



DEEPSTREAM SDK 4.0 FOR NVIDIA DGPU AND JETSON

RN-09353-001 | August 6, 2019
Advance Information | Subject to Change

4.0 Release Notes



TABLE OF CONTENTS

1.0	ABOUT THIS RELEASE	3
1.1	What's New	3
1.2	Contents of this Release.....	5
1.3	Documentation in this Release	5
1.4	Differences with Deepstream 3.0	5
2.0	LIMITATIONS	6
3.0	NOTES	7
3.1	Applications May Be Deployed in a Docker Container	7
3.2	Sample Applications Malfunction if Docker Environment Cannot Support Display	10
3.3	Installing DeepStream with Jetson Software Developer Kit	10

1.0 ABOUT THIS RELEASE

These release notes are for the NVIDIA DeepStream SDK for NVIDIA® Tesla®, NVIDIA® Jetson AGX Xavier™, NVIDIA® Jetson Nano™, and NVIDIA® Jetson™ TX2 platforms.

1.1 WHAT'S NEW

The following new features are supported in this DeepStream SDK release:

- ▶ Single release for dGPU (Tesla) and Jetson platforms
- ▶ New memory management APIs
- ▶ New APIs for image conversion and scaling
- ▶ New meta data design for ease of use and customization
- ▶ Gst-nvinfer plugin enhancements:
 - Encoded model support from Transfer Learning Toolkit SDK. (For information, see <https://developer.nvidia.com/transfer-learning-toolkit>.)
 - Gray/monochrome model support
 - Support for flowing Tensor output as meta data in pipeline
 - Segmentation model support
 - Configurable support for maintaining aspect ratio when scaling to network resolution
 - New interface for generating an NVIDIA® TensorRT™ engine for custom models which are not UFF/ONNX/Caffe models
 - FP16 and INT8 DLA (Jetson Xavier) support
 - Source code available in DeepStream package
- ▶ New plugins:
 - V4L2-based accelerated H265 + H264 encode and decode
 - Accelerated JPEG + MJPEG decoder support
 - Gst-nvvideoconvert for accelerated conversion and scaling

- Gst-nvof, a new Gst plugin for optical flow
 - Gst-nvofvisual, a new plugin for visualizing optical flow output
 - Gst-nvsegvisual, a new plugin for visualizing segmentation model output
- ▶ New sample applications:
- Yolo: An example Yolo object detector (supporting Yolo v2, v2 tiny, v3, and v3 tiny detectors) showing use of the `IPugin` interface, custom output parsing, and the CUDA engine generation interface of Gst-nvinfer.
 - Custom meta data example: An example to demonstrate how to add custom/user specific meta data to any component of DeepStream.
 - Dewarper test example: Demonstrates dewarper functionality for single or multiple 360 degree camera streams. Reads camera calibration parameters from a CSV file and renders aisle and spot surfaces on screen.
 - Optical flow test example (for Turing GPU and Jetson Xavier only): Demonstrates optical flow functionality for single or multiple streams. This example uses two GStreamer plugins (Gst-nvof and Gst-nvofvisual). Gst-nvof generates MV (motion vector) data and attaches it as user meta data. Gst-nvofvisual enables visualizing the MV data using a predefined color wheel matrix.
 - Image decode test example: Demonstrates JPEG/MJPEG decode for inferencing.
 - Industrial use case of segmentation: Demonstrates segmentation of multi-stream video or images using a semantic or industrial neural network, and displays output.
 - Example for handling metadata before Gst-nvstreammux: Demonstrates how to set metadata upstream from the Gst-nvstreammux plugin in DeepStream pipeline, and how to access it downstream from that plugin.
 - Gst-nvinfer tensor meta flow example: Demonstrates how to flow and access Gst-nvinfer tensor output as metadata.
 - Performance demo example: Performs single channel cascaded inferencing and object tracking sequentially on all streams in a directory.
- ▶ Miscellaneous features
- Message broker: Support for Azure IoT.
 - Remote display support for headless server through RTSP out.
- ▶ Jetson Docker
- ▶ Source Release
- Gst-nvinfer GStreamer plugin and low level library
 - GStreamer V4L2 codecs plugin

Note: Due to major feature additions and API changes after DeepStream 3.0, NVIDIA recommends that users of that version migrate their applications and plugins to DeepStream 4.0 from scratch.

1.2 CONTENTS OF THIS RELEASE

This release includes the following:

- ▶ The DeepStream SDK. Refer to *NVIDIA DeepStream SDK V4.0 Development Guide* for a detailed description of the contents of the DeepStream release package. The *Development Guide* also contains other information to help you get started with DeepStream, including information about system software and hardware requirements and external software dependencies that you must install before you use the SDK.

For detailed information about GStreamer plugins, metadata usage, Dockers, application and plugin migration to DeepStream 4.0, troubleshooting, and a FAQ, see the *DeepStream 4.0 Plugin Manual*.

- ▶ DeepStream SDK for dGPU and Jetson Software License Agreement (SLA).
- ▶ `LICENSE.txt` contains the license terms of third party libraries used.

1.3 DOCUMENTATION IN THIS RELEASE

This release contains the following documentation.

- ▶ *NVIDIA DeepStream SDK Development Guide*
- ▶ *NVIDIA DeepStream SDK API Reference*
- ▶ *NVIDIA DeepStream SDK Plugin Manual*

1.4 DIFFERENCES WITH DEEPSTREAM 3.0

DeepStream 4.0 uses a unified code base for Jetson and dGPU platforms. These are the major differences from DeepStream 3.0:

- ▶ The 360° camera use case is not supported.
- ▶ There are significant modifications to meta data, memory management, conversion/transformation APIs, and the Gst-nvinfer plugin. See *DeepStream 4.0 Plugin Manual* for more details.

2.0 LIMITATIONS

This section provides details about issues discovered during development and QA but not resolved in this release.

- ▶ With V4L2 codecs only MAX 1024 (decode + encode) instances are provided. The maximum number of instances can be increased by doing changes in open source code.
- ▶ Small memory leak observed; fixes in progress.
- ▶ `detected-min-w` and `detected-min-h` must be set to values larger than 32 in the primary inference configuration file (`config_infer_primary.txt`) for `gst-dsexample` on Jetson.
- ▶ The Kafka protocol adapter sometimes does not automatically reconnect when the Kafka Broker to which it is connected goes down and comes back up, thereby requiring application restart.
- ▶ The Kafka protocol adapter currently can be used to connect to Kafka brokers only on port 9092.
- ▶ On NVIDIA® Jetson AGX Xavier™, more than 50 instances of certain 1080p H.265 streams are not working due to limited memory for the decoder.
- ▶ On Jetson, running a DeepStream application over SSH (via putty) with X11 forwarding does not work.
- ▶ DeepStream currently expects model network width to be a multiple of 4 and network height to be a multiple of 2.
- ▶ Docker Container on Jetson does not support executing inference using DLA, or video input using a CSI camera, or use of video encode functionality in a DeepStream application.
- ▶ `Gst-v4l2` decoder does not support Dynamic Resolution Change (DRC) streams due to non-availability of a DRC implementation.

3.0 NOTES

- ▶ Optical flow is only supported on dGPUs having Turing architecture and on NVIDIA® Jetson AGX Xavier™.
- ▶ NVIDIA® Jetson Nano™ and NVIDIA® Jetson™ TX2 support only FP16 and FP32 network precisions with NVIDIA® TensorRT™.
- ▶ Jetson AGX Xavier supports INT8, FP16 and FP32 network precisions with TensorRT.

3.1 APPLICATIONS MAY BE DEPLOYED IN A DOCKER CONTAINER

Applications built with DeepStream can be deployed using a Docker container, available on NGC (<https://ngc.nvidia.com/>). Sign up for an NVIDIA GPU Cloud account and look for `DeepStream` containers to get started.

After you sign in to your NGC account, go to Dashboard → Setup → Get API key to get your `nvcr.io` authentication details.

As an example, you can use the DeepStream 4.0 docker containers on NGC and run the `deepstream-test4-app` sample application as an Azure edge runtime module on your edge device.

The following procedure deploys `deepstream-test4-app`:

- ▶ Using a sample video stream (`sample_720p.h264`)
- ▶ Sending messages with minimal schema
- ▶ Running with display disabled
- ▶ Using message topic `mytopic` (message topic may not be empty)

Set up and install Azure IoT Edge on your system with the instructions provided in the Azure module client README file in the `deepstream4.0` package:

```
<deepstream-4.0_package>/sources/libs/azure_protocol_adaptor/module_client/README
```

Note: For the Jetson platform, omit installation of the moby packages. Moby is currently incompatible with NVIDIA Container Runtime.

See the Azure documentation for information about prerequisites for creating an Azure edge device on the Azure portal:

<https://docs.microsoft.com/en-us/azure/iot-edge/how-to-deploy-modules-portal#prerequisites>

To deploy deepstream-test4-app as an Azure IoT edge runtime module

1. On the Azure portal, click the IoT edge device you have created and click Set Modules.
2. Enter these settings:

Container Registry Settings:

Name: NGC

Address: nvcr.io

User name: \$oauthtoken

Password: *use the password or API key from your NGC account*

Deployment modules:

Add a new module with the name `ds4`

Image URI:

For x86: `nvcr.io/nvidia/deepstream:4.0-19.07`

For Jetson: `nvcr.io/nvidia/deepstream-14t:4.0-19.07`

Container Create options:

- For Jetson:

```
{
  "HostConfig": {
    "Runtime": "nvidia"
  },
  "WorkingDir":
"/root/deepstream_sdk_v4.0_jetson/sources/apps/sample_apps/deepstream-test4",
  "ENTRYPOINT": [
    "/opt/nvidia/deepstream/deepstream-4.0/bin/deepstream-test4-app",
    "-i",
    "/root/deepstream_sdk_v4.0_jetson/samples/streams/sample_720p.h264",
    "-p",

```



```

        "/opt/nvidia/deepstream/deepstream-
4.0/lib/libnvs_azure_edge_proto.so",
        "--no-display",
        "-s",
        "1",
        "--topic",
        "mytopic"
    ]
}

```

- For X86:

```

{
    "HostConfig": {
        "Runtime": "nvidia"
    },
    "WorkingDir":
"/root/deepstream_sdk_v4.0_x86_64/sources/apps/sample_apps/deepstream
-test4",
    "ENTRYPOINT": [
        "/opt/nvidia/deepstream/deepstream-4.0/bin/deepstream-test4-
app",
        "-i",
"/root/deepstream_sdk_v4.0_x86_64/samples/streams/sample_720p.h264",
        "-p",
"/opt/nvidia/deepstream/deepstream-
4.0/lib/libnvs_azure_edge_proto.so",
        "--no-display",
        "-s",
        "1",
        "--topic",
        "mytopic"
    ]
}

```

3. Specify route options for the module:

- Option 1: Use a default route where every message from every module is sent upstream.

```

{
    "routes": {
        "route": "FROM /messages/* INTO $upstream"
    }
}

```

- Option 2: Specific routes where messages sent upstream can be filtered based on topic name. For example, in the sample test programs, topic name `mytopic` is used for the module name `ds4`:

```

{
    "routes": {
        "route": "FROM /messages/modules/ds4/outputs/mytopic INTO
$upstream"
    }
}

```

```
}  
}
```

3.2 SAMPLE APPLICATIONS MALFUNCTION IF DOCKER ENVIRONMENT CANNOT SUPPORT DISPLAY

If the Docker environment cannot support display, the sample applications `deepstream-test1`, `deepstream-test2`, `deepstream-test3`, and `deepstream-test4` do not work as expected.

To correct this problem, you must recompile the test applications after replacing `nveglglessink` with `fakesink`. With `deepstream-test4`, you also have the option to specify `fakesink` by adding the `--no-display` command line switch.

3.3 INSTALLING DEEPSTREAM WITH JETSON SOFTWARE DEVELOPER KIT

If you try to install DeepStream alone using a Jetson software developer kit, DeepStream fails, due to DeepStream dependencies on other components.

To install DeepStream with a software developer kit, select all of the JetPack SDK components.

Notice

THE INFORMATION IN THIS DOCUMENT AND ALL OTHER INFORMATION CONTAINED IN NVIDIA DOCUMENTATION REFERENCED IN THIS DOCUMENT IS PROVIDED "AS IS." NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE INFORMATION FOR THE PRODUCT, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA's aggregate and cumulative liability towards customer for the product described in this document shall be limited in accordance with the NVIDIA terms and conditions of sale for the product. THE NVIDIA PRODUCT DESCRIBED IN THIS DOCUMENT IS NOT FAULT TOLERANT AND IS NOT DESIGNED, MANUFACTURED OR INTENDED FOR USE IN CONNECTION WITH THE DESIGN, CONSTRUCTION, MAINTENANCE, AND/OR OPERATION OF ANY SYSTEM WHERE THE USE OR A FAILURE OF SUCH SYSTEM COULD RESULT IN A SITUATION THAT THREATENS THE SAFETY OF HUMAN LIFE OR SEVERE PHYSICAL HARM OR PROPERTY DAMAGE (INCLUDING, FOR EXAMPLE, USE IN CONNECTION WITH ANY NUCLEAR, AVIONICS, LIFE SUPPORT OR OTHER LIFE CRITICAL APPLICATION). NVIDIA EXPRESSLY DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY OF FITNESS FOR SUCH HIGH RISK USES. NVIDIA SHALL NOT BE LIABLE TO CUSTOMER OR ANY THIRD PARTY, IN WHOLE OR IN PART, FOR ANY CLAIMS OR DAMAGES ARISING FROM SUCH HIGH RISK USES.

NVIDIA makes no representation or warranty that the product described in this document will be suitable for any specified use without further testing or modification. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer's sole responsibility to ensure the product is suitable and fit for the application planned by customer and to do the necessary testing for the application in order to avoid a default of the application or the product. Weaknesses in customer's product designs may affect the quality and reliability of the NVIDIA product and may result in additional or different conditions and/or requirements beyond those contained in this document. NVIDIA does not accept any liability related to any default, damage, costs or problem which may be based on or attributable to: (i) the use of the NVIDIA product in any manner that is contrary to this document, or (ii) customer product designs.

Other than the right for customer to use the information in this document with the product, no other license, either expressed or implied, is hereby granted by NVIDIA under this document. Reproduction of information in this document is permissible only if reproduction is approved by NVIDIA in writing, is reproduced without alteration, and is accompanied by all associated conditions, limitations, and notices.

Trademarks

NVIDIA, the NVIDIA logo, TensorRT, Jetson Nano, Jetson AGX Xavier, and NVIDIA Tesla are trademarks and/or registered trademarks of NVIDIA Corporation in the United States and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

Copyright © 2019 NVIDIA Corporation. All rights reserved.