deti universidade de aveiro
departamento de eletrónica,
telecomunicações e informática

Nvidia Jetson Nano

Architecture, Hardware and Software aspects

TP1 - Grupo 7
Lúcia Sousa 93086
Raquel Pinto 92948

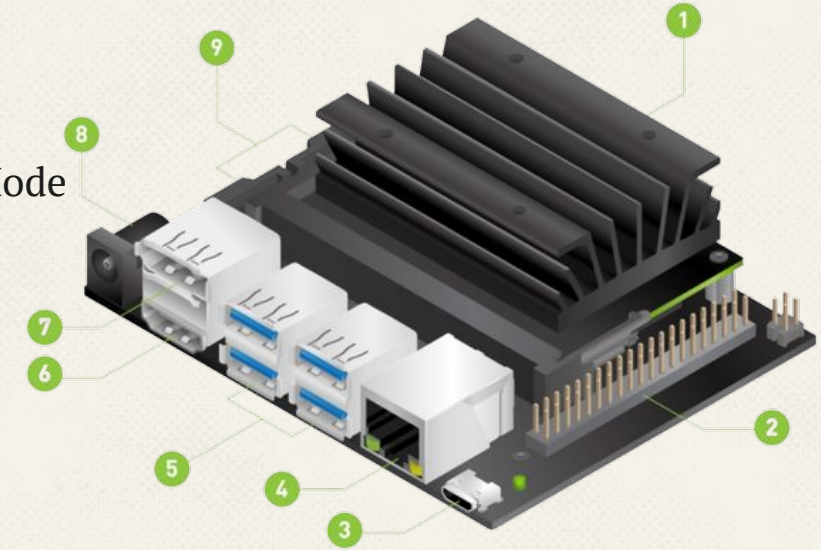
31/05/2022 ASE - Arquitetura de Sistemas Embutidos

Nvidia Jetson Nano

NVIDIA Jetson Nano is a small, powerful computer that lets you run multiple neural networks in parallel for applications like image classification, object detection, segmentation and speech processing.

COMPONENTS

1 microSD card slot for main storage
2 40-pin expansion header
3 Micro-USB port for 5V power input, or for Device Mode
4 Gigabit Ethernet port
5 4 USB 3.0 ports
6 HDMI output port
7 DisplayPort connector
8 DC Barrel jack for 5V power input
9 MIPI CSI-2 camera connectors



Specifications

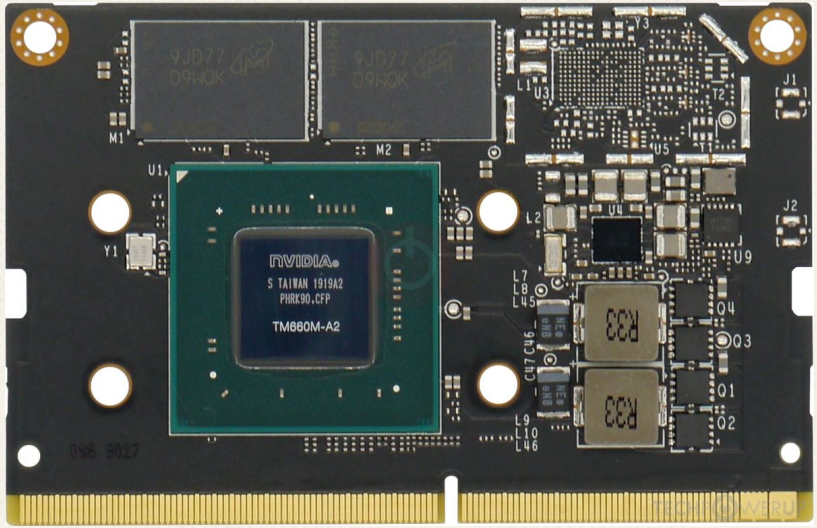
GPU	128-core Maxwell
CPU	Quad-core ARM A57 at 1.43 GHz
Memory	2 GB 64-bit LPDDR4 25.6 GB/s
Storage	microSD
Video Encode	4K at 30 | 4x 1080p at 30 | 9x 720p at 30 (H.264/H.265)
Video Decode	4K at 60 | 2x 4K at 30 | 8x 1080p at 30 | 18x 720p at 30 (H.264/H.265)
Camera	2x MIPI CSI-2 DPHY lanes
Connectivity	Gigabit Ethernet
Display	HDMI and display port
USB	4x USB 3.0, USB 2.0 Micro-B
Others	Low-bandwidth peripheral controllers: I2C, I2S, SPI, UART, PWM

Maxwell GPU

- GPU’s core graphics functions are performed inside de GPC.
- GPC (Graphics Processing Cluster) is a dedicated hardware block for rasterization, shading, texturing and compute.
- Within the GPC are multiple SM (Streaming Multiprocessor) units and a Raster Engine.
- Each SM includes a Polymorph Engine and Texture Units.
- The SM scheduler architecture and algorithms were rewritten to be more intelligent and avoid unnecessary stalls, while further reducing the energy per instruction required for scheduling.
- The organization of the SM also changed; each Maxwell SM (SMM) is now partitioned into four separate processing blocks, each with its own instruction buffer, scheduler and 32 CUDA cores.

GPU - Features

- End-to-end lossless compression
- Tile Caching
- Support for OpenGL 4.6, OpenGL ES 3.2, Vulkan 1.1, DirectX 12, CUDA 10 (FP16)
- Adaptive Scalable Texture Compression (ATSC) LDR profile supported
- Iterated blend, ROP OpenGL-ES blend modes
- 2D color compression
- Constant color render SM bypass
- 2x, 4x, 8x MSAA with color and Z compression

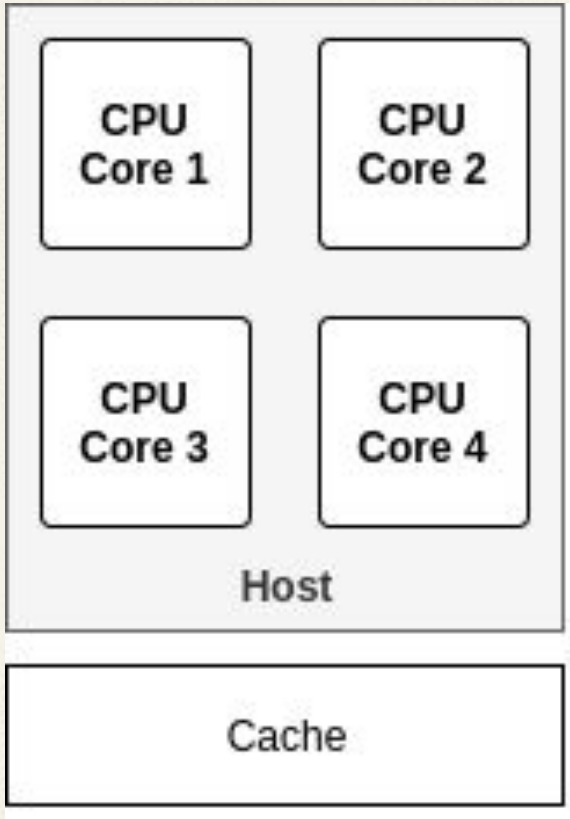


GPU - Features

- FP16 texture filtering
- FP16 shader support
- Geometry and Vertex attribute Instancing
- Parallel pixel processing
- Early-z reject: Fast rejection of occluded pixels acts as multiplier on pixel shader and texture performance while saving
- power and bandwidth
- Video protection region
- Power saving: Multiple levels of clock gating for linear scaling of power

CPU Complex

- Multi-Core SMP cluster of four ARM Cortex-A57 with 2MB of L2 cache (shared by all cores).
- Features:
  - Superscalar, variable-length, out-of-order pipeline
  - Dynamic branch prediction with Branch Target Buffer (BTB) and Global History Buffer RAMs, a return stack, and an indirect predictor
  - 48-entry fully-associative L1 instruction TLB with native support for 4KB, 64KB, and 1MB page sizes.
  - 32-entry fully-associative L1 data TLB with native support for 4KB, 64KB, and 1MB pages sizes.
  - 4-way set-associative unified 1024-entry Level 2 (L2) TLB in each processor



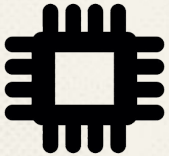
CPU Complex

- Features:
  - 48Kbyte I-cache and 32Kbyte D-cache for each core.
  - Full implementation of ARMv8 architecture instruction set
  - Embedded Trace Microcell (ETM) based on the ETMv4 architecture
  - Performance Monitor Unit (PMU) based on the PMUv3 architecture
  - Cross Trigger Interface (CTI) for multiprocessor debugging
  - Cryptographic Engine for crypto function support
  - Interface to an external Generic Interrupt Controller (vGIC-400)
  - Power management with multiple power domains

SCU and L2 Cache

- The CPU cluster includes an integrated snoop control unit (SCU) that maintains coherency between the CPUs within the cluster and a tightly coupled L2 cache.
- The L2 cache also provides a 128-bit AXI master interface to access DRAM.
- L2 cache features include:
  - 2MB L2
  - Fixed line length of 64 bytes
  - 16-way set-associative cache structure
  - Duplicate copies of the L1 data cache directories for coherency support
  - ECC (error-correcting code) support

Memory



- Integrates 2GB of LPDDR4 over a four-channel x 16-bit interface.
- The memory frequency can be 204 MHz or 1600 MHz;
- Memory Controller (MC) maximizes memory utilization while providing minimum latency access for critical CPU requests providing access to main memory for all internal devices.
- An arbiter is used to prioritise requests (increases efficiency when accessing memory and minimises system power).

Key Features

- High-Definition Audio-Video Subsystem
  - Multi-Standard Video Decoder and Encoder
- JPEG Processing Block
- Video Image Compositor (VIC)
- Image Signal Processor (ISP)
- Display Controller Complex

High-Definition Audio-Video Subsystem

- The audio-video subsystem off-loads audio and video processing activities from the CPU subsystem resulting in faster, fully concurrent, highly efficient operation.
- **Multi-Standard Video Decoder**
  - Supporting low resolution content, Standard Definition (SD), High Definition (HD) and UltraHD (2160p, or 4k video) profiles.
  - The video decoder communicates with the memory controller through the video DMA which supports a variety of memory format output options.
  - H.265, WEBM, H.264, VC-1, MPEG-4, H.263, DiVX, XviD, MPEG-2

High-Definition Audio-Video Subsystem

- **Multi-Standard Video Encoder**
  - The multi-standard video encoder enables full hardware acceleration of various encoding standards.
  - It performs high-quality video encoding operations for applications such as video recording and video conferencing.
  - H.265, H.264, VP8, MPEG4, MPEG2, VC1

JPEG Processing Block

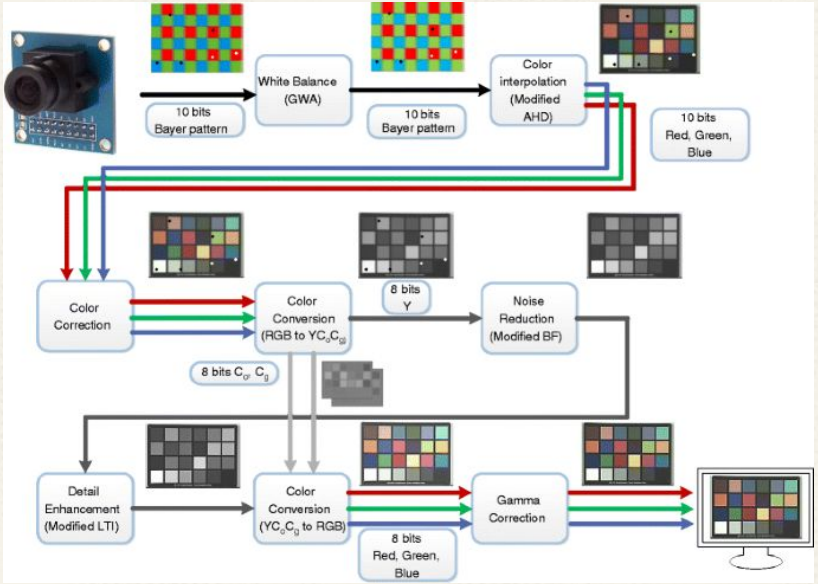
- The JPEG processing block is responsible for JPEG (de)compression calculations (based on JPEG still image standard), image scaling, decoding (YUV420, YUV422H/V, YUV444, YUV400) and color space conversion (RGB to YUV; decode only).

Video Image Compositor

- The Video Image Compositor implements various 2D image and video operations in a power-efficient manner.
- It handles various system UI scaling, blending and rotation operations, video post-processing functions needed during video playback, and advanced de-noising functions used for camera capture.
- Features
  - Color Decompression, High-quality Deinterlacing, Inverse Teleciné, Temporal Noise Reduction, High-quality video playback, Reduces camera sensor noise, Scaling, Color Conversion, Memory Format Conversion, Blend/Composite, 2D Bit BLIT operation, Rotation

Image Signal Processor

- The ISP module takes data from the VI/CSI module or memory in raw Bayer format and processes it to YUV output.
- Features
  - Bayer domain hardware noise reduction
  - Per-channel black-level compensation
  - High-order lens-shading compensation
  - 3 x 3 color transform
  - Bad pixel correction
  - Color Artifact Reduction
  - Color and gamma correction
  - Color-space conversion (RGB to YUV)
  - Image statistics gathering (per-channel)



Display Controller Complex

- The Display Controller Complex integrates two independent display controllers.
- Each display controller is capable of interfacing to an external display device and can drive the same or different display contents at different resolutions and refresh rates.
- Each controller supports a cursor and three windows (Window A, B, and C).
- Controller A supports two additional simple windows (Window D, T).
- The display controller reads rendered graphics or video frame buffers in memory, blends them and sends them to the display.

Key Features

Power Supplies

5V power supply capable of supplying 2A current via the Micro-USB connector

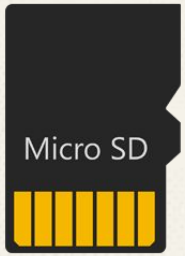
Memory Devices

Memory

2 GB 64-bit LPDDR4, 1600MHz 25.6 GB/s

Storage

MicroSD slot

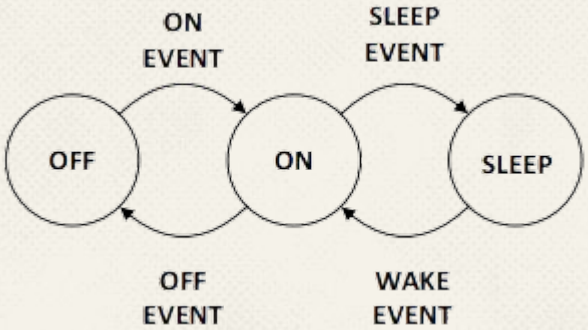


Power and System Management

- Power Management Controller (PMC) and Real Time Clock (RTC)
- Power Gating
- Clock Gating
- Dynamic Voltage and Frequency Scaling (DVFS)

PMC

System power states and transitions



PMC and RTC

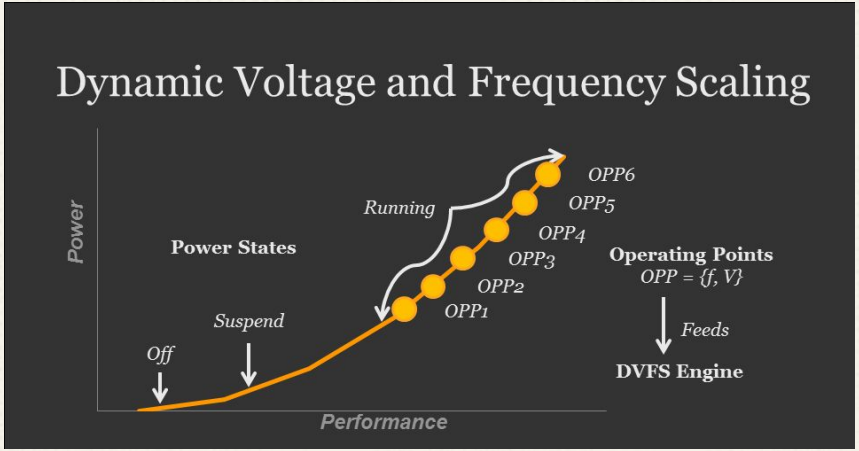
- These blocks reside in an Always On (not power gated) partition. The PMC provides an interface to an external power manager IC or PMU.
- It primarily controls voltage transitions for the SoC as it transitions to/from different low power modes; it also acts as a slave receiving dedicated power/clock request signals as well as wake events from various sources (e.g., SPI, I2C, RTC, USB attach) which can wake the system from a deep-sleep state. The RTC maintains the ability to wake the system based on either a timer event or an external trigger.

Power Gating and Clock Gating

- **Power Gating**
  - The SoC uses power-gating (controlled by PMC) to power-off modules which are idle.
  - CPU cores are on a separate power rail to allow complete removal of power and eliminate leakage.
  - Each CPU can be power gated independently. Software provides context save/restore to/from DRAM.
- **Clock Gating**
  - Used to reduce dynamic power in a variety of power states.

DVFS

- DVFS is used to change the voltage and frequencies in the following power domains: CPU and GPU.
- Raises voltages and clock frequencies when demand requires, lowers them when less is sufficient, and removes them when none is needed.



Peripherals

- Can be used as software controlled input, output and interrupt.
- GPIO pins can be switched (multiplexed) into various other modes backed by dedicated peripheral blocks such as I2C, I2S, UART, SPI, and PWM.
- All pins that can support GPIO functionality have this exposed in the Pinmux (wake up the pins that are in sleep mode).

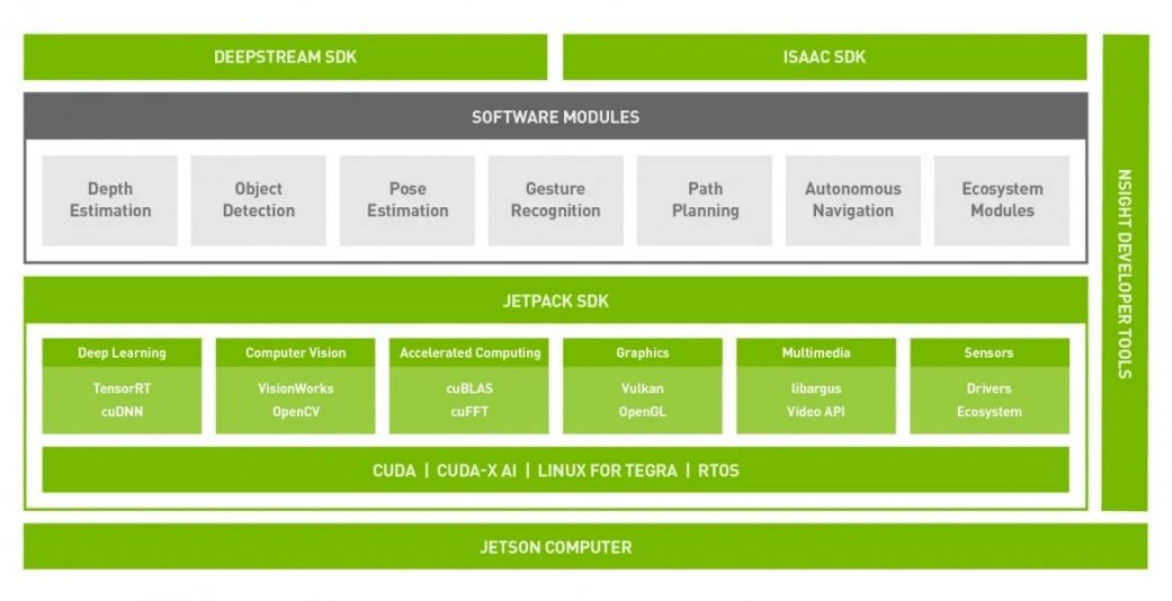
Sysfs	Name	Pin	Pin	Name	Sysfs
	3.3V DC	1	2	5V DC	
	I2C_2_SDA	3	4	5V DC	
	I2C_2_SCL	5	6	GND	
GPIO216	AUDIO_MCLK	7	8	UART_2_TX	
	GND	9	10	UART_2_RX	
GPIO50	UART_2_RTS	11	12	I2S_4_CLK	GPIO79
GPIO14	SPI_2_SCK	13	14	GND	
GPIO194	LCD_TE	15	16	SPI_2_CS1	GPIO232
	3.3VDC	17	18	SPI_2_CS0	GPIO15
GPIO16	SPI_1_MOSI	19	20	GND	
GPIO17	SPI_1_MISO	21	22	SPI_2_MISO	GPIO13
GPIO18	SPI_1_SCK	23	24	SPI_2_CS0	GPIO19
	GND	25	26	SPI_2_CS1	GPIO20
	IC2_1_SDA	27	28	I2C_1_SCL	
GPIO149	CAM_AF_EN	29	30	GND	
GPIO200	GPIO_PZO	31	32	LCD_BL_PWM	GPIO168
GPIO38	GPIO_PE6	33	34	GND	
GPIO76	I2S_4_LRCK	35	36	UART_2_CTSI	GPIO51
GPIO12	SPI_2_MOSI	37	38	I2S_4_SDIN	GPIO77
	GND	39	40	I2S_4_SDOUT	GPIO78

Operating systems

- **Linux4Tegra**
  - Is the board support package for Jetson (NVIDIA JetPack SDK)
  - Based on Ubuntu 18.04
  - Linux Kernel 4.9
  - Bootloader
  - NVIDIA drivers
  - flashing utilities

NVIDIA JetPack SDK

- JetPack SDK provides a full development environment for hardware-accelerated AI-at-the-edge development.
- Includes Jetson Linux Driver Package with bootloader, Linux kernel, Ubuntu desktop environment, and a complete set of libraries for acceleration of GPU computing, multimedia, graphics, computer vision, security features, over-the-air update capabilities.



NVIDIA JetPack SDK

CUDNN:

- Deep neural network library in CUDA (Used by several deep learning frameworks - MATLAB).

VisionWorks and OpenCV:

- Software library that provides Computer Vision / Image Processing algorithms.

Multimedia API:

- Video encoding/decoding.
- Control of camera parameters every frame.

Nsight Developer Tools (NVIDIA Nsight Systems):

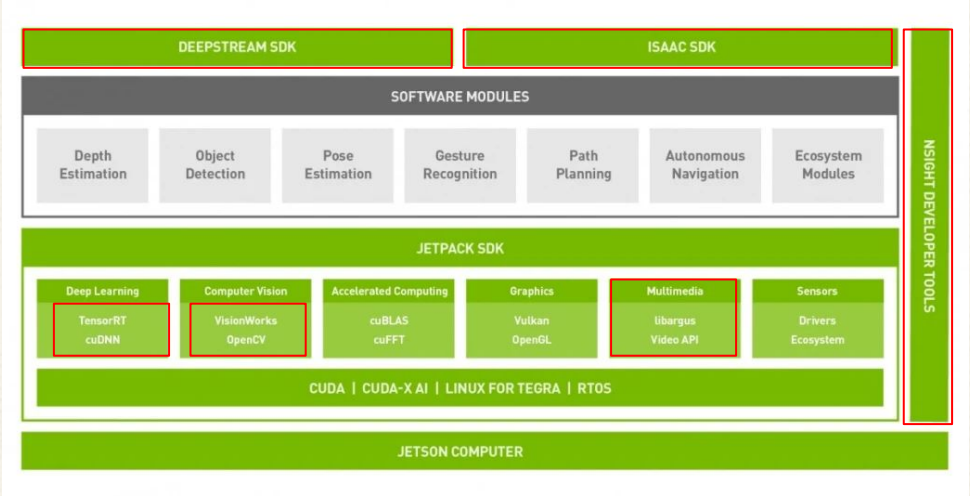
- Allows you to optimise software performance.

ISAAC SDK:

- Robotics in health, agriculture, industry …

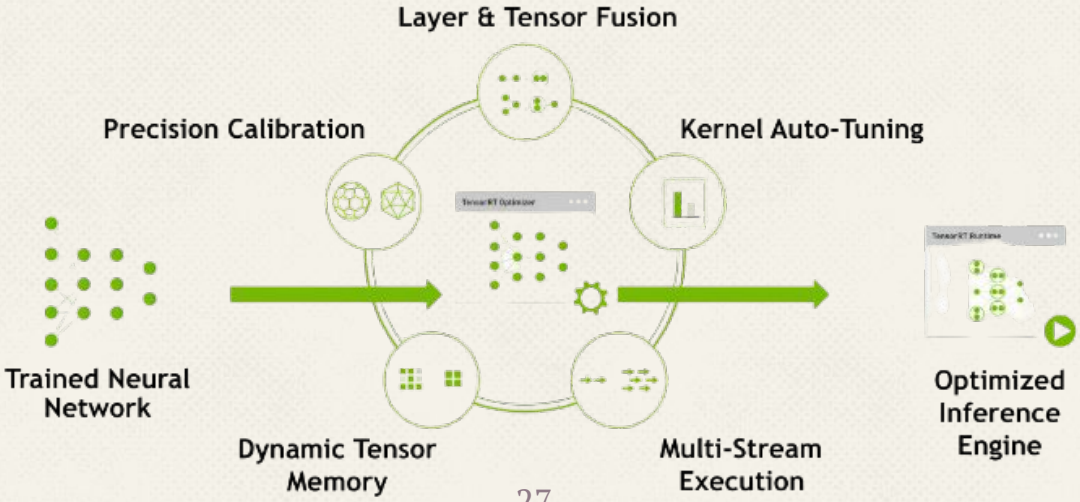
Deepstream SDK:

- Analyse data from cameras, sensors and IoT gateways in real-time.



JetPack SDK : TENSOR RT

- Is a high performance deep learning inference for image classification, segmentation, and object detection neural networks.
- Is built on CUDA, NVIDIA's parallel programming model, and enables you to optimize inference for all deep learning frameworks.
- Includes an optimiser that provides low latency and high performance for applications.



Languages and development tools

- Programming languages: Python, C, C++
- IDEs: Visual Studio Code, Eclipse
- The Nano is capable of running **CUDA**, NVIDIA's programming language for general purpose computing on graphics processor units (GPUs).

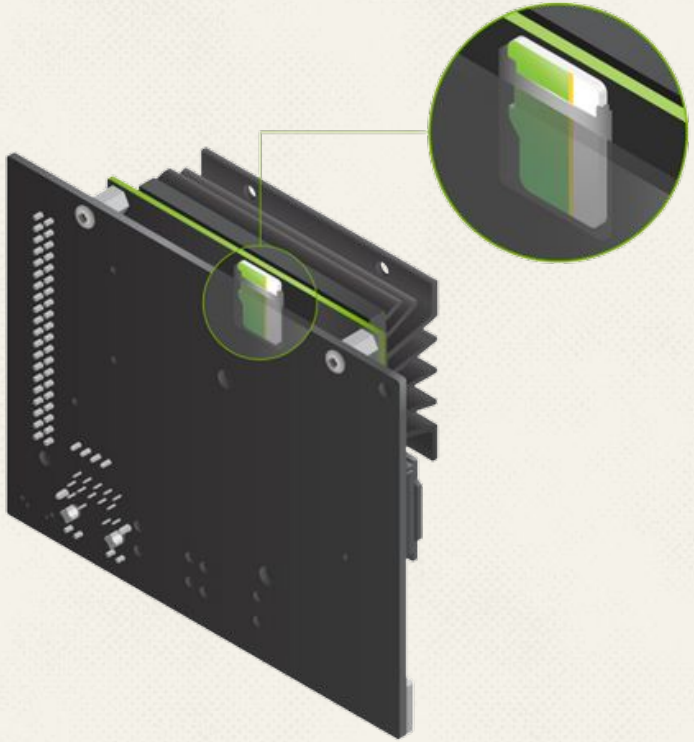
Applications

- Low cost PC/tablet/laptop
- IoT applications
- Media center
- Robotics
- Industrial/Home automation
- Cloud server
- Print server
- Security monitoring
- Web camera
- Wireless access point
- Environmental sensing/monitoring (e.g. WEATHER STATION)
- Machining Learning and AI applications.

Jetson Nano used in robotics

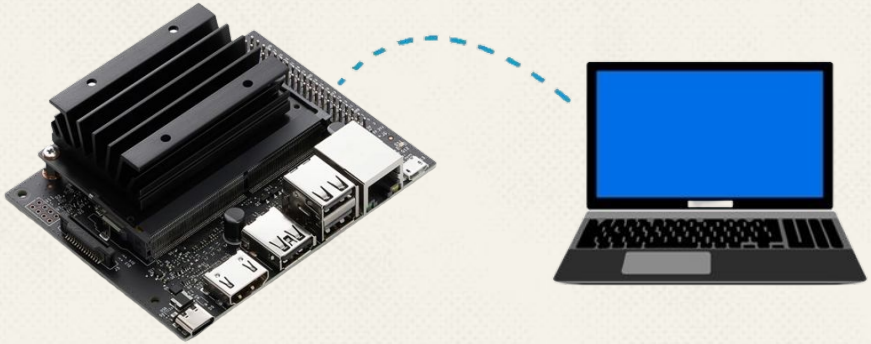
Card Images

- To prepare your microSD card, it is necessary a computer with Internet connection and the ability to read and write SD cards.
- Download the Jetson Nano Developer Kit SD Card Image.
- It is possible to write the SD card image using a graphical program like Etcher, or via command line.



Remote Access and Upgrades

- You can connect a monitor, a keyboard and a mouse to Jetson Nano to work.
- Sometimes you need to access a Jetson Nano without connecting it to this.
- It is possible to connect to the Jetson Nano from another machine. To do this, it requires knowing the IP address.
- It is possible to connect to the Jetson Nano from another machine using SSH or VNC having the IP address of the board.



Bibliography

- https://developer.nvidia.com/embedded/learn/get-started-jetson-nano-devkit
- https://developer.download.nvidia.com/assets/embedded/secure/jetson/Nano/docs/JetsonNano_DataSheet_DS09366001v1.0.pdf
- https://developer.nvidia.com/blog/jetson-nano-ai-computing/
- https://developer.download.nvidia.com/embedded/L4T/r32-3-1_Release_v1.0/Jetson_Nano_Developer_Kit_User_Guide.pdf
- https://components101.com/development-boards/nvidia-jetson-nano-developer-kit
- https://developer.nvidia.com/embedded/linux-tegra-r3251
- https://sbesc.lisha.ufsc.br/sbesc2019/tiki-download_file.php?fileId=2527
- https://developer.nvidia.com/embedded/jetpack
- https://www.digikey.com/en/maker/projects/getting-started-with-the-nvidia-jetson-nano-part-1-setup
- https://developer.nvidia.com/embedded/learn/get-started-jetson-nano-devkit#prepare
- https://components101.com/development-boards/nvidia-jetson-nano-developer-kit