



CAFFE2

RN-08515-001_v18.08 | January 2020

Release Notes



TABLE OF CONTENTS

Chapter 1. Caffe2 Overview.....	1
Chapter 2. Pulling A Container.....	2
Chapter 3. Running Caffe2.....	3
Chapter 4. Caffe2 Release 18.08.....	5
Chapter 5. Caffe2 Release 18.07.....	7
Chapter 6. Caffe2 Release 18.06.....	9
Chapter 7. Caffe2 Release 18.05.....	11
Chapter 8. Caffe2 Release 18.04.....	13
Chapter 9. Caffe2 Release 18.03.....	15
Chapter 10. Caffe2 Release 18.02.....	17
Chapter 11. Caffe2 Release 18.01.....	19
Chapter 12. Caffe2 Release 17.12.....	21
Chapter 13. Caffe2 Release 17.11.....	23
Chapter 14. Caffe2 Release 17.10.....	25
Chapter 15. Caffe2 Release 17.09.....	27
Chapter 16. Caffe2 Release 17.07.....	29
Chapter 17. Caffe2 Release 17.06.....	31
Chapter 18. Caffe2 Release 17.05.....	33
Chapter 19. Caffe2 Release 17.04.....	35

Chapter 1.

CAFFE2 OVERVIEW

The NVIDIA Deep Learning SDK accelerates widely-used deep learning frameworks such as [Caffe2™](#).

Caffe2 is a deep-learning framework designed to easily express all model types, for example, convolutional neural networks (CNNs), recurrent neural networks (RNNs), and more, in a friendly Python-based API, and execute them using a highly efficient C++ and CUDA® backend.

Caffe2 includes a flexible API that lets users define models for inference or training using expressive, high-level operations. The Python interface allows easy control and visualization of the inference or training process.

Caffe2 supports single and multi-GPU execution, along with support for multi-node execution.

See `/workspace/README.md` inside the container for information on customizing your Caffe2 image. For more information about Caffe2, including tutorials, documentation, and examples, see:

- ▶ [Caffe2 tutorials](#)
- ▶ [Caffe2 project](#)

This document describes the key features, software enhancements and improvements, any known issues, and how to run this container.

Chapter 2.

PULLING A CONTAINER

Before you can pull a container from the NGC container registry, you must have Docker installed. For DGX users, this is explained in [Preparing to use NVIDIA Containers Getting Started Guide](#).

For users other than DGX, follow the NVIDIA[®] GPU Cloud[™] (NGC) container registry [installation documentation](#) based on your platform.

You must also have access and logged into the NGC container registry as explained in the [NGC Getting Started Guide](#).

There are four repositories where you can find the NGC docker containers.

`nvcr.io/nvidia`

The deep learning framework containers are stored in the **`nvcr.io/nvidia`** repository.

`nvcr.io/hpc`

The HPC containers are stored in the **`nvcr.io/hpc`** repository.

`nvcr.io/nvidia-hpcvis`

The HPC visualization containers are stored in the **`nvcr.io/nvidia-hpcvis`** repository.

`nvcr.io/partner`

The partner containers are stored in the **`nvcr.io/partner`** repository. Currently the partner containers are focused on Deep Learning or Machine Learning, but that doesn't mean they are limited to those types of containers.

Chapter 3.

RUNNING CAFFE2

Before you can run an NGC deep learning framework container, your Docker environment must support NVIDIA GPUs. To run a container, issue the appropriate command as explained in the [Running A Container](#) chapter in the *NVIDIA Containers And Frameworks User Guide* and specify the registry, repository, and tags.

On a system with GPU support for NGC containers, the following occurs when running a container:

- ▶ The Docker engine loads the image into a container which runs the software.
- ▶ You define the runtime resources of the container by including additional flags and settings that are used with the command. These flags and settings are described in [Running A Container](#).
- ▶ The GPUs are explicitly defined for the Docker container (defaults to all GPUs, can be specified using **NV_GPU** environment variable).

The method implemented in your system depends on the DGX OS version installed (for DGX systems), the specific NGC Cloud Image provided by a Cloud Service Provider, or the software that you have installed in preparation for running NGC containers on TITAN PCs, Quadro PCs, or vGPUs.

1. Issue the command for the applicable release of the container that you want. The following command assumes you want to pull the latest container.

```
docker pull nvcr.io/nvidia/caffe2:18.08-py<x>
```

2. Open a command prompt and paste the pull command. The pulling of the container image begins. Ensure the pull completes successfully before proceeding to the next step.
3. Run the container image. A typical command to launch the container is:

```
docker run --gpus all -it --rm -v local_dir:container_dir nvcr.io/nvidia/caffe2:<xx.xx>-py<x>
```

You might want to pull in data and model descriptions from locations outside the container for use by Caffe2™ or save results to locations outside the container. To

accomplish this, the easiest method is to mount one or more host directories as Docker[®] data volumes.

You have pulled the latest files and run the container image.



In order to share data between ranks, NVIDIA[®] Collective Communications Library[™] (NCCL) may require shared system memory for IPC and pinned (page-locked) system memory resources. The operating system's limits on these resources may need to be increased accordingly. Refer to your system's documentation for details.

In particular, Docker containers default to limited shared and pinned memory resources. When using NCCL inside a container, it is recommended that you increase these resources by issuing:

```
--shm-size=1g --ulimit memlock=-1
```

in the command line to:

```
docker run --gpus all
```

Chapter 4.

CAFFE2 RELEASE 18.08

The NVIDIA container image of Caffe2, release 18.08, is available.

Contents of Caffe2

This container image contains the complete source of the version of Caffe2 in `/opt/caffe2`. It is pre-built and installed into the `/opt/caffe2/[binaries,lib]` directories in the container image.

The container also includes the following:

- ▶ [Ubuntu 16.04](#)



Container image `18.08-py2` contains [Python 2.7](#); `18.08-py3` contains [Python 3.5](#).

- ▶ [NVIDIA CUDA 9.0.176](#) (see Errata section and 2.1) including [CUDA[®] Basic Linear Algebra Subroutines library[™] \(cuBLAS\) 9.0.425](#)
- ▶ [NVIDIA CUDA[®] Deep Neural Network library[™] \(cuDNN\) 7.2.1](#)
- ▶ [NCCL 2.2.13](#) (optimized for [NVLink[™]](#))
- ▶ [OpenMPI[™] 1.10.3](#)

Driver Requirements

Release 18.08 is based on [CUDA 9](#), which requires [NVIDIA Driver](#) release 384.xx.

Key Features and Enhancements

This Caffe2 release includes the following key features and enhancements.

- ▶ [Caffe2](#) container image version 18.08 is based on [Caffe2 0.8.1](#).
- ▶ Latest version of [cuDNN 7.2.1](#).
- ▶ [Ubuntu 16.04](#) with July 2018 updates

Announcements

[Caffe2 is now merged into PyTorch](#). Due to lack of upstream changes in the separate Caffe2 container and with the merging of Caffe2 into a separate repository with PyTorch, we will discontinue updates for the separate Caffe2 container at the next version of CUDA.

Known Issues

There are no known issues in this release.

Chapter 5.

CAFFE2 RELEASE 18.07

The NVIDIA container image of Caffe2, release 18.07, is available.

Contents of Caffe2

This container image contains the complete source of the version of Caffe2 in `/opt/caffe2`. It is pre-built and installed into the `/opt/caffe2/[binaries,lib]` directories in the container image.

The container also includes the following:

- ▶ Ubuntu 16.04



Container image `18.07-py2` contains Python 2.7; `18.07-py3` contains Python 3.5.

- ▶ NVIDIA CUDA 9.0.176 (see Errata section and 2.1) including CUDA[®] Basic Linear Algebra Subroutines library[™] (cuBLAS) 9.0.425
- ▶ NVIDIA CUDA[®] Deep Neural Network library[™] (cuDNN) 7.1.4
- ▶ NCCL 2.2.13 (optimized for NVLink[™])
- ▶ OpenMPI[™] 1.10.3

Driver Requirements

Release 18.07 is based on CUDA 9, which requires NVIDIA Driver release 384.xx.

Key Features and Enhancements

This Caffe2 release includes the following key features and enhancements.

- ▶ Caffe2 container image version 18.07 is based on Caffe2 0.8.1.
- ▶ No new upstream changes in Caffe2 for June
- ▶ Latest version of CUDA[®] Basic Linear Algebra Subroutines library[™] (cuBLAS) 9.0.425.

- ▶ Ubuntu 16.04 with June 2018 updates

Announcements

Caffe2 is now merged into PyTorch. Due to lack of upstream changes in the separate Caffe2 container and with the merging of Caffe2 into a separate repository with PyTorch, we will discontinue updates for the separate Caffe2 container at the next version of CUDA.

Known Issues

There are no known issues in this release.

Chapter 6.

CAFFE2 RELEASE 18.06

The NVIDIA container image of Caffe2, release 18.06, is available.

Contents of Caffe2

This container image contains the complete source of the version of Caffe2 in `/opt/caffe2`. It is pre-built and installed into the `/opt/caffe2/[binaries,lib]` directories in the container image.

The container also includes the following:

- ▶ [Ubuntu 16.04](#)



Container image `18.06-py2` contains [Python 2.7](#); `18.06-py3` contains [Python 3.5](#).

- ▶ [NVIDIA CUDA 9.0.176](#) (see Errata section and 2.1) including [CUDA[®] Basic Linear Algebra Subroutines library[™] \(cuBLAS\) 9.0.333](#) (see section 2.3.1)
- ▶ [NVIDIA CUDA[®] Deep Neural Network library[™] \(cuDNN\) 7.1.4](#)
- ▶ [NCCL 2.2.13](#) (optimized for [NVLink[™]](#))
- ▶ [OpenMPI[™] 1.10.3](#)

Driver Requirements

Release 18.06 is based on [CUDA 9](#), which requires [NVIDIA Driver](#) release 384.xx.

Key Features and Enhancements

This Caffe2 release includes the following key features and enhancements.

- ▶ [Caffe2](#) container image version 18.06 is based on [Caffe2 0.8.1](#).
- ▶ No new upstream changes in Caffe2 for May
- ▶ [Ubuntu 16.04](#) with May 2018 updates

Announcements

[Caffe2 is now merged into PyTorch](#). Due to lack of upstream changes in the separate Caffe2 container and with the merging of Caffe2 into a separate repository with PyTorch, we will discontinue updates for the separate Caffe2 container at the next version of CUDA.

Known Issues

There are no known issues in this release.

Chapter 7.

CAFFE2 RELEASE 18.05

The NVIDIA container image of Caffe2, release 18.05, is available.

Contents of Caffe2

This container image contains the complete source of the version of Caffe2 in `/opt/caffe2`. It is pre-built and installed into the `/opt/caffe2/[binaries,lib]` directories in the container image.

The container also includes the following:

- ▶ [Ubuntu 16.04](#)



Container image `18.05-py2` contains [Python 2.7](#); `18.05-py3` contains [Python 3.5](#).

- ▶ [NVIDIA CUDA 9.0.176](#) (see Errata section and 2.1) including [CUDA[®] Basic Linear Algebra Subroutines library[™] \(cuBLAS\) 9.0.333](#) (see section 2.3.1)
- ▶ [NVIDIA CUDA[®] Deep Neural Network library[™] \(cuDNN\) 7.1.2](#)
- ▶ [NCCL 2.1.15](#) (optimized for [NVLink[™]](#))
- ▶ [OpenMPI[™] 1.10.3](#)

Driver Requirements

Release 18.05 is based on [CUDA 9](#), which requires [NVIDIA Driver](#) release 384.xx.

Key Features and Enhancements

This Caffe2 release includes the following key features and enhancements.

- ▶ [Caffe2](#) container image version 18.05 is based on [Caffe2 0.8.1](#).
- ▶ No new upstream changes in Caffe2 for April
- ▶ [Ubuntu 16.04](#) with April 2018 updates

Announcements

Starting with the next major version of CUDA release, we will no longer provide Python 2 containers and will only maintain Python 3 containers.

[Caffe2 is now merged into PyTorch](#). Due to lack of upstream changes in the separate Caffe2 container and merges into a separate repository with PyTorch, we will discontinue updates at the next version of CUDA. Until we discontinue, we will continue to update NVIDIA libraries and base OS updates.

Known Issues

There are no known issues in this release.

Chapter 8.

CAFFE2 RELEASE 18.04

The NVIDIA container image of Caffe2, release 18.04, is available.

Contents of Caffe2

This container image contains the complete source of the version of Caffe2 in `/opt/caffe2`. It is pre-built and installed into the `/opt/caffe2/[binaries,lib]` directories in the container image.

The container also includes the following:

- ▶ [Ubuntu 16.04](#)



Container image `18.04-py2` contains [Python 2.7](#); `18.04-py3` contains [Python 3.5](#).

- ▶ [NVIDIA CUDA 9.0.176](#) (see [Errata section](#) and [2.1](#)) including [CUDA[®] Basic Linear Algebra Subroutines library[™] \(cuBLAS\) 9.0.333](#) (see [section 2.3.1](#))
- ▶ [NVIDIA CUDA[®] Deep Neural Network library[™] \(cuDNN\) 7.1.1](#)
- ▶ [NCCL 2.1.15](#) (optimized for [NVLink[™]](#))
- ▶ [OpenMPI[™] 1.10.3](#)

Driver Requirements

Release 18.04 is based on [CUDA 9](#), which requires [NVIDIA Driver](#) release 384.xx.

Key Features and Enhancements

This Caffe2 release includes the following key features and enhancements.

- ▶ [Caffe2](#) container image version 18.04 is based on [Caffe2 0.8.1](#).
- ▶ Fixed a bug in `nvidia-examples/seq2seq` that was reported in the [18.03 Release Notes](#). This bug caused executing the wrong number of iterations or epochs when loading snapshots.

- ▶ Caffe2 runtime binary compiled with NVTX support.
- ▶ Support NVTX ranges for CPU operators (GPU operators were already supported).
- ▶ Latest version of NCCL 2.1.15
- ▶ Ubuntu 16.04 with March 2018 updates

Announcements

Starting with the next major version of CUDA release, we will no longer provide Python 2 containers and will only maintain Python 3 containers.

Known Issues

There are no known issues in this release.

Chapter 9.

CAFFE2 RELEASE 18.03

The NVIDIA container image of Caffe2, release 18.03, is available.

Contents of Caffe2

This container image contains the complete source of the version of Caffe2 in `/opt/caffe2`. It is pre-built and installed into the `/opt/caffe2/[binaries,lib]` directories in the container image.

The container also includes the following:

- ▶ [Ubuntu 16.04](#)



Container image `18.03-py2` contains [Python 2.7](#); `18.03-py3` contains [Python 3.5](#).

- ▶ [NVIDIA CUDA 9.0.176](#) (see [Errata section](#) and [2.1](#)) including [CUDA[®] Basic Linear Algebra Subroutines library[™] \(cuBLAS\) 9.0.333](#) (see [section 2.3.1](#))
- ▶ [NVIDIA CUDA[®] Deep Neural Network library[™] \(cuDNN\) 7.1.1](#)
- ▶ [NCCL 2.1.2](#) (optimized for [NVLink[™]](#))
- ▶ [OpenMPI[™] 1.10.3](#)

Driver Requirements

Release 18.03 is based on [CUDA 9](#), which requires [NVIDIA Driver](#) release 384.xx.

Key Features and Enhancements

This Caffe2 release includes the following key features and enhancements.

- ▶ [Caffe2](#) container image version 18.03 is based on [Caffe2 0.8.1](#).
- ▶ When using ImageNet training scripts in `nvidia-examples` on multiple GPUs, the printed metrics in the log for weak scaling was wrong. Also, the number of epochs the model is trained for was wrong. Both of these issues are fixed in this release.

- ▶ Gradient clipping used to be done by executing a series of small operators that compute a ratio by which the learning rate gets scaled, which has the same effect as gradient clipping for SGD optimizers. However, that method is wrong with optimizers that use momentum or history such as AdaGrad and Adam. In this release, we added a new operator **ClipByGlobalNorm** that explicitly clips the gradient. This operator also supports mixed precision for inputs and outputs.
- ▶ Caffe2 already supported cuDNN RNN, however that integration does not provide enough features and flexibility to use cuDNN RNN in seq2seq. We improved this integration and also enabled using cuDNN RNN in the seq2seq example in **nvidia-examples**.
- ▶ Incorporated GitHub Caffe2 code as of February 16, 2018.
- ▶ Latest version of cuBLAS 9.0.333
- ▶ Latest version of cuDNN 7.1.1
- ▶ Ubuntu 16.04 with February 2018 updates

Announcements

Starting with the next major version of CUDA release, we will no longer provide Python 2 containers and will only maintain Python 3 containers.

Known Issues

In **nvidia-examples/seq2seq**, there is a bug that causes training to skip one epoch in case of loading a snapshot. This bug will be fixed in 18.04.

Chapter 10.

CAFFE2 RELEASE 18.02

The NVIDIA container image of Caffe2, release 18.02, is available.

Caffe2 container image version 18.02 is based on [Caffe2 0.8.1](#).

Contents of Caffe2

This container image contains the complete source of the version of Caffe2 in /`opt/caffe2`. It is pre-built and installed into the /`opt/caffe2/[binaries,lib]` directories in the container image.

The container also includes the following:

- ▶ [Ubuntu 16.04](#)



Container image `18.02-py2` contains [Python 2.7](#); `18.02-py3` contains [Python 3.5](#).

- ▶ [NVIDIA CUDA 9.0.176](#) including:
 - ▶ [CUDA[®] Basic Linear Algebra Subroutines library[™] \(cuBLAS\) 9.0.282 Patch 2](#) which is installed by default
 - ▶ [cuBLAS 9.0.234 Patch 1](#) as a debian file. Installing Patch 1 by issuing the `dpkg -i /opt/cuda-cublas-9-0_9.0.234-1_amd64.deb` command is the workaround for the known issue described below.
- ▶ [NVIDIA CUDA[®] Deep Neural Network library[™] \(cuDNN\) 7.0.5](#)
- ▶ [NCCL 2.1.2 \(optimized for NVLink[™]\)](#)
- ▶ [OpenMPI[™] 1.10.3](#)

Driver Requirements

Release 18.02 is based on CUDA 9, which requires [NVIDIA Driver](#) release 384.xx.

Key Features and Enhancements

This Caffe2 release includes the following key features and enhancements.

- ▶ Enhanced debug mode in `nvidia-examples/seq2seq`
- ▶ Added `nvidia-examples/onnx-inference` for testing ONNX
- ▶ Enabled FP16 in `TransposeOp` that uses cuDNN
- ▶ Latest version of cuBLAS
- ▶ Ubuntu 16.04 with January 2018 updates

Known Issues

cuBLAS 9.0.282 regresses RNN seq2seq FP16 performance for a small subset of input sizes. This issue should be fixed in the next update. As a workaround, install cuBLAS 9.0.234 Patch 1 by issuing the `dpkg -i /opt/cuda-cublas-9-0_9.0.234-1_amd64.deb` command.

Chapter 11.

CAFFE2 RELEASE 18.01

The NVIDIA container image of Caffe2, release 18.01, is available.

Caffe2 container image version 18.01 is based on [Caffe2 0.8.1](#).

Contents of Caffe2

This container image contains the complete source of the version of Caffe2 in /`opt/caffe2`. It is pre-built and installed into the /`opt/caffe2/[binaries,lib]` directories in the container image.

The container also includes the following:

- ▶ [Ubuntu 16.04](#)



Container image `18.01-py2` contains [Python 2.7](#); `18.01-py3` contains [Python 3.5](#).

- ▶ [NVIDIA CUDA 9.0.176](#) including [CUDA[®] Basic Linear Algebra Subroutines library[™] \(cuBLAS\) 9.0.282](#)
- ▶ [NVIDIA CUDA[®] Deep Neural Network library[™] \(cuDNN\) 7.0.5](#)
- ▶ [NCCL 2.1.2](#) (optimized for [NVLink[™]](#))
- ▶ [OpenMPI[™] 1.10.3](#)

Driver Requirements

Release 18.01 is based on CUDA 9, which requires [NVIDIA Driver](#) release 384.xx.

Key Features and Enhancements

This Caffe2 release includes the following key features and enhancements.

- ▶ Addition of Python 3 container
- ▶ Addition of ONNX package for inference

- ▶ FP16 and Tensor Core support for RNNs
- ▶ Latest version of cuBLAS
- ▶ Latest version of cuDNN
- ▶ Latest version of NCCL
- ▶ Ubuntu 16.04 with December 2017 updates

Known Issues

cuBLAS 9.0.282 regresses RNN seq2seq FP16 performance for a small subset of input sizes. As a workaround, revert back to the 11.12 container.

Chapter 12.

CAFFE2 RELEASE 17.12

The NVIDIA container image of Caffe2, release 17.12, is available.

Caffe2 container image version 17.12 is based on [Caffe2 0.8.1](#).

Contents of Caffe2

This container image contains the complete source of the version of Caffe2 in /`opt/caffe2`. It is pre-built and installed into the /`opt/caffe2/[binaries,lib]` directories in the container image.

The container also includes the following:

- ▶ Ubuntu 16.04
- ▶ NVIDIA CUDA 9.0.176 including CUDA[®] Basic Linear Algebra Subroutines library[™] (cuBLAS) 9.0.234
- ▶ NVIDIA CUDA[®] Deep Neural Network library[™] (cuDNN) 7.0.5
- ▶ NCCL 2.1.2 (optimized for NVLink[™])
- ▶ OpenMPI[™] 1.10.3

Driver Requirements

Release 17.12 is based on CUDA 9, which requires [NVIDIA Driver](#) release 384.xx.

Key Features and Enhancements

This Caffe2 release includes the following key features and enhancements.

- ▶ Latest version of CUDA
- ▶ Latest version of cuDNN
- ▶ Latest version of NCCL
- ▶ Ubuntu 16.04 with November 2017 updates

Known Issues

There are no known issues in this release.

Chapter 13.

CAFFE2 RELEASE 17.11

The NVIDIA container image of Caffe2, release 17.11, is available.

Caffe2 container image version 17.11 is based on [Caffe2 0.8.1](#).

Contents of Caffe2

This container image contains the complete source of the version of Caffe2 in /`opt/caffe2`. It is pre-built and installed into the /`opt/caffe2/[binaries,lib]` directories in the container image.

The container also includes the following:

- ▶ Ubuntu 16.04
- ▶ NVIDIA CUDA 9.0.176 including CUDA[®] Basic Linear Algebra Subroutines library[™] (cuBLAS) 9.0.234
- ▶ NVIDIA CUDA[®] Deep Neural Network library[™] (cuDNN) 7.0.4
- ▶ NCCL 2.1.2 (optimized for NVLink[™])
- ▶ OpenMPI[™] 1.10.3

Driver Requirements

Release 17.11 is based on CUDA 9, which requires [NVIDIA Driver](#) release 384.xx.

Key Features and Enhancements

This Caffe2 release includes the following key features and enhancements.

- ▶ Reworked mixed precision optimizers and fused mixed precision-aware optimization operators
- ▶ Latest version of CUDA
- ▶ Latest version of cuDNN
- ▶ Latest version of NCCL
- ▶ Ubuntu 16.04 with October 2017 updates

Known Issues

There are no known issues in this release.

Chapter 14.

CAFFE2 RELEASE 17.10

The NVIDIA container image of Caffe2, release 17.10, is available.

Caffe2 container image version 17.10 is based on [Caffe2 0.8.1](#).

Contents of Caffe2

This container image contains the complete source of the version of Caffe2 in /`opt/caffe2`. It is pre-built and installed into the `/opt/caffe2/[binaries,lib]` directories in the container image.

The container also includes the following:

- ▶ [Ubuntu 16.04](#)
- ▶ [NVIDIA CUDA 9.0](#)
- ▶ [NVIDIA CUDA[®] Deep Neural Network library[™] \(cuDNN\) 7.0.3](#)
- ▶ [NCCL 2.0.5 \(optimized for NVLink[™]\)](#)
- ▶ [OpenMPI[™] 1.10.3](#)

Driver Requirements

Release 17.10 is based on CUDA 9, which requires [NVIDIA Driver](#) release 384.xx.

Key Features and Enhancements

This Caffe2 release includes the following key features and enhancements.

- ▶ Improved data pipeline using `libturbojpeg` for faster JPEG decoding
- ▶ Enhanced and fixed bugs in `nvidia-examples/seq2seq`
- ▶ Added GradientSlice support to MultiPrecisionSGD
- ▶ Added FP16 support to gradient clipping
- ▶ Latest version of CUDA
- ▶ Latest version of cuDNN
- ▶ Latest version of NCCL

- ▶ Ubuntu 16.04 with September 2017 updates

Known Issues

There are no known issues in this release.

Chapter 15.

CAFFE2 RELEASE 17.09

The NVIDIA container image of Caffe2, release 17.09, is available.

Caffe2 container image version 17.09 is based on [Caffe2 0.8.1](#).

Contents of Caffe2

This container image contains the complete source of the version of Caffe2 in /`opt/caffe2`. It is pre-built and installed into the /`opt/caffe2/[binaries,lib]` directories in the container image.

The container also includes the following:

- ▶ [Ubuntu 16.04](#)
- ▶ [NVIDIA CUDA 9.0](#)
- ▶ [NVIDIA CUDA[®] Deep Neural Network library[™] \(cuDNN\) 7.0.2](#)
- ▶ [NCCL 2.0.5 \(optimized for NVLink[™]\)](#)
- ▶ [OpenMPI[™] 1.10.3](#)

Driver Requirements

Release 17.09 is based on CUDA 9, which requires [NVIDIA Driver](#) release 384.xx.

Key Features and Enhancements

This Caffe2 release includes the following key features and enhancements.

- ▶ Support Tensor Core Operations for Volta
- ▶ Improved the **ImageNet** training script in `nvidia-examples`
- ▶ Data input pipeline improvements - use of `libturbojpeg` directly for JPEG decode, and resize work-saving improvements.
- ▶ Improved the sequence-to-sequence (seq2seq) example in `caffe2/python/models/seq2seq`
- ▶ Added FP16 support for RNNs via the updated seq2seq example

- ▶ Latest version of CUDA
- ▶ Latest version of cuDNN
- ▶ Latest version of NCCL
- ▶ Ubuntu 16.04 with August 2017 updates

Known Issues

There are no known issues in this release.

Chapter 16.

CAFFE2 RELEASE 17.07

The NVIDIA container image of Caffe2, release 17.07, is available.

Caffe2 container image version 17.07 is based on [Caffe2 0.7.0](#).

Contents of Caffe2

This container image contains the complete source of the version of Caffe2 in `/opt/caffe2`. It is pre-built and installed into the `/opt/caffe2/[binaries,lib]` directories in the container image.

The container also includes the following:

- ▶ Ubuntu 16.04
- ▶ NVIDIA CUDA 8.0.61.2 including CUDA[®] Basic Linear Algebra Subroutines library[™] (cuBLAS) Patch 2
- ▶ NVIDIA CUDA[®] Deep Neural Network library[™] (cuDNN) 6.0.21
- ▶ NCCL 2.0.3 (optimized for NVLink[™])
- ▶ OpenMPI[™] 1.10.3

Key Features and Enhancements

This Caffe2 release includes the following key features and enhancements.

- ▶ Updated version of the `imagenet` script in the `nvidia-examples` directory.
- ▶ Ubuntu 16.04 with June 2017 updates

Known Issues

You may come across unexpected behavior, including a runtime crash, when using the `predictor_exporter.py` script to save or load models or checkpoints. However, in this release, there may be cases where the script can save or load a model incorrectly. The `imagenet` script at `nvidia-examples/imagenet/train.py` uses the `predictor_exporter.py` script, therefore, when using the `imagenet` script, you

may come across unexpected behavior. Alternatively, use the `SaveOp` and `LoadOp` operations, with the correct argument, to save and load the models and checkpoints.

Chapter 17.

CAFFE2 RELEASE 17.06

The NVIDIA container image of Caffe2, release 17.06, is available.

Caffe2 container image version 17.06 is based on [Caffe2 0.7.0](#).

Contents of Caffe2

This container image contains the complete source of the version of Caffe2 in `/opt/caffe2`. It is pre-built and installed into the `/opt/caffe2/[binaries,lib]` directories in the container image.

The container also includes the following:

- ▶ [Ubuntu 16.04](#)
- ▶ [NVIDIA CUDA 8.0.61](#)
- ▶ [NVIDIA CUDA[®] Deep Neural Network library[™] \(cuDNN\) 6.0.21](#)
- ▶ [NCCL 1.6.1 \(optimized for NVLink[™]\)](#)
- ▶ [OpenMPI[™] 1.10.3](#)

Key Features and Enhancements

This Caffe2 release includes the following key features and enhancements.

- ▶ [Ubuntu 16.04 with May 2017 updates](#)

Known Issues

The NCCL library version 1.6.1 included in this image, modifies the output buffers on all GPUs during in-place `ncclReduce()` operations, whereas normally only the "root" (target) device's output buffer should be modified. This is fixed in later versions of NCCL, as will be packaged in later versions of this image. As a workaround, either use `ncclAllReduce()`, which correctly modifies output buffers of all GPUs to the same values, or use out-of-place `ncclReduce()`, wherein the output buffer is distinct from the input buffer.

In some cases, the use of GPU transforms in the `ImageInput` operator can cause out-of-bound memory accesses. To avoid this, set `use_gpu_transform=False` in the `ImageInput` operator of your network definition.

Chapter 18.

CAFFE2 RELEASE 17.05

The NVIDIA container image of Caffe2, release 17.05, is available.

Caffe2 container image version 17.05 is based on [Caffe2 0.5.0+](#).

Contents of Caffe2

This container image contains the complete source of the version of Caffe2 in /`opt/caffe2`. It is pre-built and installed into the /`opt/caffe2/[binaries,lib]` directories in the container image.

The container also includes the following:

- ▶ [Ubuntu 16.04](#)
- ▶ [NVIDIA CUDA 8.0.61](#)
- ▶ [NVIDIA CUDA[®] Deep Neural Network library[™] \(cuDNN\) 6.0.21](#)
- ▶ [NCCL 1.6.1 \(optimized for NVLink[™]\)](#)
- ▶ [OpenMPI[™] 1.10.3](#)

Key Features and Enhancements

This Caffe2 release includes the following key features and enhancements.

- ▶ Latest cuDNN release
- ▶ Ubuntu 16.04 with April 2017 updates

Known Issues

The NCCL library version 1.6.1 included in this image, modifies the output buffers on all GPUs during in-place `ncclReduce()` operations, whereas normally only the "root" (target) device's output buffer should be modified. This is fixed in later versions of NCCL, as will be packaged in later versions of this image. As a workaround, either use `ncclAllReduce()`, which correctly modifies output buffers of all GPUs to the same

values, or use out-of-place `ncclReduce()`, wherein the output buffer is distinct from the input buffer.

Chapter 19.

CAFFE2 RELEASE 17.04

The NVIDIA container image of Caffe2, release 17.04, is available.

Caffe2 container image version 17.04 is based on [Caffe2 0.5.0+](#).

Contents of Caffe2

This container image contains the complete source of the version of Caffe2 in `/opt/caffe2`. It is pre-built and installed into the `/opt/caffe2/[binaries,lib]` directories in the container image.

The container also includes the following:

- ▶ Ubuntu 16.04
- ▶ NVIDIA CUDA 8.0.61
- ▶ NVIDIA CUDA[®] Deep Neural Network library[™] (cuDNN) 6.0.20
- ▶ NCCL 1.6.1 (optimized for NVLink[™])
- ▶ OpenMPI[™] 1.10.3

Key Features and Enhancements

This Caffe2 release includes the following key features and enhancements.

- ▶ Examples that support multi-GPU using NCCL for computer vision, such as, image classification, segmentation, and more:
 - ▶ MNIST, CIFAR-10, and IMAGENET datasets



IMAGENET is the dataset on which neural networks, such as, AlexNet, VGG, OverFeat, ResNet, and Inception are trained on. These neural networks are comprised of different depths and complexities.

The complete list of examples are in the `/workspace/nvidia-examples` directory.

- ▶ Ubuntu 16.04 with March 2017 updates

Known Issues

There are no known issues in this release.

Notice

THE INFORMATION IN THIS GUIDE AND ALL OTHER INFORMATION CONTAINED IN NVIDIA DOCUMENTATION REFERENCED IN THIS GUIDE 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 guide shall be limited in accordance with the NVIDIA terms and conditions of sale for the product.

THE NVIDIA PRODUCT DESCRIBED IN THIS GUIDE 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 guide 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 guide. 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 guide, or (ii) customer product designs.

Other than the right for customer to use the information in this guide with the product, no other license, either expressed or implied, is hereby granted by NVIDIA under this guide. Reproduction of information in this guide 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, and cuBLAS, CUDA, cuDNN, DALI, DIGITS, DGX, DGX-1, DGX-2, DGX Station, DLProf, Jetson, Kepler, Maxwell, NCCL, Nsight Compute, Nsight Systems, NvCaffe, PerfWorks, Pascal, SDK Manager, Tegra, TensorRT, TensorRT 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

© 2020 NVIDIA Corporation. All rights reserved.

www.nvidia.com

