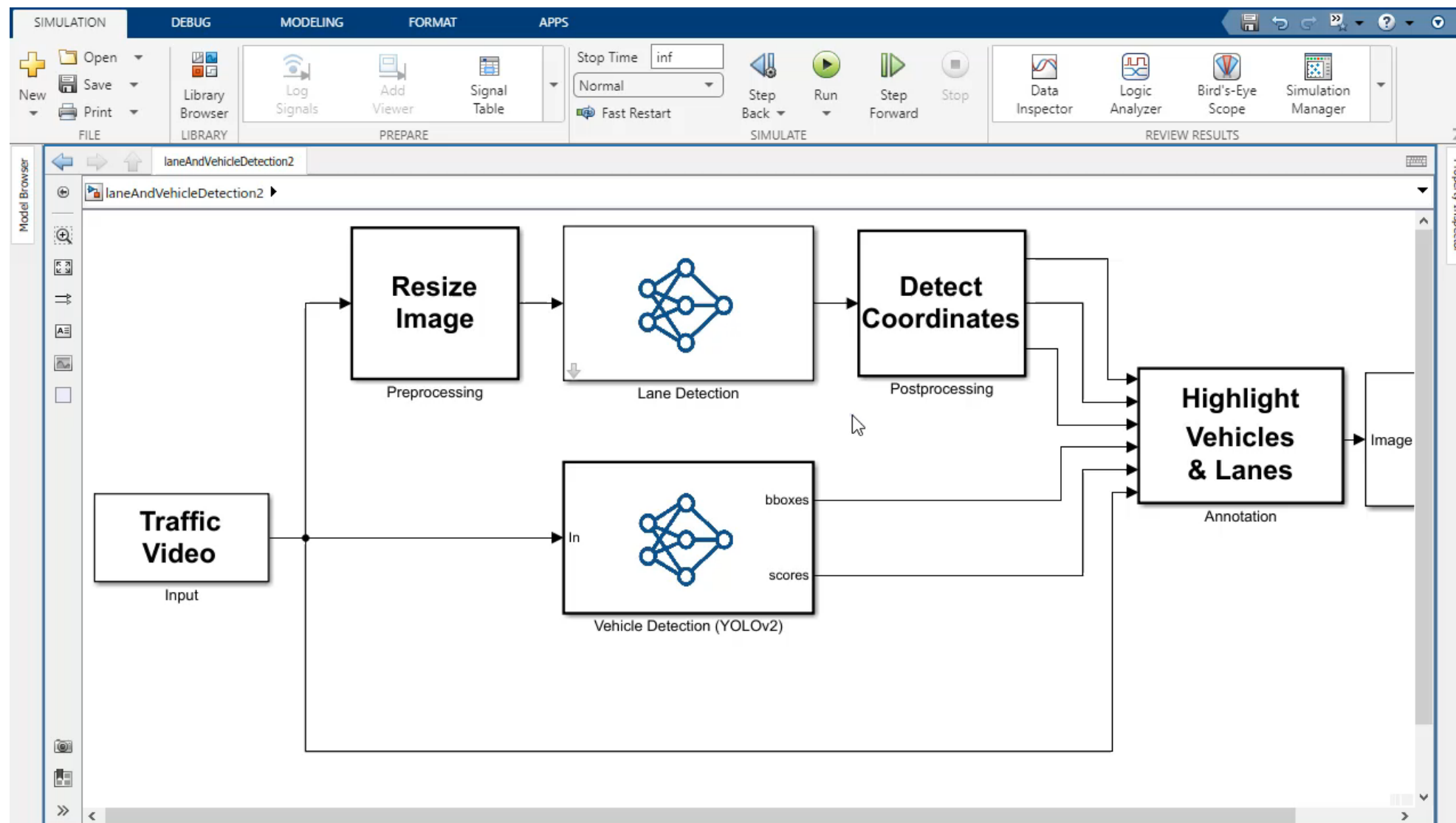


使用 GPU CODER 生成CUDA代码并部署到 JETSON AGX ORIN



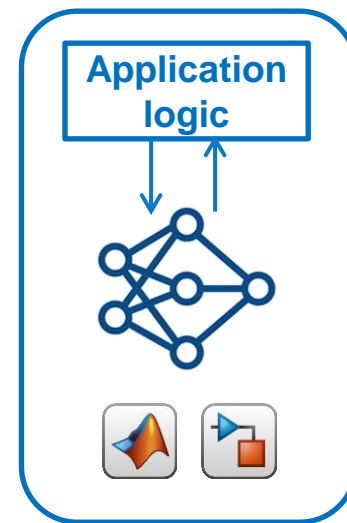


使用 GPU CODER 生成 CUDA 代码并部署到 JETSON AGX ORIN



快速开始创建 AI 应用并使用 SIMULINK 部署到 JETSON AGX ORIN

- 了解 NVIDIA Jetson AGX Orin
 - tinyurl.com/jetson-orin
- 查看详细示例说明:
 - tinyurl.com/gpu-coder-blog
- 快速开始使用 GPU Coder
 - tinyurl.com/gpu-coder



MATLAB EXPO

谢谢

