Chapter 1. MICROSOFT COGNITIVE TOOLKIT OVERVIEW

The NVIDIA Deep Learning SDK accelerates widely-used deep learning frameworks such as the Microsoft® Cognitive Toolkit™, formerly referred to as CNTK.

The Microsoft Cognitive Toolkit empowers you to harness the intelligence within massive datasets through deep learning by providing uncompromised scaling, speed and accuracy with commercial-grade quality and compatibility with the programming languages and algorithms you already use.

This document describes the key features, software enhancements and improvements, any known issues, and how to run this container.
Chapter 2.
PULLING A CONTAINER

You can access NVIDIA’s GPU accelerated containers for all three products, the NVIDIA DGX-1™, NVIDIA DGX Station™, and the NVIDIA® GPU Cloud™ (NGC). If you own a DGX-1 or DGX Station then you should use the NVIDIA® DGX™ container registry at https://compute.nvidia.com. This is a web interface to the Docker hub, nvcr.io (NVIDIA DGX container registry). You can pull the containers from there and you can also push containers there into your own account in the registry.

If you are accessing the NVIDIA containers from the NVIDIA® GPU Cloud™ (NGC) container registry via a cloud services provider such as Amazon® Web Services™ (AWS), then you should use the NGC container registry at https://ngc.nvidia.com. This is also a web interface to the same Docker repository as for the DGX-1 and DGX Station. After you create an account, the commands to pull containers are the same as if you had a DGX-1 in your own data center. However, currently, you cannot save any containers to the NGC container registry. Instead, you have to save the containers to your own Docker repository that is either on-premise or in the Cloud.

The containers are exactly the same, whether you pull them from the NVIDIA DGX container registry or the NGC container registry.

For all three products, the DGX-1, DGX Station, and the NVIDIA NGC Cloud Services, the location of the framework source is in /opt/<framework> in the container.

Before you can pull a container from the NGC container registry, you must have Docker and nvidia-docker installed as explained in Preparing to use NVIDIA Containers Getting Started Guide. You must also have access and logged into the NGC container registry as explained in the NGC Getting Started Guide.

For step-by-step instructions, see Container User Guide.
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.

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.

   \[
   \text{docker pull nvcr.io/nvidia/cntk:18.07}
   \]

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:

   \[
   \text{nvidia-docker run -it --rm -v local_dir:container_dir nvcr.io/nvidia/cntk:<xx.xx>}
   \]

   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:

   \[
   -v /home/jsmith/data/mnist:/data/mnist
   \]

   - `<xx.xx>` is the tag. For example, 18.01.
a) When running on a single GPU, the Microsoft Cognitive Toolkit can be invoked using a command similar to the following:

```
cntk configFile=myscript.cntk ...
```

b) When running on multiple GPUs, run the Microsoft Cognitive Toolkit through MPI. The following example uses 4 GPUs, numbered 0..3, for training:

```
export OMP_NUM_THREADS=10
export CUDA_VISIBLE_DEVICES=0,1,2,3
mpirun --allow-run-as-root --oversubscribe --npernode 4 \ -x OMP_NUM_THREADS -x CUDA_VISIBLE_DEVICES \   -x CUDA_DEVICE_ORDER -x
    cntk configFile=myscript.cntk ...
```

c) When running with all 8 GPUs together, it is even more simple:

```
export OMP_NUM_THREADS=10
mpirun --allow-run-as-root --oversubscribe --npernode 8 \ -x OMP_NUM_THREADS cntk configFile=myscript.cntk ...
```

You can vary the number of GPUs with the option `--npernode X` where `X` is the number of GPUs. For the DGX-1™ this is a maximum of 8 GPUs per node. For the DGX Station™ it is a maximum of 4 GPUs. For NVIDIA® GPU Cloud™ (NGC) the number of GPUs depends upon the instance type that you have selected.

You might want to pull in data and model descriptions from locations outside the container for use by Microsoft Cognitive Toolkit 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.

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 `nvidia-docker run`.

4. See `/workspace/README.md` inside the container for information on customizing your the DGX-1 image.

For more information about the Microsoft Cognitive Toolkit, including tutorials, documentation, and examples, see the Microsoft Cognitive Toolkit wiki.
Chapter 4.
MICROSOFT COGNITIVE TOOLKIT RELEASE 18.07

The NVIDIA container image of Microsoft Cognitive Toolkit, release 18.07, is available.

Contents of the Microsoft Cognitive Toolkit

This container image contains the source files that are used to build the Microsoft Cognitive Toolkit in the /opt/cntk directory. The Microsoft Cognitive Toolkit binaries are pre-built and installed into the /usr/local/cntk directory.

A Python 3.6 Conda environment (including the Microsoft Cognitive Toolkit module) is installed in the /opt/conda/envs/cntk-py36 directory. This is included in PATH, so by default the python command executes a Python 3.6 interpreter. To use the base Ubuntu Python 2.7 binary, you can either adjust PATH or explicitly call /usr/bin/python.

This image of the Microsoft Cognitive Toolkit 2.5, is built with the following options enabled:

- Async SGD
- Python 3 API
- OpenCV
- Kaldi

The container also includes the following:

- Ubuntu 16.04 including Python 3.6 environment
- 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™ 3.0.0
Driver Requirements

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

Key Features and Enhancements

This Microsoft Cognitive Toolkit release includes the following key features and enhancements.

- Microsoft Cognitive Toolkit container image version 18.07 is based on Microsoft Cognitive Toolkit 2.5.
- Latest version of CUDA® Basic Linear Algebra Subroutines library™ (cuBLAS) 9.0.425.
- Ubuntu 16.04 with June 2018 updates

Known Issues

There are no known issues in this release.
Chapter 5.
MICROSOFT COGNITIVE TOOLKIT RELEASE 18.06

The NVIDIA container image of Microsoft Cognitive Toolkit, release 18.06, is available.

Contents of the Microsoft Cognitive Toolkit

This container image contains the source files that are used to build the Microsoft Cognitive Toolkit in the `/opt/cntk` directory. The Microsoft Cognitive Toolkit binaries are pre-built and installed into the `/usr/local/cntk` directory.

A Python 3.6 Conda environment (including the Microsoft Cognitive Toolkit module) is installed in the `/opt/conda/envs/cntk-py36` directory. This is included in PATH, so by default the `python` command executes a Python 3.6 interpreter. To use the base Ubuntu Python 2.7 binary, you can either adjust PATH or explicitly call `/usr/bin/python`.

This image of the Microsoft Cognitive Toolkit 2.5, is built with the following options enabled:

- Async SGD
- Python 3 API
- OpenCV
- Kaldi

The container also includes the following:

- Ubuntu 16.04 including Python 3.6 environment
- 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™ 3.0.0
Driver Requirements

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

Key Features and Enhancements

This Microsoft Cognitive Toolkit release includes the following key features and enhancements.

‣ Microsoft Cognitive Toolkit container image version 18.06 is based on Microsoft Cognitive Toolkit 2.5.
‣ Ubuntu 16.04 with May 2018 updates

Known Issues

There are no known issues in this release.
Chapter 6.
MICROSOFT COGNITIVE TOOLKIT RELEASE 18.05

The NVIDIA container image of Microsoft Cognitive Toolkit, release 18.05, is available.

Contents of the Microsoft Cognitive Toolkit

This container image contains the source files that are used to build the Microsoft Cognitive Toolkit in the /opt/cntk directory. The Microsoft Cognitive Toolkit binaries are pre-built and installed into the /usr/local/cntk directory.

A Python 3.6 Conda environment (including the Microsoft Cognitive Toolkit module) is installed in the /opt/conda/envs/cntk-py36 directory. This is included in PATH, so by default the python command executes a Python 3.6 interpreter. To use the base Ubuntu Python 2.7 binary, you can either adjust PATH or explicitly call /usr/bin/python.

This image of the Microsoft Cognitive Toolkit 2.5, is built with the following options enabled:

- Async SGD
- Python 3 API
- OpenCV
- Kaldi

The container also includes the following:

- Ubuntu 16.04 including Python 3.6 environment
- 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™ 3.0.0
**Driver Requirements**

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

**Key Features and Enhancements**

This Microsoft Cognitive Toolkit release includes the following key features and enhancements.

- Microsoft Cognitive Toolkit container image version 18.05 is based on Microsoft Cognitive Toolkit 2.5.
- Incorporated all upstream GitHub Microsoft Cognitive Toolkit code as of Microsoft Cognitive Toolkit 2.5.
- Ubuntu 16.04 with April 2018 updates

**Known Issues**

1-bit SGD is now built by default since it is now part of Microsoft Cognitive Toolkit. There is a known error when running OpenMPI 3.0. If using 1-bit SGD, downgrade your OpenMPI version to 1.10.
Chapter 7.
MICROSOFT COGNITIVE TOOLKIT RELEASE 18.04

The NVIDIA container image of Microsoft Cognitive Toolkit, release 18.04, is available.

Contents of the Microsoft Cognitive Toolkit

This container image contains the source files that are used to build the Microsoft Cognitive Toolkit in the `/opt/cntk` directory. The Microsoft Cognitive Toolkit binaries are pre-built and installed into the `/usr/local/cntk` directory.

A Python 3.6 Conda environment (including the Microsoft Cognitive Toolkit module) is installed in the `/opt/conda/envs/cntk-py36` directory. This is included in PATH, so by default the `python` command executes a Python 3.6 interpreter. To use the base Ubuntu Python 2.7 binary, you can either adjust PATH or explicitly call `/usr/bin/python`.

This image of the Microsoft Cognitive Toolkit 2.4, is built with the following options enabled:

- Async SGD
- Python 3 API
- OpenCV
- Kaldi

The container also includes the following:

- Ubuntu 16.04 including Python 3.6 environment
- 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™ 3.0.0
**Driver Requirements**

Release 18.04 is based on CUDA 9, which requires [NVIDIA Driver](https://www.nvidia.com) release 384.xx.

**Key Features and Enhancements**

This Microsoft Cognitive Toolkit release includes the following key features and enhancements.

- Microsoft Cognitive Toolkit container image version 18.04 is based on Microsoft Cognitive Toolkit 2.4.
- Incorporated all upstream GitHub Microsoft Cognitive Toolkit code as of Microsoft Cognitive Toolkit 2.4.
- Latest version of NCCL 2.1.15
- Ubuntu 16.04 with March 2018 updates

**Known Issues**

There are no known issues in this release.
The NVIDIA container image of Microsoft Cognitive Toolkit, release 18.03, is available.

Contents of the Microsoft Cognitive Toolkit

This container image contains the source files that are used to build the Microsoft Cognitive Toolkit in the /opt/cntk directory. The Microsoft Cognitive Toolkit binaries are pre-built and installed into the /usr/local/cntk directory.

A Python 3.6 Conda environment (including the Microsoft Cognitive Toolkit module) is installed in the /opt/conda/envs/cntk-py36 directory. This is included in PATH, so by default the python command executes a Python 3.6 interpreter. To use the base Ubuntu Python 2.7 binary, you can either adjust PATH or explicitly call /usr/bin/python.

This image of the Microsoft Cognitive Toolkit 2.4, is built with the following options enabled:

- Async SGD
- Python 3 API
- OpenCV
- Kaldi

The container also includes the following:

- Ubuntu 16.04 including Python 3.6 environment
- 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™ 3.0.0
Driver Requirements

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

Key Features and Enhancements

This Microsoft Cognitive Toolkit release includes the following key features and enhancements.

‣ Microsoft Cognitive Toolkit container image version 18.03 is based on Microsoft Cognitive Toolkit 2.4.
‣ Added support for Python 3.6. Python 3.4 is now deprecated.
‣ Added support for FP16 natively with Microsoft Cognitive Toolkit upstream version 2.4.
‣ Latest version of cuBLAS 9.0.333
‣ Latest version of cuDNN 7.1.1
‣ Ubuntu 16.04 with February 2018 updates

Known Issues

There are no known issues in this release.
Chapter 9.
MICROSOFT COGNITIVE TOOLKIT RELEASE 18.02

The NVIDIA container image of Microsoft Cognitive Toolkit, release 18.02, is available.

Microsoft Cognitive Toolkit container image version 18.02 is based on Microsoft Cognitive Toolkit 2.3.1.

Contents of the Microsoft Cognitive Toolkit

This container image contains the source files that are used to build the Microsoft Cognitive Toolkit in the /opt/cntk directory. The Microsoft Cognitive Toolkit binaries are pre-built and installed into the /usr/local/cntk directory.

A Python 3.4 Conda environment (including the Microsoft Cognitive Toolkit module) is installed in the /opt/conda/envs/cntk-py34 directory. This is included in PATH, so by default the python command executes a Python 3.4 interpreter. To use the base Ubuntu Python 2.7 binary, you can either adjust PATH or explicitly call /usr/bin/python.

This image of the Microsoft Cognitive Toolkit 2.3.1, is built with the following options enabled:

- Async SGD
- Python 3 API
- OpenCV
- Kaldi

The container also includes the following:

- Ubuntu 16.04 including Python 3.4 environment
- 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™ 3.0.0

**Driver Requirements**

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

**Key Features and Enhancements**

This Microsoft Cognitive Toolkit 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 10. MICROSOFT COGNITIVE TOOLKIT RELEASE 18.01

The NVIDIA container image of Microsoft Cognitive Toolkit, release 18.01, is available.

Microsoft Cognitive Toolkit container image version 18.01 is based on Microsoft Cognitive Toolkit 2.3.1.

Contents of the Microsoft Cognitive Toolkit

This container image contains the source files that are used to build the Microsoft Cognitive Toolkit in the /opt/cntk directory. The Microsoft Cognitive Toolkit binaries are pre-built and installed into the /usr/local/cntk directory.

A Python 3.4 Conda environment (including the Microsoft Cognitive Toolkit module) is installed in the /opt/conda/envs/cntk-py34 directory. This is included in PATH, so by default the python command executes a Python 3.4 interpreter. To use the base Ubuntu Python 2.7 binary, you can either adjust PATH or explicitly call /usr/bin/python.

This image of the Microsoft Cognitive Toolkit 2.3.1, is built with the following options enabled:

- Async SGD
- Python 3 API
- OpenCV
- Kaldi

The container also includes the following:

- Ubuntu 16.04 including Python 3.4 environment
- 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™ 3.0.0

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

Key Features and Enhancements
This Microsoft Cognitive Toolkit release includes the following key features and enhancements.

- Preview mixed precision training support and ResNet-50 ImageNet example
- 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 11.
MICROSOFT COGNITIVE TOOLKIT RELEASE 17.12

The NVIDIA container image of Microsoft Cognitive Toolkit, release 17.12, is available. Microsoft Cognitive Toolkit container image version 17.12 is based on Microsoft Cognitive Toolkit 2.2.

Contents of the Microsoft Cognitive Toolkit

This container image contains the source files that are used to build the Microsoft Cognitive Toolkit in the `/opt/cntk` directory. The Microsoft Cognitive Toolkit binaries are pre-built and installed into the `/usr/local/cntk` directory.

A Python 3.4 Conda environment (including the Microsoft Cognitive Toolkit module) is installed in the `/opt/conda/envs/cntk-py34` directory. This is included in `PATH`, so by default the `python` command executes a Python 3.4 interpreter. To use the base Ubuntu Python 2.7 binary, you can either adjust `PATH` or explicitly call `/usr/bin/python`.

This image of the Microsoft Cognitive Toolkit 2.2, is built with the following options enabled:

- Async SGD
- Python 3 API
- OpenCV
- Kaldi

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™ 3.0.0

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

Key Features and Enhancements
This Microsoft Cognitive Toolkit 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 12.
MICROSOFT COGNITIVE TOOLKIT RELEASE 17.11

The NVIDIA container image of Microsoft Cognitive Toolkit, release 17.11, is available.

Microsoft Cognitive Toolkit container image version 17.11 is based on Microsoft Cognitive Toolkit 2.2.

Contents of the Microsoft Cognitive Toolkit

This container image contains the source files that are used to build the Microsoft Cognitive Toolkit in the /opt/cntk directory. The Microsoft Cognitive Toolkit binaries are pre-built and installed into the /usr/local/cntk directory.

A Python 3.4 Conda environment (including the Microsoft Cognitive Toolkit module) is installed in the /opt/conda/envs/cntk-py34 directory. This is included in PATH, so by default the python command executes a Python 3.4 interpreter. To use the base Ubuntu Python 2.7 binary, you can either adjust PATH or explicitly call /usr/bin/python.

This image of the Microsoft Cognitive Toolkit 2.2, is built with the following options enabled:

- Async SGD
- Python 3 API
- OpenCV
- Kaldi

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™ 3.0.0

Driver Requirements

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

Key Features and Enhancements

This Microsoft Cognitive Toolkit release includes the following key features and enhancements.

- The version of MPI included in the container was bumped from 1.10 to 3.0.0 in order to correct a bug in which GPU memory exhaustion could cause the framework to hang.
- 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 13.
MICROSOFT COGNITIVE TOOLKIT RELEASE 17.10

The NVIDIA container image of Microsoft Cognitive Toolkit, release 17.10, is available. Microsoft Cognitive Toolkit container image version 17.10 is based on Microsoft Cognitive Toolkit 2.2.

Contents of the Microsoft Cognitive Toolkit

This container image contains the source files that are used to build the Microsoft Cognitive Toolkit in the /opt/cntk directory. The Microsoft Cognitive Toolkit binaries are pre-built and installed into the /usr/local/cntk directory.

A Python 3.4 Conda environment (including the Microsoft Cognitive Toolkit module) is installed in the /opt/conda/envs/cntk-py34 directory. This is included in PATH, so by default the python command executes a Python 3.4 interpreter. To use the base Ubuntu Python 2.7 binary, you can either adjust PATH or explicitly call /usr/bin/python.

This image of the Microsoft Cognitive Toolkit 2.2, is built with the following options enabled:

- Async SGD
- Python 3 API
- OpenCV
- Kaldi

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 Microsoft Cognitive Toolkit release includes the following key features and enhancements.

‣ Auto padding for convolutional layers correctly accounts for dilation
‣ 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 14.
MICROSOFT COGNITIVE TOOLKIT RELEASE 17.09

The NVIDIA container image of Microsoft Cognitive Toolkit, release 17.09, is available. Microsoft Cognitive Toolkit container image version 17.09 is based on Microsoft Cognitive Toolkit 2.1.

Contents of the Microsoft Cognitive Toolkit

This container image contains the source files that are used to build the Microsoft Cognitive Toolkit in the /opt/cntk directory. The Microsoft Cognitive Toolkit binaries are pre-built and installed into the /usr/local/cntk directory.

A Python 3.4 Conda environment (including the Microsoft Cognitive Toolkit module) is installed in the /opt/conda/envs/cntk-py34 directory. This is included in PATH, so by default the python command executes a Python 3.4 interpreter. To use the base Ubuntu Python 2.7 binary, you can either adjust PATH or explicitly call /usr/bin/python.

This image of the Microsoft Cognitive Toolkit 2.1, is built with the following options enabled:

- Async SGD
- Python 3 API
- OpenCV
- Kaldi

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 Microsoft Cognitive Toolkit release includes the following key features and enhancements.

- 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 15.
MICROSOFT COGNITIVE TOOLKIT RELEASE 17.07

The NVIDIA container image of Microsoft Cognitive Toolkit, release 17.07, is available.

Microsoft Cognitive Toolkit container image version 17.07 is based on Microsoft Cognitive Toolkit 2.0.

Contents of the Microsoft Cognitive Toolkit

This container image contains the source files that are used to build the Microsoft Cognitive Toolkit in the /opt/cntk directory. The Microsoft Cognitive Toolkit binaries are pre-built and installed into the /usr/local/cntk directory.

A Python 3.4 Conda environment (including the Microsoft Cognitive Toolkit module) is installed in the /opt/conda/envs/cntk-py34 directory. This is included in PATH, so by default the python command executes a Python 3.4 interpreter. To use the base Ubuntu Python 2.7 binary, you can either adjust PATH or explicitly call /usr/bin/python.

This image of the Microsoft Cognitive Toolkit 2.0, is built with the following options enabled:

- 1-bit SGD
- Async SGD
- Python 3 API
- OpenCV
- Kaldi

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
Key Features and Enhancements

This Microsoft Cognitive Toolkit release includes the following key features and enhancements.

- Ubuntu 16.04 with June 2017 updates

Known Issues

There are no known issues in this release.
Chapter 16.
MICROSOFT COGNITIVE TOOLKIT RELEASE
17.06

The NVIDIA container image of Microsoft Cognitive Toolkit, release 17.06, is available. Microsoft Cognitive Toolkit container image version 17.06 is based on Microsoft Cognitive Toolkit 2.0.

Contents of the Microsoft Cognitive Toolkit

This container image contains the source files that are used to build the Microsoft Cognitive Toolkit in the /opt/cntk directory. The Microsoft Cognitive Toolkit binaries are pre-built and installed into the /usr/local/cntk directory.

A Python 3.4 Conda environment (including the Microsoft Cognitive Toolkit module) is installed in the /opt/conda/envs/cntk-py34 directory. This is included in PATH, so by default the python command executes a Python 3.4 interpreter. To use the base Ubuntu Python 2.7 binary, you can either adjust PATH or explicitly call /usr/bin/python.

This image of the Microsoft Cognitive Toolkit 2.0, is built with the following options enabled:

- 1-bit SGD
- Async SGD
- Python 3 API
- OpenCV
- Kaldi

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 Microsoft Cognitive Toolkit release includes the following key features and enhancements.

- Improved distributed Python validation performance by avoiding unnecessary NCCL initializations.
- Ubuntu 16.04 with May 2017 updates

Known Issues

There are no known issues in this release.
Chapter 17. MICROSOFT COGNITIVE TOOLKIT RELEASE 17.05

The NVIDIA container image of Microsoft Cognitive Toolkit, release 17.05, is available.

Microsoft Cognitive Toolkit container image version 17.05 is based on Microsoft Cognitive Toolkit 2.0.rc2.

Contents of the Microsoft Cognitive Toolkit

This container image contains the source files that are used to build the Microsoft Cognitive Toolkit in the /opt/cntk directory. The Microsoft Cognitive Toolkit binaries are pre-built and installed into the /usr/local/cntk directory.

A Python 3.4 Conda environment (including the Microsoft Cognitive Toolkit module) is installed in the /opt/conda/envs/cntk-py34 directory. This is included in PATH, so by default the python command executes a Python 3.4 interpreter. To use the base Ubuntu Python 2.7 binary, you can either adjust PATH or explicitly call /usr/bin/python.

This image of the Microsoft Cognitive Toolkit 2.0.rc2, is built with the following options enabled:

- 1-bit SGD
- Async SGD
- Python 3 API
- OpenCV
- Kaldi

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 Microsoft Cognitive Toolkit release includes the following key features and enhancements.

- Latest cuDNN release
- Ubuntu 16.04 with April 2017 updates

Known Issues

There are no known issues in this release.
Chapter 18.
MICROSOFT COGNITIVE TOOLKIT RELEASE 17.04

The NVIDIA container image of Microsoft Cognitive Toolkit, release 17.04, is available. Microsoft Cognitive Toolkit container image version 17.04 is based on Microsoft Cognitive Toolkit 2.0.beta15.0.

Contents of the Microsoft Cognitive Toolkit

This container image contains the source files that are used to build the Microsoft Cognitive Toolkit in the `/opt/cntk` directory. The Microsoft Cognitive Toolkit binaries are pre-built and installed into the `/usr/local/cntk` directory.

A Python 3.4 Conda environment (including the Microsoft Cognitive Toolkit module) is installed in the `/opt/conda/envs/cntk-py34` directory. This is included in PATH, so by default the `python` command executes a Python 3.4 interpreter. To use the base Ubuntu Python 2.7 binary, you can either adjust PATH or explicitly call `/usr/bin/python`.

This image of the Microsoft Cognitive Toolkit, 2.0beta15.0, is built with the following options enabled:

- 1-bit SGD
- Async SGD
- Python 3 API
- OpenCV
- Kaldi

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 Microsoft Cognitive Toolkit release includes the following key features and enhancements.

- Dilated convolution support (carried over from 17.03)
- Fixed major bug from Microsoft Cognitive Toolkit 2.0 beta 15 affecting multi-GPU training with NCCL
- Fixed minor bug in end-of-sweep detection
- Ubuntu 16.04 with March 2017 updates

Known Issues

There are no known issues in this release.
Chapter 19.
MICROSOFT COGNITIVE TOOLKIT RELEASE 17.03

The NVIDIA container image of Microsoft Cognitive Toolkit, release 17.03, is available.

Microsoft Cognitive Toolkit container image version 17.03 is based on Microsoft Cognitive Toolkit 2.0.beta12.0.

Contents of Microsoft Cognitive Toolkit

This container image contains the source files that are used to build the Microsoft Cognitive Toolkit in the /opt/cntk directory. The Microsoft Cognitive Toolkit binaries are pre-built and installed into the /usr/local/cntk directory.

A Python 3.4 Conda environment (including the Microsoft Cognitive Toolkit module) is installed in the /opt/conda/envs/cntk-py34 directory. This is included in PATH, so by default the python command executes a Python 3.4 interpreter. To use the base Ubuntu Python 2.7 binary, you can either adjust PATH or explicitly call /usr/bin/python.

This image of the Microsoft Cognitive Toolkit, 2.0beta12.0, is built with the following options enabled:

‣ 1-bit SGD
‣ Async SGD
‣ Python 3 API
‣ OpenCV
‣ Kaldi

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 Microsoft Cognitive Toolkit release includes the following key features and enhancements.

- Dilated convolution support (carried over from 17.02)
- GPU memory consumption optimizations
- NCCL integration for MASGD Solver
- Fixes in provided example and tutorial scripts
- Ubuntu 16.04 with February 2017 updates

Known Issues

There is a known issue in 17.03 that causes corruption in networks incorporating residual connections, such as ResNet. As a result, they fail to train to full accuracy.
Chapter 20.
MICROSOFT COGNITIVE TOOLKIT RELEASE
17.02

The NVIDIA container image of Microsoft Cognitive Toolkit, release 17.02, is available. Microsoft Cognitive Toolkit container image version 17.02 is based on Microsoft Cognitive Toolkit 2.0.beta9.0.

Contents of the Microsoft Cognitive Toolkit

This container image contains the source files that are used to build the Microsoft Cognitive Toolkit in the `/opt/cntk` directory. The Microsoft Cognitive Toolkit binaries are pre-built and installed into the `/usr/local/cntk` directory.

A Python 3.4 Conda environment (including the Microsoft Cognitive Toolkit module) is installed in the `/opt/conda/envs/cntk-py34` directory. This is included in PATH, so by default the `python` command executes a Python 3.4 interpreter. To use the base Ubuntu Python 2.7 binary, you can either adjust PATH or explicitly call `/usr/bin/python`.

This image of the Microsoft Cognitive Toolkit, 2.0beta9.0, is built with the following options enabled:

- 1-bit SGD
- Async SGD
- Python API
- OpenCV
- Kaldi

The container also includes the following:

- Ubuntu 14.04
- NVIDIA CUDA® 8.0.61
- NVIDIA CUDA® Deep Neural Network library™ (cuDNN) 6.0.13
- NCCL 1.6.1 (optimized for NVLink™)
Key Features and Enhancements

This Microsoft Cognitive Toolkit release includes the following key features and enhancements.

- Dilated convolution support added to new features
- Ubuntu 14.04 with January 2017 updates

Known Issues

The Microsoft Cognitive Toolkit 17.02 container has a known issue in which the `v2librarytest TrainSequenceToSequenceTranslator` unit test fails due to a bug in the Microsoft Cognitive Toolkit C++ Library API's `TextFormatMinibatchSource` object.

The Microsoft Cognitive Toolkit 2.0 Beta9 introduced `RandomArea` cropping. This is useful for Inception style networks. To avoid confusion, the `Random` crop mode available in previous releases was renamed to `RandomSide`. Users will need to update `BrainScripts` accordingly.
The NVIDIA container image of Microsoft Cognitive Toolkit, release 17.01, is available.

Microsoft Cognitive Toolkit container image version 17.01 is based on Microsoft Cognitive Toolkit 2.0.beta5.0.

Contents of the Microsoft Cognitive Toolkit

This container image contains the source files that are used to build Microsoft Cognitive Toolkit in the /opt/cntk directory. The Microsoft Cognitive Toolkit binaries are pre-built and installed into the /usr/local/cntk directory.

A Python 3.4 Conda environment (including the Microsoft Cognitive Toolkit module) is installed in the /opt/conda/envs/cntk-py34 directory. This is included in PATH, so by default the python command executes a Python 3.4 interpreter. To use the base Ubuntu Python 2.7 binary, you can either adjust PATH or explicitly call /usr/bin/python.

This image of the Microsoft Cognitive Toolkit, 2.0beta5.0, is built with the following options enabled:

- 1-bit SGD
- Async SGD
- Python API
- OpenCV
- Kaldi

The container also includes the following:

- Ubuntu 14.04
- NVIDIA CUDA® 8.0.54
- NVIDIA CUDA® Deep Neural Network library™ (cuDNN) 6.0.10
- NCCL 1.6.1 (optimized for NVLink™)
Key Features and Enhancements

This Microsoft Cognitive Toolkit release includes the following key features and enhancements.

➤ Ubuntu 14.04 with December 2016 updates

Known Issues

There are no known issues with this release.
Chapter 22.  
MICROSOFT COGNITIVE TOOLKIT RELEASE 16.12

The NVIDIA container image of Microsoft Cognitive Toolkit, release 16.12, is available.  
Microsoft Cognitive Toolkit container image version 16.12 is based on Microsoft Cognitive Toolkit 2.0.beta5.0.

Contents of the Microsoft Cognitive Toolkit

This container image contains the source files that are used to build the Microsoft Cognitive Toolkit in the /opt/cntk directory. The Microsoft Cognitive Toolkit binaries are pre-built and installed into the /usr/local/cntk directory.

A Python 3.4 Conda environment (including the Microsoft Cognitive Toolkit module) is installed in the /opt/conda/envs/cntk-py34 directory. This is included in PATH, so by default the python command executes a Python 3.4 interpreter. To use the base Ubuntu Python 2.7 binary, you can either adjust PATH or explicitly call /usr/bin/python.

This image of the Microsoft Cognitive Toolkit, 2.0beta5.0, is built with the following options enabled:

- 1-bit SGD
- Async SGD
- Python API
- OpenCV
- Kaldi

The container also includes the following:

- Ubuntu 14.04
- NVIDIA CUDA® 8.0.54
- NVIDIA CUDA® Deep Neural Network library™ (cuDNN) 6.0.5
- NCCL 1.6.1 (optimized for NVLink™)
Key Features and Enhancements

This Microsoft Cognitive Toolkit release includes the following key features and enhancements.

- Supports FP32 arithmetic and storage
- Optimized multi-GPU training
  - NCCL integration for improved multi-GPU scaling
  - Supports quantized (1-bit) communication
- Supports recurrent neural networks
  - Supports cuDNN recurrent neural networks (RNN) layers

Requires explicit use by model script

- BrainScript (text file), Python, and C++ frontends
- Ubuntu 14.04 with November 2016 updates

Known Issues

There are no known issues with 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 Unites 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