



# PyTorch for Jetson Platform

## Release Notes

# Table of Contents

Chapter 1. Overview.....	1
Chapter 2. PyTorch for Jetson Platform.....	2

---

# Chapter 1. Overview

## PyTorch on Jetson Platform

[PyTorch](#) (for JetPack) is an optimized tensor library for deep learning, using GPUs and CPUs. Automatic differentiation is done with a tape-based system at both a functional and neural network layer level. This functionality brings a high level of flexibility, speed as a deep learning framework, and provides accelerated NumPy-like functionality. These NVIDIA-provided redistributables are Python pip wheel installers for PyTorch, with GPU-acceleration and support for cuDNN. The packages are intended to be installed on top of the specified version of JetPack as in the provided documentation.

## Jetson AGX Xavier

The NVIDIA Jetson AGX Xavier developer kit for Jetson platform is the world's first AI computer for autonomous machines. The Jetson AGX Xavier delivers the performance of a GPU workstation in an embedded module under 30W.

## Jetson AGX Orin

The NVIDIA Jetson AGX Orin Developer Kit includes a high-performance, power-efficient Jetson AGX Orin module, and can emulate the other Jetson modules. You now have up to 275 TOPS and 8X the performance of NVIDIA Jetson AGX Xavier in the same compact form-factor for developing advanced robots and other autonomous machine products.

---

# Chapter 2. PyTorch for Jetson Platform

This document describes the key features, software enhancements and improvements, and known issues regarding PyTorch [2.0.0](#) on the Jetson platform.

## Key Features and Enhancements

This release includes the following key features and enhancements.

- ▶ The TF32 numerical format is enabled by default for cuBLAS and cuDNN operations on Ampere GPUs starting with the 22.06 release. If you encounter training issues especially for regression, generative or higher-order models, or by using TF32 operations in pre- or post-processing steps, try to disable TF32 by setting the following:

```
torch.set_float32_matmul_precision('highest')
```

## Compatibility

Table 1. PyTorch compatibility with NVIDIA containers and Jetpack

PyTorch Version	NVIDIA Framework Container	JetPack Version
<a href="#">2.0.0</a>	23.06, 23.05	5.1.x
<a href="#">2.0.0a0+fe05266f</a>	23.04	
<a href="#">2.0.0a0+8aa34602</a>	23.03	
<a href="#">1.14.0a0+44dac51c</a>	23.02, 23.01	5.0.2
<a href="#">1.13.0a0+936e930</a>	22.12, 22.11	
<a href="#">1.13.0a0+d0d6b1f</a>	22.10, 22.09	
<a href="#">1.13.0a0+08820cb</a>	22.07	
<a href="#">1.13.0a0+340c412</a>	22.06	5.0.1
<a href="#">1.12.0a0+8a1a93a9</a>	22.05	5.0

PyTorch Version	NVIDIA Framework Container	JetPack Version
<a href="#">1.12.0a0+bd13bc66</a>	22.04	
<a href="#">1.12.0a0+2c916ef</a>	22.03	
<a href="#">1.11.0a0+bfe5ad28</a>	22.01	4.6.1

## Using PyTorch with the Jetson Platform

### Storage

If you need more storage, we recommend connecting an external SSD via SATA on TX2 or Xavier devices, or USB on Jetson Nano.

### Known Issues

- ▶ If you receive a `CUPTI_ERROR_INSUFFICIENT_PRIVILEGES` error while profiling your code, run the script via `sudo` or ensure that your current user has the appropriate permissions to run CUPTI profiling.
- ▶ Building custom CUDA extensions may break due to symbol leaking. This will be fixed in a future release.
- ▶ A functional regression might be observed on Orin devices when calling into `torch.linalg.ldl_solve` showing a memory violation.
- ▶ Some ops have numerical mismatches on ARM CPUs.
- ▶ The wheel does not currently support flash attention.
- ▶ [torch.cuda.synchronize](#) does not behave as expected when used between [torch.mem\\_get\\_info](#).

## 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.

## HDMI

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

## OpenCL

OpenCL is a trademark of Apple Inc. used under license to the Khronos Group Inc.



**Trademarks**

NVIDIA, the NVIDIA logo, and cuBLAS, CUDA, DALI, DGX, DGX-1, DGX-2, DGX Station, DLProf, Jetson, Kepler, Maxwell, NCCL, Nsight Compute, Nsight Systems, NvCaffe, PerfWorks, Pascal, SDK Manager, Tegra, TensorRT, Triton Inference Server, Tesla, TF-TRT, and Volta are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

**Copyright**

© 2022-2023 NVIDIA Corporation & Affiliates. All rights reserved.

