



NVIDIA TensorRT

Support Matrix | NVIDIA Docs

Table of Contents

| | |
|---|---|
| Chapter 1. Features for Platforms and Software..... | 1 |
| Chapter 2. Hardware and Precision..... | 4 |
| Chapter 3. Compute Capability Per Platform..... | 6 |
| Chapter 4. Software Versions Per Platform..... | 7 |
| Chapter 5. ONNX Operator Support..... | 8 |

Chapter 1. Features for Platforms and Software

This section lists the supported NVIDIA® TensorRT™ features based on which platform and software.

Table 1. List of Supported Features per Platform

| | Linux x86-64 | Windows x64 | Linux ppc64le | Linux SBSA | JetPack |
|---|--|--|----------------------|-------------------------------|----------------------|
| | 10.3.x | 10.3.x | 8.5.x | 10.3.x | 10.3.x |
| Supported NVIDIA CUDA® versions | 12.5 update 1 12.4 update 1 12.3 update 2¹ 12.2 update 2² 12.1 update 1³ 12.0 update 1⁴ | 12.5 update 1 12.4 update 1 12.3 update 2¹³ 12.2 update 2¹⁴ 12.1 update 1¹⁵ 12.0 update 1¹⁶ | 11.8 | 12.5 update 1 | 12.6 |

¹ These CUDA versions are supported using a single build, built with CUDA toolkit 12.5. It is compatible with all CUDA 12.x versions and only requires driver 525.x.

¹³ These CUDA versions are supported using a single build, built with CUDA toolkit 12.5. It is compatible with all CUDA 12.x versions and only requires driver 525.x.

² These CUDA versions are supported using a single build, built with CUDA toolkit 12.5. It is compatible with all CUDA 12.x versions and only requires driver 525.x.

¹⁴ These CUDA versions are supported using a single build, built with CUDA toolkit 12.5. It is compatible with all CUDA 12.x versions and only requires driver 525.x.

³ These CUDA versions are supported using a single build, built with CUDA toolkit 12.5. It is compatible with all CUDA 12.x versions and only requires driver 525.x.

¹⁵ These CUDA versions are supported using a single build, built with CUDA toolkit 12.5. It is compatible with all CUDA 12.x versions and only requires driver 525.x.

⁴ These CUDA versions are supported using a single build, built with CUDA toolkit 12.5. It is compatible with all CUDA 12.x versions and only requires driver 525.x.

| | Linux x86-64 | Windows x64 | Linux ppc64le | Linux SBSA | JetPack |
|--|--|--|---------------|------------|---------|
| | 10.3.x | 10.3.x | 8.5.x | 10.3.x | 10.3.x |
| | 11.8 | 11.8 | | | |
| | 11.7 update 1 ⁵ | 11.7 update 1 ¹⁷ | | | |
| | 11.6 update 2 ⁶ | 11.6 update 2 ¹⁸ | | | |
| | 11.5 update 2 ⁷ | 11.5 update 2 ¹⁹ | | | |
| | 11.4 update 4 ⁸ | 11.4 update 4 ²⁰ | | | |
| | 11.3 update 1 ⁹ | 11.3 update 1 ²¹ | | | |
| | 11.2 update 2 ¹⁰ | 11.2 update 2 ²² | | | |
| | 11.1 update 1 ¹¹ | 11.1 update 1 ²³ | | | |

¹⁶ These CUDA versions are supported using a single build, built with CUDA toolkit 12.5. It is compatible with all CUDA 12.x versions and only requires driver 525.x.

⁵ These CUDA versions are supported using a single build, built with CUDA toolkit 11.8. It is compatible with all CUDA 11.x versions and only requires driver 450.x.

¹⁷ These CUDA versions are supported using a single build, built with CUDA toolkit 11.8. It is compatible with all CUDA 11.x versions and only requires driver 450.x.

⁶ These CUDA versions are supported using a single build, built with CUDA toolkit 11.8. It is compatible with all CUDA 11.x versions and only requires driver 450.x.

¹⁸ These CUDA versions are supported using a single build, built with CUDA toolkit 11.8. It is compatible with all CUDA 11.x versions and only requires driver 450.x.

⁷ These CUDA versions are supported using a single build, built with CUDA toolkit 11.8. It is compatible with all CUDA 11.x versions and only requires driver 450.x.

¹⁹ These CUDA versions are supported using a single build, built with CUDA toolkit 11.8. It is compatible with all CUDA 11.x versions and only requires driver 450.x.

⁸ These CUDA versions are supported using a single build, built with CUDA toolkit 11.8. It is compatible with all CUDA 11.x versions and only requires driver 450.x.

²⁰ These CUDA versions are supported using a single build, built with CUDA toolkit 11.8. It is compatible with all CUDA 11.x versions and only requires driver 450.x.

⁹ These CUDA versions are supported using a single build, built with CUDA toolkit 11.8. It is compatible with all CUDA 11.x versions and only requires driver 450.x.

²¹ These CUDA versions are supported using a single build, built with CUDA toolkit 11.8. It is compatible with all CUDA 11.x versions and only requires driver 450.x.


¹⁰ These CUDA versions are supported using a single build, built with CUDA toolkit 11.8. It is compatible with all CUDA 11.x versions and only requires driver 450.x.

²² These CUDA versions are supported using a single build, built with CUDA toolkit 11.8. It is compatible with all CUDA 11.x versions and only requires driver 450.x.

¹¹ These CUDA versions are supported using a single build, built with CUDA toolkit 11.8. It is compatible with all CUDA 11.x versions and only requires driver 450.x.

²³ These CUDA versions are supported using a single build, built with CUDA toolkit 11.8. It is compatible with all CUDA 11.x versions and only requires driver 450.x.

| | Linux x86-64 | Windows x64 | Linux ppc64le | Linux SBSA | JetPack |
|---|--|--|-----------------------------|-----------------------------|-----------------------------|
| | 10.3.x | 10.3.x | 8.5.x | 10.3.x | 10.3.x |
| | 11.0 update 3¹² | 11.0 update 3²⁴ | | | |
| Supported cuDNN versions (Optional) | cuDNN 8.9.7 | cuDNN 8.9.7 | cuDNN 8.6.0 | cuDNN 8.9.7 | cuDNN 8.9.6 |
| TensorRT Python API | Yes | Yes | Yes | Yes | Yes |
| NvOnnxParser | Yes | Yes | Yes | Yes | Yes |
| Loops | Yes | Yes | Yes | Yes | Yes |

 **Note:**

- ▶ Serialized engines are not portable across platforms.
- ▶ If a serialized engine was created using the version-compatible flag, it could run with newer versions of TensorRT within the same major version.
- ▶ If a serialized engine was created with hardware compatibility mode enabled, it can run on more than one kind of GPU architecture; the specifics depend on the hardware compatibility level used. Otherwise, serialized engines are not portable across devices.
- ▶ Refer to the minimum compatible driver versions in the [NVIDIA CUDA Release Notes](#) for specific [NVIDIA Driver](#) versions.

¹² These CUDA versions are supported using a single build, built with CUDA toolkit 11.8. It is compatible with all CUDA 11.x versions and only requires driver 450.x.

²⁴ These CUDA versions are supported using a single build, built with CUDA toolkit 11.8. It is compatible with all CUDA 11.x versions and only requires driver 450.x.

Chapter 2. Hardware and Precision

The following table lists NVIDIA hardware and the precision modes each hardware supports. TensorRT has been compiled to support all NVIDIA hardware with SM 7.0 or higher capability. The table also lists the availability of DLA on this hardware. Refer to the following tables for the specifics.

Table 2. Supported Hardware

| CUDA Compu Capabi | Exempl Device | TF32 | FP32 | FP16 | FP8 | BF16 | INT8 | FP16 Tensor Cores | INT8 Tensor Cores | DLA |
|----------------------------------|--|-------------|-------------|-------------|------------|-------------|-------------|----------------------------------|----------------------------------|------------|
| 9.0 | NVIDIA H100 NVIDIA GH200 480GB | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No |
| 8.9 | NVIDIA L40S | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No |
| 8.7 | NVIDIA DRIVE AGX Orin™ | Yes | Yes | Yes | No | No | Yes | Yes | Yes | Yes |
| 8.6 | NVIDIA A10 | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | No |
| 8.0 | NVIDIA A100 | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | No |
| 7.5 | NVIDIA T4 | No | Yes | Yes | No | No | Yes | Yes | Yes | No |

Deprecated Hardware

Table 3. Deprecated Hardware

| CUDA Compu Capabi | Examp Device | TF32 | FP32 | FP16 | FP8 | BF16 | INT8 | FP16 Tensor Cores | INT8 Tensor Cores | DLA |
|----------------------------------|-------------------------|-------------|-------------|-------------|------------|-------------|-------------|----------------------------------|----------------------------------|------------|
| 7.0 | NVIDIA V100 | No | Yes | Yes | No | No | Yes | Yes | No | No |

Chapter 3. Compute Capability Per Platform

The section lists the supported computing capability based on the platform.

Table 4. Compute Capability per Platform

| Platform | Compute capability |
|------------------------|--|
| Linux x86-64 | 7.0, 7.5, 8.0, 8.6 ²⁵ , 8.9 ²⁶ , 9.0 ²⁷ |
| Windows 10 x64 | 7.0, 7.5, 8.0, 8.6 ²⁸ , 8.9 ²⁹ , 9.0 ³⁰ |
| CentOS 8.5 ppc64le | 7.0, 7.5, 8.0, 8.6, 9.0 |
| Ubuntu 22.04 SBSA | 7.0, 7.5, 8.0, 8.6, 9.0 |
| NVIDIA JetPack AArch64 | 8.7 |

²⁵ Requires CUDA Toolkit 11.1 or newer.

²⁶ Requires CUDA Toolkit 11.8 or newer.

²⁷ Requires CUDA Toolkit 11.8 or newer.

²⁸ Requires CUDA Toolkit 11.1 or newer.

²⁹ Requires CUDA Toolkit 11.8 or newer.

³⁰ Requires CUDA Toolkit 11.8 or newer.

Chapter 4. Software Versions Per Platform

The section lists the supported software versions based on platform.

Table 5. List of Supported Platforms per Software Version

| Platform | Compiler version | Python version |
|---|-----------------------------------|----------------------|
| Ubuntu 20.04 x86-64 | gcc 8.3.1 | 3.8 |
| Ubuntu 22.04 x86-64 | gcc 8.3.1 | 3.10 |
| Rocky Linux 8.9 x86-64 | gcc 8.3.1 | 3.8 |
| Rocky Linux 9.3 x86-64 | gcc 8.3.1 | |
| SLES 15 x86-64 | gcc 8.3.1 | N/A |
| <ul style="list-style-type: none">▶ Windows 10 x64▶ Windows 11 x64▶ Windows Server 2019 x64▶ Windows Server 2022 x64 | MSVC 2019 v16.9.2 | N/A |
| CentOS 8.5 ppc64le | Clang 14.0.6 | 3.8 |
| Ubuntu 22.04 SBSA | gcc 8.4 | 3.10 |
| NVIDIA JetPack AArch64 | gcc 11.4 | 3.10 |



Note: Python versions 3.8, 3.9, 3.10, 3.11, and 3.12 are supported using Debian or RPM packages and when using Python wheel files.

Chapter 5. ONNX Operator Support

The ONNX operator support list for TensorRT can be found [here](#).

Notice

This document is provided for information purposes only and shall not be regarded as a warranty of a certain functionality, condition, or quality of a product. NVIDIA Corporation ("NVIDIA") makes no representations or warranties, expressed or implied, as to the accuracy or completeness of the information contained in this document and assumes no responsibility for any errors contained herein. NVIDIA shall have no liability for the consequences or use of such information or for any infringement of patents or other rights of third parties that may result from its use. This document is not a commitment to develop, release, or deliver any Material (defined below), code, or functionality.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and any other changes to this document, at any time without notice.

Customer should obtain the latest relevant information before placing orders and should verify that such information is current and complete.

NVIDIA products are sold subject to the NVIDIA standard terms and conditions of sale supplied at the time of order acknowledgement, unless otherwise agreed in an individual sales agreement signed by authorized representatives of NVIDIA and customer ("Terms of Sale"). NVIDIA hereby expressly objects to applying any customer general terms and conditions with regards to the purchase of the NVIDIA product referenced in this document. No contractual obligations are formed either directly or indirectly by this document.

NVIDIA products are not designed, authorized, or warranted to be suitable for use in medical, military, aircraft, space, or life support equipment, nor in applications where failure or malfunction of the NVIDIA product can reasonably be expected to result in personal injury, death, or property or environmental damage. NVIDIA accepts no liability for inclusion and/or use of NVIDIA products in such equipment or applications and therefore such inclusion and/or use is at customer's own risk.

NVIDIA makes no representation or warranty that products based on this document will be suitable for any specified use. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer's sole responsibility to evaluate and determine the applicability of any information contained in this document, ensure the product is suitable and fit for the application planned by customer, and perform 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 accepts no 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.

No license, either expressed or implied, is granted under any NVIDIA patent right, copyright, or other NVIDIA intellectual property right under this document. Information published by NVIDIA regarding third-party products or services does not constitute a license from NVIDIA to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property rights of the third party, or a license from NVIDIA under the patents or other intellectual property rights of NVIDIA.

Reproduction of information in this document is permissible only if approved in advance by NVIDIA in writing, reproduced without alteration and in full compliance with all applicable export laws and regulations, and accompanied by all associated conditions, limitations, and notices.

THIS DOCUMENT AND ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, "MATERIALS") ARE BEING PROVIDED "AS IS." NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. TO THE EXTENT NOT PROHIBITED BY LAW, IN NO EVENT WILL NVIDIA BE LIABLE FOR ANY DAMAGES, INCLUDING WITHOUT LIMITATION ANY DIRECT, INDIRECT, SPECIAL, INCIDENTAL, PUNITIVE, OR CONSEQUENTIAL DAMAGES, HOWEVER CAUSED AND REGARDLESS OF THE THEORY OF LIABILITY, ARISING OUT OF ANY USE OF THIS DOCUMENT, EVEN IF NVIDIA HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA's aggregate and cumulative liability towards customer for the products described herein shall be limited in accordance with the Terms of Sale for the product.

Arm

Arm, AMBA and Arm Powered are registered trademarks of Arm Limited. Cortex, MPCore and Mali are trademarks of Arm Limited. "Arm" is used to represent Arm Holdings plc; its operating company Arm Limited; and the regional subsidiaries Arm Inc.; Arm KK; Arm Korea Limited.; Arm Taiwan Limited; Arm France SAS; Arm Consulting (Shanghai) Co. Ltd.; Arm Germany GmbH; Arm Embedded Technologies Pvt. Ltd.; Arm Norway, AS and Arm Sweden AB.

HDMI

HDMI, the HDMI logo, and High-Definition Multimedia Interface are trademarks or registered trademarks of HDMI Licensing LLC.

BlackBerry/QNX

Copyright © 2020 BlackBerry Limited. All rights reserved.

Trademarks, including but not limited to BLACKBERRY, EMBLEM Design, QNX, AVIAGE, MOMENTICS, NEUTRINO and QNX CAR are the trademarks or registered trademarks of BlackBerry Limited, used under license, and the exclusive rights to such trademarks are expressly reserved.

Google

Android, Android TV, Google Play and the Google Play logo are trademarks of Google, Inc.

Trademarks

NVIDIA, the NVIDIA logo, and BlueField, CUDA, DALI, DRIVE, Hopper, JetPack, Jetson AGX Xavier, Jetson Nano, Maxwell, NGC, Nsight, Orin, Pascal, Quadro, Tegra, TensorRT, Triton, Turing and Volta 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

© 2018-2024 NVIDIA Corporation & affiliates. All rights reserved.

