NVIDIA Jetson
Software-Defined AI Platform

AI AT THE EDGE
Sensor Fusion & Compute Performance

Sense

Act

Reason

SOFTWARE DEFINED
SDK, Design Tools, Libs, GEMs

JetPack SDK ∙ CUDA ∙ TensorRT ∙ TensorFlow ∙ ONNX ∙ ROS

ECOSYSTEM
Expertise, Time to Market

© 2022 NVIDIA
Jetson Software
Accelerates AI Applications and Time-to-Market

**AI MODEL DEVELOPMENT**
Production-ready AI pre-trained models and Toolkits to accelerate development by up to 10X

- Pre-trained models
- TAO Toolkit
- NVIDIA Omniverse
- Synthetic Data Generator

**APPLICATION FRAMEWORKS**
Easy-to-use libraries that support the development of applications

- DeepStream
- Isaac
- Riva

**JETPACK SDK**
Comprehensive SDK for building end-to-end accelerated AI pipeline

JetPack SDK
Linux | RTOS

CUDA-X
- Deep Learning
- Multimedia
- Accelerated Computing
- Computer Vision
- Sensors

© 2022 NVIDIA
CV Is More than AI Inference
DeepStream SDK
Accelerated and Optimized Applications from Edge to Cloud

Sensors
Capture and Decode
Pre-processing and Batching
AI Inference
Tracking
Business Rules and Analytics
Composition
Actions

Hardware Acceleration
ISP
GPU
GigE
DLA
GigE
NVDEC
GPU
GPU
CPU
CPU
VIC
GPU

Edge to Cloud Deployments
Jetson Appliances
EGX Servers
Cloud

© 2022 NVIDIA
Pipeline Efficiency with 0 Memory Copies

Capture → Decode → Batching → Inference → ... → Next Plugin

1. CPU Memory
2. Allocated CUDA HW Buffers (GPU Memory)
3. Uses shared CUDA HW Buffers (GPU Memory)

Rivermax (Optional)

CPU to GPU Memory

CUDA HW Buffer Sharing

Returns CUDA HW Buffer for Reuse

Zero Buffer Copy Path
DeepStream Features

Graph Composer
Develop production ready apps

New Models Ready to Use
Action Recognition, Pose, Gaze, ASR, ...

New Unified Tracker Architecture

State-of-the-art Trackers
Bring Your Own Tracker

DeepStream SDK

DeepStream SDK

Triton Inference Server

gRPC Support for Triton
Build apps that leverage multiple frameworks

DeepStream Features

From NIC to GPU Memory

Rivermax Support
Optimize uncompressed data pipelines

Any Cloud All Your Metadata

IOT and Cloud Connectivity
Redis, Azure Percept and many others

© 2022 NVIDIA
Developing with DeepStream

Building Blocks

- DeepStream Developers
  - Python Bindings (source code available now)
  - Hardware Accelerated DeepStream Plugins
  - Plugins to Graph Composer Extensions

Application Development

- Develop in Python
  - Python
  - Develop in C/C++
  - DeepStream Applications

Application Deployment

- Container Builder
  - Cloud
  - EGX Servers
  - Jetson Appliances
  - Cloud Native Deployment

© 2022 NVIDIA
Why Graph Composer?
Benefits of Low-Code Development

Graph Composer for AI Beginners

- Makes NVIDIA Vision AI technology more accessible to developers and organizations
- Makes it easier for beginners to get started with DeepStream on EGX or Jetson
- No prior Gstreamer/DeepStream knowledge required

Graph Composer for Developers

- Allows developers to focus on the application’s functionality and optimization
- Provides a platform for quick evaluation and PoC tasks
- Makes it easier to create, integrate, and share custom functionality using extensions
Graph Composer Workflow Steps

1. Sync extension from NVIDIA Cloud repository

2. Develop custom extension
   - Generate extension from GStreamer plugin (Automatic)
   - Generate extension from codelets (Manual/Advanced)

3. Build and extensions to registry

4. Create and test graph

5. Create deployment container
   - Generate containers for both x86 (dGPU) and Arm64 (Jetson)
Graph Composer Demo
Easier and Faster Deployment
- Eliminates complex, time-consuming builds and installs

Agile and Easier Development
- Update specific modules not the whole system

Scalability
- Push the right container to the right platform

Portable
- Deploy across various environments, from test to production with minimal changes

Consistency Across Portfolio
- Update once and push broadly
Call to Action

Developer Resources

DeepStream SDK
https://developer.nvidia.com/deepstream-sdk

NVIDIA Deep Learning Institute

Jetson Module Family

2022 Embedded Vision Summit

• NVIDIA - Booth 806

• NVIDIA Partner Booths
  • 324 - deci.
  • 318 - nX NetworkOptix
  • 417 - vision components®
  • 618 - AVW
  • 609 -
  • 608 -
  • 709 -
  • 604 - BASLER
  • 607 - e-con Systems
  • 523 - SEQUITUR LABS

© 2022 NVIDIA
Questions and Answer