TENSORRT

SWE-SWDOCTRT-001-APIS_v001 | December 2019

API Guide

TABLE OF CONTENTS

Chapter 1. C++ API

The C++ API allows developers to import, calibrate, generate and deploy networks using C++. Networks can be imported directly from NVCaffe, or from other frameworks via the UFF format. They may also be created programmatically by instantiating individual layers and setting parameters and weights directly.

Within the core C++ API in **`NvInfer.h`**, the following APIs are included:

- Builder API
- Execution API
- Network Definition API
- Plugin API

To view this API, see TensorRT C++ API.

For more information about the C++ API, including sample code, see TensorRT Developer Guide.

Chapter 2. PYTHON API

The TensorRT Python API enables developers, (in Python based development environments and those looking to experiment with TensorRT) to easily parse models (for example, from NVCaffe, TensorFlow, ONNX, and NumPy compatible frameworks) and generate and run PLAN files. Currently, all functionality except for Int8Calibrators and RNNs are available to use in Python.

To view this API, see TensorRT Python API.

For more information about the Python API, including sample code, see TensorRT Developer Guide.

2.1. Graph Surgeon API

Included within the Python API is the Graph Surgeon API; which enables you to transform TensorFlow graphs.

The Graph Surgeon API is located in **`graphsurgeon/graphsurgeon.html`** and contains three classes, **`Node Creation`**, **`Static Graph`**, and **`Dynamic Graph`**.

To view this API, see Graph Surgeon API.

For more information about the Graph Surgeon API, see TensorRT Developer Guide.

2.2. UFF API

Included within the Python API is the UFF API; a package that contains a set of utilities to convert trained models from various frameworks to a common format.

The UFF API is located in **`uff/uff.html`** and contains two conversion type tool classes called **`Tensorflow Modelstream to UFF`** and **`Tensorflow Frozen Protobuf Model to UFF`**.

To view this API, see UFF API.

For more information about the UFF API, see TensorRT Developer Guide.

Notice

Trademarks

Copyright

www.nvidia.com