



# TRANSFER LEARNING TOOLKIT INTELLIGENT VIDEO ANALYTICS RELEASE NOTES

RN-09244-002\_v2.0 | August 2020

**Release Notes**



# TABLE OF CONTENTS

Chapter 1. Transfer Learning Toolkit for Intelligent Video Analytics V2.0 Release Notes.....	1
--	---

# Chapter 1.

## TRANSFER LEARNING TOOLKIT FOR INTELLIGENT VIDEO ANALYTICS V2.0

### RELEASE NOTES

#### Description

NVIDIA Transfer Learning Toolkit (TLT) is a Python package to enable NVIDIA customers the ability to fine-tune pretrained models with customer's own data and export them for TensorRT based inference through an edge device.

#### Key Features

Features included in this release:

- ▶ Pretrained models for several public architectures and reference applications serving computer vision related object classification and detection Intelligent Video Analytics (IVA) use cases.
- ▶ Support for YOLOv3, FasterRCNN, SSD, RetinaNet and DSSD object detection models.
- ▶ Support for MaskRCNN Instance segmentation model
- ▶ Support to train highly accurate purpose-built models - PeopleNet, TrafficCamNet, DashCamNet, FaceDetectIR, VehicleTypeNet, VehicleMakeNet
- ▶ Quantization aware training for accurate INT8 models
- ▶ Support for Automatic Mixed Precision (AMP) training
- ▶ Offline augmentation tool
- ▶ Jupyter notebook examples showing how to use the pretrained models effectively.
- ▶ Model adaptation and retraining that is easy to use in heterogeneous multiple GPU environments.
- ▶ Pruning API that compresses the size of the model during training.
- ▶ Model export API for integrating the model directly into the DeepStream environment.

- ▶ Converter utility to generate a device specific optimized TensorRT engine.
- ▶ TLT uses the CUDA multi-process service which helps in optimizing GPU utilization during multiple GPU training.

## Contents

Components included in this release:

- ▶ TLT docker
- ▶ Jupyter notebook with sample workflows
- ▶ Getting Started Guide containing usage and installation instructions
- ▶ tlt-converter (for x86 it's part of the package)
- ▶ tlt-converter for Jetson (ARM64) available [here](#).
- ▶ Pre-trained weights trained on Open Image dataset available on NGC
  - ▶ [Classification](#)
  - ▶ [Object Detection](#)
  - ▶ [Object Detection - DetectNet\\_v2](#)
  - ▶ [Instance Segmentation](#)
- ▶ Unpruned and Pruned models for Purpose-built models - Pruned models can be deployed out-of-box with DeepStream and unpruned models can be used for re-training.
  - ▶ [PeopleNet](#)
  - ▶ [TrafficCamNet](#)
  - ▶ [DashCamNet](#)
  - ▶ [FaceDetectIR](#)
  - ▶ [VehicleTypeNet](#)
  - ▶ [VehicleMakeNet](#)

## Software Requirements

- ▶ Ubuntu 18.04 LTS
- ▶ NVIDIA GPU Cloud account and API key - <https://ngc.nvidia.com/>
- ▶ nvidia-docker 2 installed, instructions: [https://github.com/nvidia/nvidia-docker/wiki/Installation-\(version-2.0\)](https://github.com/nvidia/nvidia-docker/wiki/Installation-(version-2.0))
- ▶ NVIDIA GPU driver v410.xx or above



DeepStream 5.0 for inference and deployment <https://developer.nvidia.com/deepstream-sdk> is recommended.

## Hardware Requirements

### Minimum

- ▶ 4 GB system RAM
- ▶ 4 GB of GPU RAM
- ▶ Single core CPU
- ▶ 1 NVIDIA GPU
- ▶ 50 GB of HDD space

### Recommended

- ▶ 32 GB system RAM
- ▶ 32 GB of GPU RAM
- ▶ 8 core CPU
- ▶ 4 GPUs - V100
- ▶ 100 GB of SSD space

### Known Issues

1. SSD, DSSD, YOLOv3, RetinaNet, FasterRCNN, and MaskRCNN integration to DeepStream is a feature and requires custom plugins from the TensorRT Open Source Software (OSS) library. DeepStream 5.0 doesn't natively support custom plugins from TensorRT OSS. Instructions to build TensorRT OSS and custom parsing code to run with Deepstream have been provided here: [https://github.com/NVIDIA-AI-IOT/deepstream\\_tlt\\_apps](https://github.com/NVIDIA-AI-IOT/deepstream_tlt_apps).
2. Transfer Learning is not supported on pruned models across all applications.
3. When training with multiple GPUs, you might need to scale down the batch\_size and/or scale up the learning rate to get the same accuracy seen in single GPU training.
4. When resuming a DetectNet\_v2 training from checkpoint, please make sure to maintain the same number of GPUs and the same command line to restart the training.
5. When training DetectNet\_v2 for object detection use-cases with more than 10 classes, you may need to either update the cost\_weight parameter in the cost\_function\_config, or balance the number of samples per class in the dataset for better training.
6. When training a DetectNet\_v2 network for datasets with less than 20,000 images, please use smaller batch-sizes (1, 2 or 4) to get better accuracy.
7. When using MaskRCNN, please make sure GPU 0 is free.
8. ResNet101 pre-trained weights from NGC is not supported on YOLOV3, SSD, DSSD, and RetinaNet.

### Resolved Issues

1. Loading pretrained weights when retraining a pruned model.

2. Incorrect image\_extension in detectnet dataset\_config returns unrelated error during tlt-evaluate.

## 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

© 2020 NVIDIA Corporation. All rights reserved.

[www.nvidia.com](http://www.nvidia.com)

