# TABLE OF CONTENTS

Chapter 1. MXNet Overview ............................................................................................... 1
Chapter 2. Pulling A Container .......................................................................................... 2
Chapter 3. Running MXNet ............................................................................................... 3
Chapter 4. MXNet Release 18.12 .................................................................................... 5
Chapter 5. MXNet Release 18.11 .................................................................................... 7
Chapter 6. MXNet Release 18.10 .................................................................................... 9
Chapter 7. MXNet Release 18.09 ................................................................................... 11
Chapter 8. MXNet Release 18.08 ................................................................................... 13
Chapter 9. MXNet Release 18.07 ................................................................................... 15
Chapter 10. MXNet Release 18.06 ............................................................................... 17
Chapter 11. MXNet Release 18.05 ............................................................................... 19
Chapter 12. MXNet Release 18.04 ............................................................................... 21
Chapter 13. MXNet Release 18.03 ............................................................................... 23
Chapter 14. MXNet Release 18.02 ............................................................................... 25
Chapter 15. MXNet Release 18.01 ............................................................................... 27
Chapter 16. MXNet Release 17.12 ............................................................................... 29
Chapter 17. MXNet Release 17.11 ............................................................................... 31
Chapter 18. MXNet Release 17.10 ............................................................................... 33
Chapter 19. MXNet Release 17.09 ............................................................................... 35
Chapter 20. MXNet Release 17.07 ............................................................................... 37
Chapter 21. MXNet Release 17.06 ............................................................................... 38
Chapter 22. MXNet Release 17.05 ............................................................................... 39
Chapter 23. MXNet Release 17.04 ............................................................................... 40
Chapter 24. MXNet Release 17.03 ............................................................................... 41
Chapter 1.
MXNET OVERVIEW

The NVIDIA Deep Learning SDK accelerates widely-used deep learning frameworks such as MXNet™.

MXNet is a deep learning framework designed for both efficiency and flexibility. It allows you to mix the flavors of symbolic programming and imperative programming to maximize efficiency and productivity.

In its core is a dynamic dependency scheduler that automatically parallelizes both symbolic and imperative operations on the fly. A graph optimization layer on top of that makes symbolic execution fast and memory efficient. The library is portable and lightweight, and it scales to multiple GPUs and multiple machines.

MXNet is also more than a deep learning project. It is also a collection of blueprints and guidelines for building deep learning systems and interesting insights of deep learning systems for hackers.

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 and nvidia-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 nvidia-docker installation documentation based on your platform.

You must also have access and be 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 MXNET

Before running the container, use the `docker pull` command to ensure an up-to-date image is installed. Once the pull is complete, you can run the container image. This is because nvidia-docker ensures that drivers that match the host are used and configured for the container. Without nvidia-docker, you are likely to get an error when trying to run the container.

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/mxnet:18.12-py3
```

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:

```
nvidia-docker run -it --rm -v local_dir:container_dir
nvcr.io/nvidia/mxnet:<xx.xx>-py3
```

Where:

- `-it` means interactive
- `--rm` means delete the container when finished
- `-v` means mount directory
- `local_dir` is the directory or file from your host system (absolute path) that you want to access from inside your container. For example, the `local_dir` in the following path is `/home/jsmith/data/mnist`:

```
-v /home/jsmith/data/mnist:/data/mnist
```

If you are inside the container, for example, `ls /data/mnist`, you will see the same files as if you issued the `ls /home/jsmith/data/mnist` command from outside the container.

- `container_dir` is the target directory when you are inside your container. For example, `/data/mnist` is the target directory in the example:
For more information about MXNet, including tutorials, documentation, and examples, see:

- MXNet website
- MXNet project
Chapter 4.
MXNET RELEASE 18.12

The NVIDIA container image for MXNet, release 18.12, is available.

Contents of MXNet

This container image contains the complete source of the version of MXNet in /opt/mxnet. It is pre-built and installed to the Python path.

The container also includes the following:

- Ubuntu 16.04 including Python 3.5
- NVIDIA CUDA 10.0.130 including CUDA® Basic Linear Algebra Subroutines library™ (cuBLAS) 10.0.130
- NVIDIA CUDA® Deep Neural Network library™ (cuDNN) 7.4.1
- NCCL 2.3.7 (optimized for NVLink™)
- ONNX exporter 0.1 for CNN classification models

The ONNX exporter is being continuously improved. You can try the latest changes by pulling from the master branch.

- Amazon Labs Sockeye sequence-to-sequence framework 1.18.28 (for machine translation)
- TensorRT 5.0.2
- DALI 0.5.0 Beta

Driver Requirements

Release 18.12 is based on CUDA 10, which requires NVIDIA Driver release 410.xx. However, if you are running on Tesla (Tesla V100, Tesla P4, Tesla P40, or Tesla P100), you may use NVIDIA driver release 384. For more information, see CUDA Compatibility and Upgrades.
GPU Requirements

Release 18.12 supports CUDA compute capability 6.0 and higher. This corresponds to GPUs in the Pascal, Volta, and Turing families. Specifically, for a list of GPUs that this compute capability corresponds to, see CUDA GPUs. For additional support details, see Deep Learning Frameworks Support Matrix.

Key Features and Enhancements

This MXNet release includes the following key features and enhancements.

‣ MXNet container image version 18.12 is based on 1.3.0, with all upstream changes from the MXNet master branch up to and including PR 13069.
‣ Improved handling of float32 datatype in examples/image-classification/train_imagenet_runner.
‣ Enabled NVIDIA Tools Extension SDK (NVTX) instrumentation.
‣ Improved speed of metrics computation during training, especially in the case of using TopKAccuracy metric.
‣ Latest version of DALI 0.5.0 Beta.
‣ Ubuntu 16.04 with November 2018 updates

Known Issues

‣ The MXNet KVStore GPU peer-to-peer communication tree discovery, as of release 18.09, is not compatible with DGX-1V. Only users that set the environment variable MXNET_KVSTORE_USETREE=1 will experience issues, which will be resolved in a subsequent release. Issue tracked on https://github.com/apache/incubator-mxnet/issue/13341.
‣ The default setting of the environment variable MXNET_GPU_COPY_NTHREADS=1 in the container may not be optimal for all networks. Networks with a high ratio of parameters and computation, like AlexNet, may achieve greater multi-GPU training speeds with the setting MXNET_GPU_COPY_NTHREADS=2. Users are encouraged to try this setting for their own use case.
Chapter 5.
MXNET RELEASE 18.11

The NVIDIA container image for MXNet, release 18.11, is available.

Contents of MXNet

This container image contains the complete source of the version of MXNet in `/opt/mxnet`. It is pre-built and installed to the Python path.

The container also includes the following:

- Ubuntu 16.04 including Python 3.5
- NVIDIA CUDA 10.0.130 including CUDA® Basic Linear Algebra Subroutines library™ (cuBLAS) 10.0.130
- NVIDIA CUDA® Deep Neural Network library™ (cuDNN) 7.4.1
- NCCL 2.3.7 (optimized for NVLink™)
- ONNX exporter 0.1 for CNN classification models

The ONNX exporter is being continuously improved. You can try the latest changes by pulling from the master branch.

- Amazon Labs Sockeye sequence-to-sequence framework 1.18.28 (for machine translation)
- TensorRT 5.0.2
- DALI 0.4.1 Beta

Driver Requirements

Release 18.11 is based on CUDA 10, which requires NVIDIA Driver release 410.xx. However, if you are running on Tesla (Tesla V100, Tesla P4, Tesla P40, or Tesla P100), you may use NVIDIA driver release 384. For more information, see CUDA Compatibility and Upgrades.
Key Features and Enhancements

This MXNet release includes the following key features and enhancements.

‣ MXNet container image version 18.11 is based on 1.3.0, with all upstream changes from the MXNet master branch up to and including PR 12537.

‣ Added fused BatchNormAddRelu operator to the MXNet Symbol package (accessible via mx.sym.BatchNormAddRelu), which performs BatchNorm operation on data, sums the result with a tensor and performs Relu activation on the result of the sum. Currently it is limited to FP16 data type and NHWC data layout.

‣ Added MXNET_EXEC_ENABLE_ADDTO environment variable, which when set to 1 increases performance for some networks.

‣ Added MXNET_UPDATE_ON_KVSTORE environment variable, which controls whether to update parameters using KVStore (default is 1 for KVStore device and 0 for KVStore horovod).

‣ Added aggregation of SGD updates which increases performance when update on KVStore is disabled.

‣ Increased performance when training with small batch sizes.

‣ Fixed a bug that prevented matrix multiplications to overlap with other computation, which increases performance for some networks.

‣ Fixed an issue which prevented score function to respect not-full batches of data.

‣ Added resnet-v1b as possible network in the train_imagenet_runner script.

‣ Latest version of NCCL 2.3.7.

‣ Latest version of NVIDIA cuDNN 7.4.1.

‣ Latest version of TensorRT 5.0.2

‣ Latest version of DALI 0.4.1 Beta.

‣ Ubuntu 16.04 with October 2018 updates.

Known Issues

‣ The MXNet KVStore GPU peer-to-peer communication tree discovery, as of release 18.09, is not compatible with DGX-1V. Only users that set the environment variable MXNET_KVSTORE_USETREE=1 will experience issues, which will be resolved in a subsequent release.

‣ MXNet ResNet50 regresses in FP32 performance. This issue should be fixed in a later release.
Chapter 6.
MXNET RELEASE 18.10

The NVIDIA container image of MXNet, release 18.10, is available.

Contents of MXNet

This container image contains the complete source of the version of MXNet in /opt/mxnet. It is pre-built and installed to the Python path.

The container also includes the following:

- Ubuntu 16.04 including Python 3.5
- NVIDIA CUDA 10.0.130 including CUDA® Basic Linear Algebra Subroutines library™ (cuBLAS) 10.0.130
- NVIDIA CUDA® Deep Neural Network library™ (cuDNN) 7.4.0
- NCCL 2.3.6 (optimized for NVLink™)
- ONNX exporter 0.1 for CNN classification models

The ONNX exporter is being continuously improved. You can try the latest changes by pulling from the master branch.

- Amazon Labs Sockeye sequence-to-sequence framework 1.18.28 (for machine translation)
- TensorRT 5.0.0 RC
- DALI 0.4 Beta

Driver Requirements

Release 18.10 is based on CUDA 10, which requires NVIDIA Driver release 410.xx. However, if you are running on Tesla (Tesla V100, Tesla P4, Tesla P40, or Tesla P100), you may use NVIDIA driver release 384. For more information, see CUDA Compatibility and Upgrades.
Key Features and Enhancements

This MXNet release includes the following key features and enhancements.

‣ MXNet container image version 18.10 is based on 1.3.0, with all upstream changes from the MXNet master branch up to and including PR 12537.
‣ Latest version of NCCL 2.3.6.
‣ Latest version of DALI 0.4 Beta.
‣ The known issue in the prior release regarding the variable maximum GPU global memory usage has been fixed. You should now see lower and stable global memory usage from run to run, and across GPUs in multi-GPU training.
‣ Ubuntu 16.04 with September 2018 updates

Known Issues

‣ The MXNet KVStore GPU peer-to-peer communication tree discovery, as of release 18.09, is not compatible with DGX-1V. Only users that set the environment variable `MXNET_KVSTORE_USETREE=1` will experience issues, which will be resolved in a subsequent release.

‣ MXNet ResNet50 regresses in FP32 performance. This issue should be fixed in a later release.
Chapter 7.  
**MXNET RELEASE 18.09**

The NVIDIA container image of MXNet, release 18.09, is available.

**Contents of MXNet**

This container image contains the complete source of the version of MXNet in `/opt/mxnet`. It is pre-built and installed to the Python path.

The container also includes the following:

- Ubuntu 16.04 including Python 3.5
- NVIDIA CUDA 10.0 including CUDA® Basic Linear Algebra Subroutines library™ (cuBLAS) 10.0
- NVIDIA CUDA® Deep Neural Network library™ (cuDNN) 7.3.0
- NCCL 2.3.4 (optimized for NVLink™)
- ONNX exporter 0.1 for CNN classification models

The ONNX exporter is being continuously improved. You can try the latest changes by pulling from the master branch.

- Amazon Labs Sockeye sequence-to-sequence framework 1.18.28 (for machine translation)
- TensorRT 5.0.0 RC
- DALI 0.2 Beta

**Driver Requirements**

Release 18.09 is based on CUDA 10, which requires NVIDIA Driver release 410.xx. However, if you are running on Tesla (Tesla V100, Tesla P4, Tesla P40, or Tesla P100), you may use NVIDIA driver release 384. For more information, see CUDA Compatibility and Upgrades.
Key Features and Enhancements

This MXNet release includes the following key features and enhancements.

- MXNet container image version 18.09 is based on 1.3.0, with all upstream changes from the MXNet master branch up to the creation point of the v1.3.x branch (PR 12301), plus all substantive cherry-picks from master that were included in the v1.3.0 release.
- The demonstrator of mixed precision ResNet-50 training using the NHWC data layout has been expanded to work now on the Turing architecture in addition to Volta.
- Latest version of cuDNN 7.3.0.
- Latest version of CUDA 10.0 which includes support for DGX-2, Turing, and Jetson Xavier.
- Latest version of cuBLAS 10.0.
- Latest version of NCCL 2.3.4.
- Latest version of TensorRT 5.0.0 RC.
- Latest version of DALI 0.2 Beta.
- Ubuntu 16.04 with August 2018 updates

Known Issues

The multi-threaded nature of MXNet model execution may result in a variable maximum usage of GPU global memory, as discussed in earlier release notes. Users that experience sporadic out-of-GPU-memory errors should experiment with setting the environment variable `MXNET_GPU_WORKER_NTHREADS=1` as a possible remedy. We anticipate the need for this experimentation will be removed in our next release.
Chapter 8. 
MXNET RELEASE 18.08

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

Contents of MXNet

This container image contains the complete source of the version of MXNet in /opt/mxnet. It is pre-built and installed to the Python path.

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™)
- ONNX exporter 0.1 for CNN classification models

The ONNX exporter is being continuously improved. You can try the latest changes by pulling from the master branch.

- Amazon Labs Sockeye sequence-to-sequence framework 1.18.28 (for machine translation)
- TensorRT 4.0.1
- DALI 0.1.2 Beta

Driver Requirements

Release 18.08 is based on CUDA 9, which requires NVIDIA Driver release 384.xx.
Key Features and Enhancements

This MXNet release includes the following key features and enhancements.

‣ MXNet container image version 18.08 is based on 1.2.0, with all upstream changes from the MXNet master branch up to and including PR 11545.
‣ Latest version of cuDNN 7.2.1.
‣ Latest version of DALI 0.1.2 Beta.
‣ New demonstrator of increased mixed-precision ResNet-50 training speeds on Volta when processed end-to-end in the NHWC data layout. We are working to PR the code improvements to upstream MXNet. To evaluate in the meantime, type /opt/mxnet/examples/image_classification/trainImagenet_runner --batch-size N. Substitute 256 for N on systems with GPUs having 32GB global memory (or 192 with 16GB GPUs) and prepare the imagenet database as directed in nvidia-examples/imagenet_preparations. Training images should be pre-resized to 480px shorter side and validation should be pre-resized to 256px shorter side. The script expects RecordIO files to be present in the /data/imagenet/train-480-val-256-recordio/ directory.
‣ Ubuntu 16.04 with July 2018 updates

Announcements

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

Known Issues

The multi-threaded nature of MXNet model execution may result in a variable maximum usage of GPU global memory. Users that experience sporadic out-of-GPU-memory errors should experiment with setting the environment variable MXNET_GPU_WORKER_NTHREADS=1 as a possible remedy. We anticipate the need for this experimentation will be removed in a subsequent release.
The NVIDIA container image of MXNet, release 18.07, is available.

**Contents of MXNet**

This container image contains the complete source of the version of MXNet in `/opt/mxnet`. It is pre-built and installed to the Python path.

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™)
  - ONNX exporter 0.1 for CNN classification models

  The ONNX exporter is being continuously improved. You can try the latest changes by pulling from the master branch.

- Amazon Labs Sockeye sequence-to-sequence framework 1.18.23 (for machine translation)
- TensorRT 4.0.1
- DALI 0.1 Beta

**Driver Requirements**

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

This MXNet release includes the following key features and enhancements.

‣ MXNet container image version 18.07 is based on 1.2.0, with all upstream changes from the MXNet master branch up to and including PR 11302.
‣ Added support for DALI 0.1 Beta.
‣ Latest version of CUDA® Basic Linear Algebra Subroutines library™ (cuBLAS) 9.0.425.
‣ Ubuntu 16.04 with June 2018 updates

Announcements

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

Known Issues

‣ Some of the unit tests available in /opt/mxnet/tests/python/{gpu,unittest}/*.py require the SciPy Python library. For those that want to run the unit tests, first install SciPy by pip install scipy. There is no longer a need to specifically request the 1.0 version of SciPy.
‣ The multi-threaded nature of MXNet model execution may result in a variable maximum usage of GPU global memory. Users that experience sporadic out-of-GPU-memory errors should experiment with setting the environment variable MXNET_GPU_WORKER_NTHREADS=1 as a possible remedy. We anticipate the need for this experimentation will be removed in a subsequent release.
Chapter 10.  
MXNET RELEASE 18.06

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

Contents of MXNet

This container image contains the complete source of the version of MXNet in /opt/mxnet. It is pre-built and installed to the Python path.

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™)
- ONNX exporter 0.1 for CNN classification models
  - The ONNX exporter is being continuously improved. You can try the latest changes by pulling from the master branch.
- Amazon Labs Sockeye sequence-to-sequence framework 1.18.22 (for machine translation)
- TensorRT 4.0.1

Driver Requirements

Release 18.06 is based on CUDA 9, which requires NVIDIA Driver release 384.xx.
Key Features and Enhancements

This MXNet release includes the following key features and enhancements.

- MXNet container image version 18.06 is based on 1.2.0. Specifically, container image 18.06 has merged all commits on upstream MXNet master, up to the creation point of the v1.2.0 branch, and all commits on that branch up to the 1.2.0 tag.
- Container includes TensorRT 4.0.1
- TensorRT integration examples for in-framework inference can be found in /workspace/examples/tensorrt-integration. This includes a LeNet-5 unit test and a ResNet-50 example.
- Support added for DALI iterators.
- Ubuntu 16.04 with May 2018 updates

Announcements

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

Known Issues

Some of the unit tests available in /opt/mxnet/tests/python/{gpu,unittest}/*.py require the SciPy Python library. For those that want to run the unit tests, first install the 1.0 version of SciPy by typing pip install scipy==1.0.

The latest SciPy release, version 1.1, is not compatible with the unit tests.
Chapter 11.
MXNET RELEASE 18.05

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

Contents of MXNet

This container image contains the complete source of the version of MXNet in `/opt/mxnet`. It is pre-built and installed to the Python path.

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™)**

- **ONNX exporter 0.1 for CNN classification models**

  - The ONNX exporter is being continuously improved. You can try the latest changes by pulling from the master branch.

- **Amazon Labs Sockeye sequence-to-sequence framework 1.18.13 (for machine translation)**

Driver Requirements

Release 18.05 is based on CUDA 9, which requires NVIDIA Driver release 384.xx.

Key Features and Enhancements

This MXNet release includes the following key features and enhancements.
MXNet container image version 18.05 is based on MXNet 1.1.0.

For this month, no upstream merges as we work toward incorporating the upcoming MXNet 1.2.0 release.

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.

Known Issues

Those wishing to run the MXNet unit tests under /opt/mxnet/tests/python should install SciPy using `pip install scipy==1.0`, as the recently available SciPy v1.1 is not compatible with all the unit tests. For more information, see Broken test_sparse_operator.test_sparse_mathematical_core with scipy 1.1.0.
The NVIDIA container image of MXNet, release 18.04, is available.

**Contents of MXNet**

This container image contains the complete source of the version of MXNet in `/opt/mxnet`. It is pre-built and installed to the Python path.

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™)

- ONNX exporter 0.1 for CNN classification models

  - The ONNX exporter is being continuously improved. You can try the latest changes by pulling from the master branch.

- Amazon Labs Sockeye sequence-to-sequence framework 1.18.1 (for machine translation)

**Driver Requirements**

Release 18.04 is based on CUDA 9, which requires NVIDIA Driver release 384.xx.

**Key Features and Enhancements**

This MXNet release includes the following key features and enhancements.
MXNet container image version 18.04 is based on MXNet 1.1.0.

For this month, no upstream merges as we develop a performant approach to MXNet’s recent operator refactoring to a stateless imperative style.

ResNet-50 performance improvement based on the automatic fusion of `add_relu` and `copy_split` backward pass.

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

**Contents of MXNet**

This container image contains the complete source of the version of MXNet in /opt/mxnet. It is pre-built and installed to the Python path.

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™)
- ONNX exporter 0.1 for CNN classification models

The ONNX exporter is being continuously improved. You can try the latest changes by pulling from the master branch.

- Amazon Labs Sockeye sequence-to-sequence framework 1.17.4 (for machine translation)

**Driver Requirements**

Release 18.03 is based on CUDA 9, which requires NVIDIA Driver release 384.xx.

**Key Features and Enhancements**

This MXNet release includes the following key features and enhancements.
MXNet container image version 18.03 is based on MXNet 1.1.0.

Incorporated all upstream changes from the MXNet master branch, specifically, PR 9749.

Added compute-graph optimizations for improved ResNet performance.

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**

There are no known issues in this release.
Chapter 14.
MXNET RELEASE 18.02

The NVIDIA container image of MXNet, release 18.02, is available.
MXNet container image version 18.02 is based on MXNet 1.0.0.

Contents of MXNet

This container image contains the complete source of the version of MXNet in /opt/mxnet. It is pre-built and installed to the Python path.

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™)
  - ONNX exporter 0.1 for CNN classification models

  The ONNX exporter is being continuously improved. You can try the latest changes by pulling from the master branch.

- Amazon Labs Sockeye sequence-to-sequence framework 1.17.0 (for machine translation)
Driver Requirements

Release 18.02 is based on CUDA 9, which requires NVIDIA Driver release 384.xx.

Key Features and Enhancements

This MXNet release includes the following key features and enhancements.

‣ 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 15.
MXNET RELEASE 18.01

The NVIDIA container image of MXNet, release 18.01, is available.
MXNet container image version 18.01 is based on MXNet 1.0.0.

Contents of MXNet
This container image contains the complete source of the version of MXNet in `/opt/mxnet`. It is pre-built and installed to the Python path.
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™)
- ONNX exporter 0.1 for CNN classification models

  The ONNX exporter is being continuously improved. You can try the latest changes by pulling from the master branch.

- Amazon Labs Sockeye sequence-to-sequence framework 1.16.2 (for machine translation)

Driver Requirements
Release 18.01 is based on CUDA 9, which requires NVIDIA Driver release 384.xx.
Key Features and Enhancements

This MXNet release includes the following key features and enhancements.

▶ Addition of Python 3 package
▶ Enhanced-performance cuDNN-based batched 1D convolutions (merged to upstream)
▶ Added **MxNet-to-ONNX** exporter for classification of CNN models (tested with LeNet-5, ResNet-50, etc.).
▶ Added the Sockeye sequence-to-sequence framework, along with a German-to-English translation model, based on the WMT’15 dataset and translation task. This model’s launch script should reproduce the OpenNMT reference model when trained until convergence.
▶ 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 16.
MXNET RELEASE 17.12

The NVIDIA container image of MXNet, release 17.12, is available.
MXNet container image version 17.12 is based on MXNet 1.0.0.

Contents of MXNet

This container image contains the complete source of the version of MXNet in /opt/mxnet. It is pre-built and installed to the Python path.

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™)

Driver Requirements

Release 17.12 is based on CUDA 9, which requires NVIDIA Driver release 384.xx.

Key Features and Enhancements

This MXNet release includes the following key features and enhancements.

‣ Both the nccl and nccl_allreduce KVStore options now have the same, improved performance.
‣ 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.
The NVIDIA container image of MXNet, release 17.11, is available.

MXNet container image version 17.11 is based on MXNet 0.12.0.

Contents of MXNet

This container image contains the complete source of the version of MXNet in `/opt/mxnet`. It is pre-built and installed to the Python path.

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™)

Driver Requirements

Release 17.11 is based on CUDA 9, which requires NVIDIA Driver release 384.xx.

Key Features and Enhancements

This MXNet release includes the following key features and enhancements.

- Added Sockeye to the container, including NCCL `kvstore` option
- Enabled `mx.sym.batch_dot` to use FP16
- 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 18.
MXNET RELEASE 17.10

The NVIDIA container image of MXNet, release 17.10, is available.
MXNet container image version 17.10 is based on MXNet 0.11.0.

Contents of MXNet
This container image contains the complete source of the version of MXNet in /opt/mxnet. It is pre-built and installed to the Python path.
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™)

Driver Requirements
Release 17.10 is based on CUDA 9, which requires NVIDIA Driver release 384.xx.

Key Features and Enhancements
This MXNet release includes the following key features and enhancements.
  • Mixed precision support for all optimizers
  • New image input pipeline with faster speed and support for global shuffling after each epoch when used with IndexedRecordIO format.
  • 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 19.
MXNET RELEASE 17.09

The NVIDIA container image of MXNet, release 17.09, is available. MXNet container image version 17.09 is based on MXNet 0.11.0.rc3.

Contents of MXNet

This container image contains the complete source of the version of MXNet in `/opt/mxnet`. It is pre-built and installed to the Python path.

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™)

Driver Requirements

Release 17.09 is based on CUDA 9, which requires NVIDIA Driver release 384.xx.

Key Features and Enhancements

This MXNet release includes the following key features and enhancements.

- Tensor Core support in convolutions, deconvolutions, and fully connected layers on Volta
- Support for mixed precision training with SGD optimizer
- Streamlined FP16 examples for image classification
- Optimized input pipeline for image processing
- 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 20.
MXNET RELEASE 17.07

The NVIDIA container image of MXNet, release 17.07, is available.
MXNet container image version 17.07 is based on MXNet 0.10.0.

Contents of MXNet
This container image contains the complete source of the version of MXNet in /opt/mxnet. It is pre-built and installed to the Python path.

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™)

Key Features and Enhancements
This MXNet release includes the following key features and enhancements.

- Support for multi-precision SGD
- cuBLAS back-end for **FullyConnected** operation
- Ubuntu 16.04 with June 2017 updates

Known Issues
There are no known issues in this release.
Chapter 21.
MXNET RELEASE 17.06

The NVIDIA container image of MXNet, release 17.06, is available.
MXNet container image version 17.06 is based on MXNet 0.10.0.

Contents of MXNet
This container image contains the complete source of the version of MXNet in `/opt/mxnet`. It is pre-built and installed to the Python path.

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™)

Key Features and Enhancements
This MXNet release includes the following key features and enhancements.

- Implemented double buffering in ResNet v1 example
- Ubuntu 16.04 with May 2017 updates

Known Issues
There are no known issues in this release.
Chapter 22.
MXNET RELEASE 17.05

The NVIDIA container image of MXNet, release 17.05, is available. 
MXNet container image version 17.05 is based on MXNet 0.9.3a+.

Contents of MXNet

This container image contains the complete source of the version of MXNet in /opt/mxnet. It is pre-built and installed to the Python path.

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™)

Key Features and Enhancements

This MXNet release includes the following key features and enhancements.

- Latest cuDNN release
- Improved IO pipeline for increased multi-GPU performance
- Optimized SGD weight update
- Added the nccl_allreduce option for gradient communication
- Added support for dilated deconvolutions
- Improved convolutional neural network (CNN) performance by removing unnecessary computations
- Added options to show the cuDNN algorithms that are chosen for convolutions
- Ubuntu 16.04 with April 2017 updates

Known Issues

There are no known issues in this release.
Chapter 23.  
MXNET RELEASE 17.04

The NVIDIA container image of MXNet, release 17.04, is available. 
MXNet container image version 17.04 is based on MXNet 0.9.3a+.

Contents of MXNet
This container image contains the complete source of the version of MXNet in /opt/mxnet. It is pre-built and installed to the Python path.
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™)

Key Features and Enhancements
This MXNet release includes the following key features and enhancements.
  ▶ Support for cuDNN accelerated dilated convolutions
  ▶ Ubuntu 16.04 with March 2017 updates

Known Issues
There are no known issues in this release.
The NVIDIA container image of MXNet, release 17.03, is available.

**MXNet container image version 17.03** is based on MXNet 0.9.3.

### Contents of MXNet

This container image contains the complete source of the version of MXNet in `/opt/mxnet`. It is pre-built and installed to the Python path.

The container also includes the following:

- Ubuntu 16.04
- NVIDIA CUDA® 8.0.61
- NVIDIA CUDA® Deep Neural Network library™ (cuDNN) 6.0.20

### Key Features and Enhancements

This MXNet release includes the following key features and enhancements.

- Ubuntu 16.04 with February 2017 updates
- Improved input pipeline for image processing
- Support for FP16 training of AlexNet
- Optimized embedding layer of CUDA kernels
- Optimized tensor broadcast and reduce CUDA kernels

### 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, cuFFT, cuSPARSE, DIGITS, DGX, DGX-1, DGX Station, GRID, Jetson, Kepler, NVIDIA GPU Cloud, Maxwell, NCCL, NVLink, Pascal, Tegra, TensorRT, Tesla 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 NVIDIA Corporation. All rights reserved.

www.nvidia.com