

NVIDIA DeepStream SDK 学习摘要

DeepStream 是 NVIDIA 提供的一个流分析工具包，专为构建 AI 驱动的多传感器处理、视频、图像分析应用而设计。它利用 NVIDIA 的 GPU 加速技术，提供从边缘到云的高性能视频分析能力。DeepStream 采用模块化流水线（Pipeline）设计，以 GStreamer 框架为基础，通过插件（Plugins）实现多级数据处理。DeepStream 提供多平台、可扩展、TLS 加密的安全，可以部署在本地、边缘和云端。

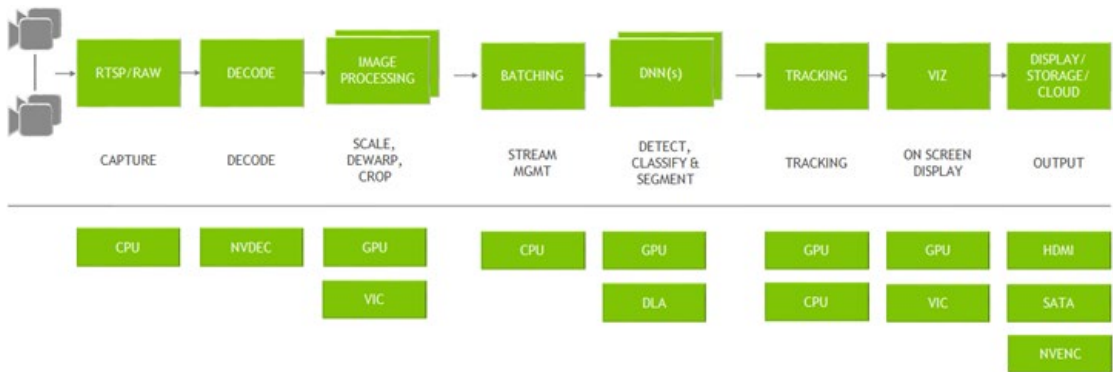


图 1：视频分析应用程序流程

(from https://docs.nvidia.com/metropolis/deepstream/dev-guide/text/DS_Overview.html)

DeepStream 的视频分析流程是一个高度优化的数据处理链，它从视频数据的输入开始，经过多个处理阶段，最终输出分析结果。这个流程主要包括：

- 数据输入：** 视频数据可以通过多种方式输入到 DeepStream 中，包括 USB/CSI 摄像头、文件中的视频或基于 RTSP 的流。
- 解码：** 使用 NVDEC 硬件加速器进行视频解码，确保高效处理。
- 预处理：** 解码后的视频帧可能需要进行预处理，如图像畸变校正或颜色空间转换，这些操作通常由 GPU 或 VIC（视觉图像合成器）执行。
- 批处理：** 使用 Gst-nvstreammux 插件对视频帧进行批处理，以优化后续的推理性能。
- 推理：** 使用 TensorRT 或 Triton 推理服务器进行 AI 推理，这可以在 GPU 或 DLA（深度学习加速器）上执行。
- 对象跟踪：** 推理后，使用 Gst-nvtracker 插件进行对象跟踪，以维持场景中对象的连续性。
- 可视化：** 使用 Gst-nvdsosd 插件创建可视化输出，如边界框、分割掩码和标签。
- 结果输出：** 分析结果可以渲染到屏幕上，保存到本地磁盘，通过 RTSP 流输出，或通过云服务发送元数据。

表 1：TAO Toolkit 和 Deepstream 的区别

组件	TAO Toolkit	Deepstream
用途	AI 模型训练与优化工具	实时视频分析推理框架
阶段	训练和优化	部署
核心技术	迁移学习、预训练模型、模型压缩	GStreamer, TensorRT
输入/输出	输入数据集（图像、视频），输出模型	处理视频流，实施传感器数据

deepstream-app 是 NVIDIA DeepStream SDK 提供的一个预构建的可执行文件，用于快速测试和部署基于配置文件的 AI 视频分析流水线。它封装了 DeepStream 的核心功能，用户无

需编写代码即可通过修改配置文件实现多路视频流的分析任务。

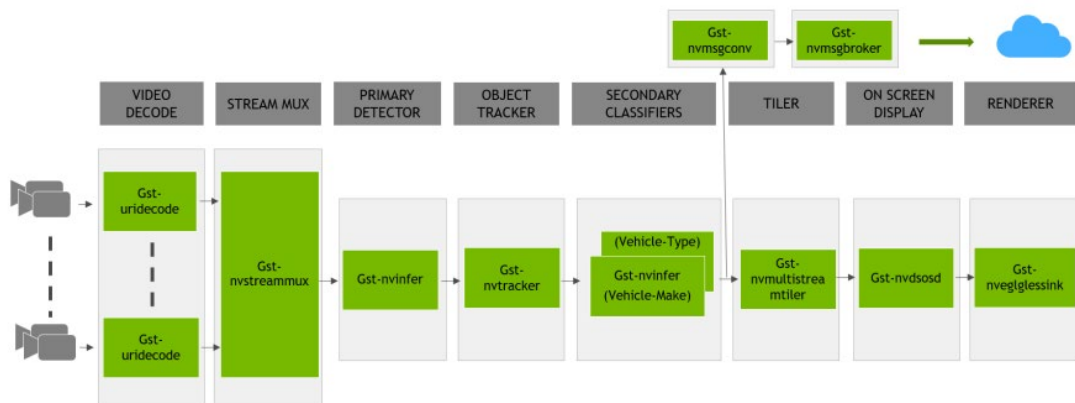


图 2: deepstream-app 架构

(from https://docs.nvidia.com/metropolis/deepstream/dev-guide/text/DS_ref_app_deepstream.html#application-architecture)