NVIDIA TRANSFER LEARNING TOOLKIT

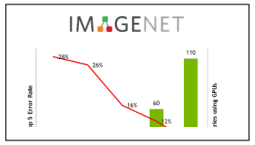
NVIDIA 开发者社区 何琨

POWERING THE DEEP LEARNING ECOSYSTEM

COMPUTER VISION

OBJECT DETECTION

IMAGE CLASSIFICATION



SPEECH & AUDIO

VOICE RECOGNITION

LANGUAGE TRANSLATION

NATURAL LANGUAGE PROCESSING

RECOMMENDATION ENGINES

SENTIMENT ANALYSIS

DEEP LEARNING FRAMEWORKS

Caffe

Chainer

DL4J
Deeplearning4j

Mocha.jl
julia

KERAS

Microsoft CNTK

mxnet

Purine

TensorFlow

torch

MatConvNet

MINERVA

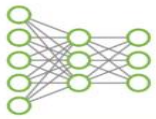
OpenDeep

Pylearn2

theano

NVIDIA DEEP LEARNING SDK

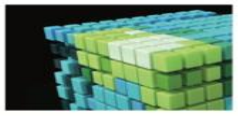
cuDNN



TensorRT

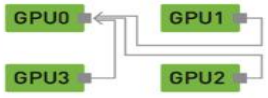
DeepStream SDK

cuBLAS



cuSPARSE

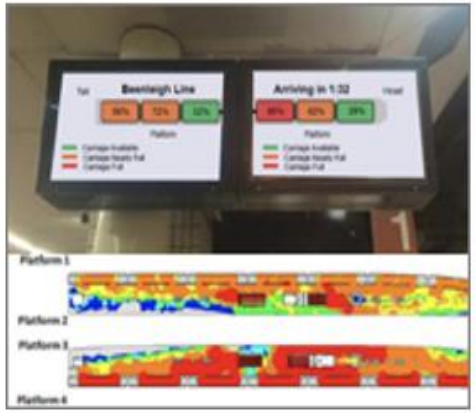
NCCL



developer.nvidia.com/deep-learning-software

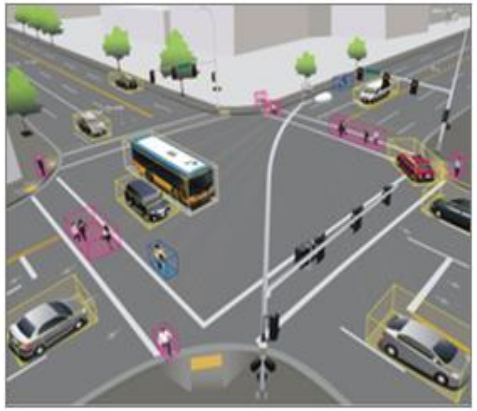
CV IN DEEPLEARNING

Access Control

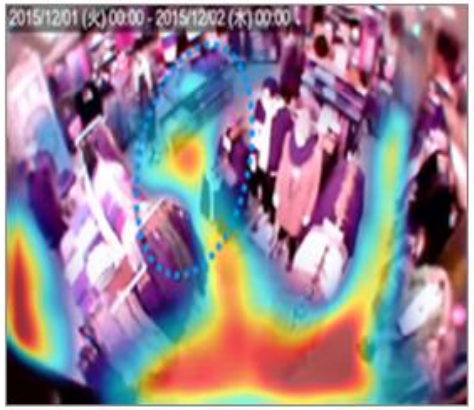


Public Transit

Parking Management



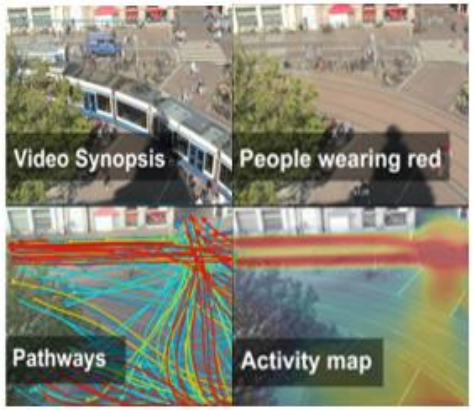
Traffic Engineering



Retail Analytics

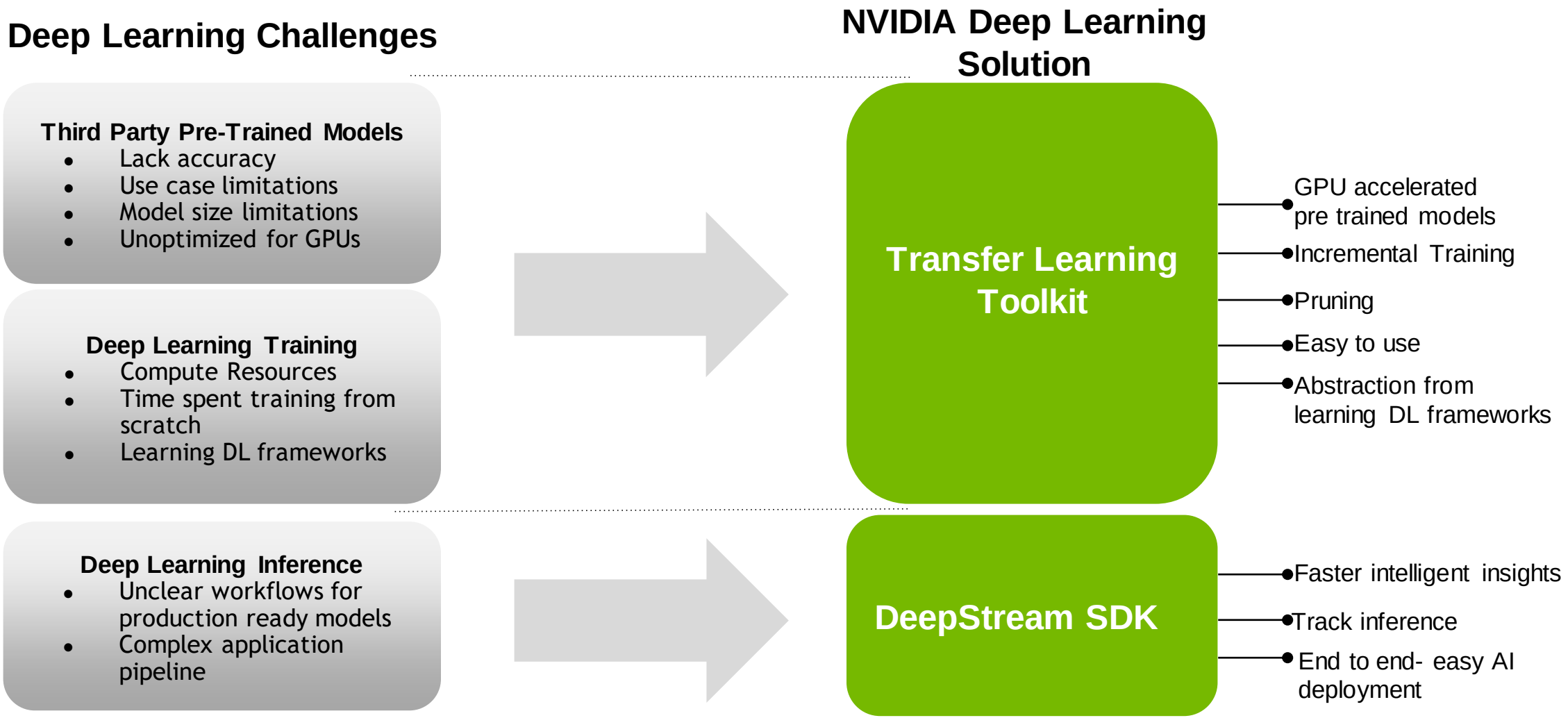
Securing Critical Infrastructure

Managing Logistics



Forensic Analysis

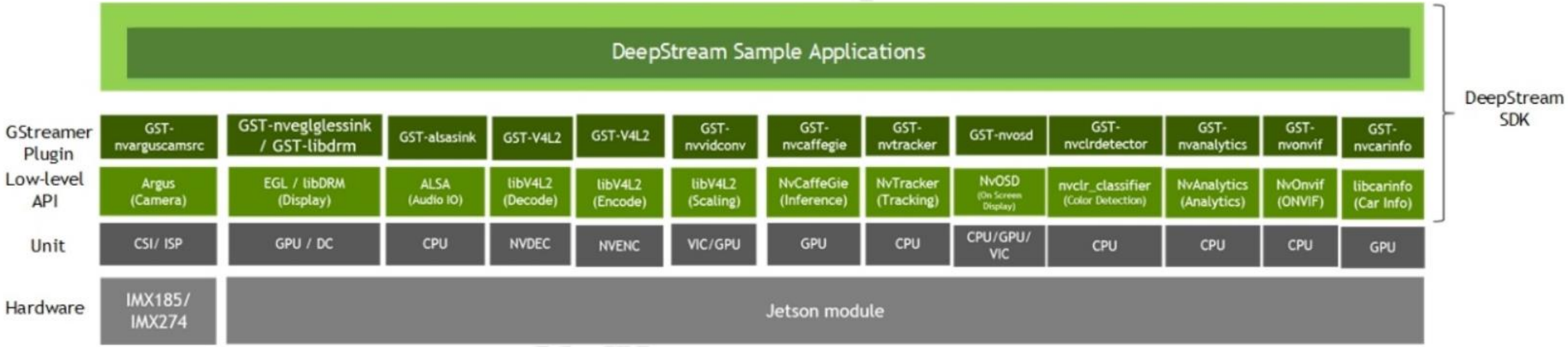
Deep Learning Workflow Management



DeepStream

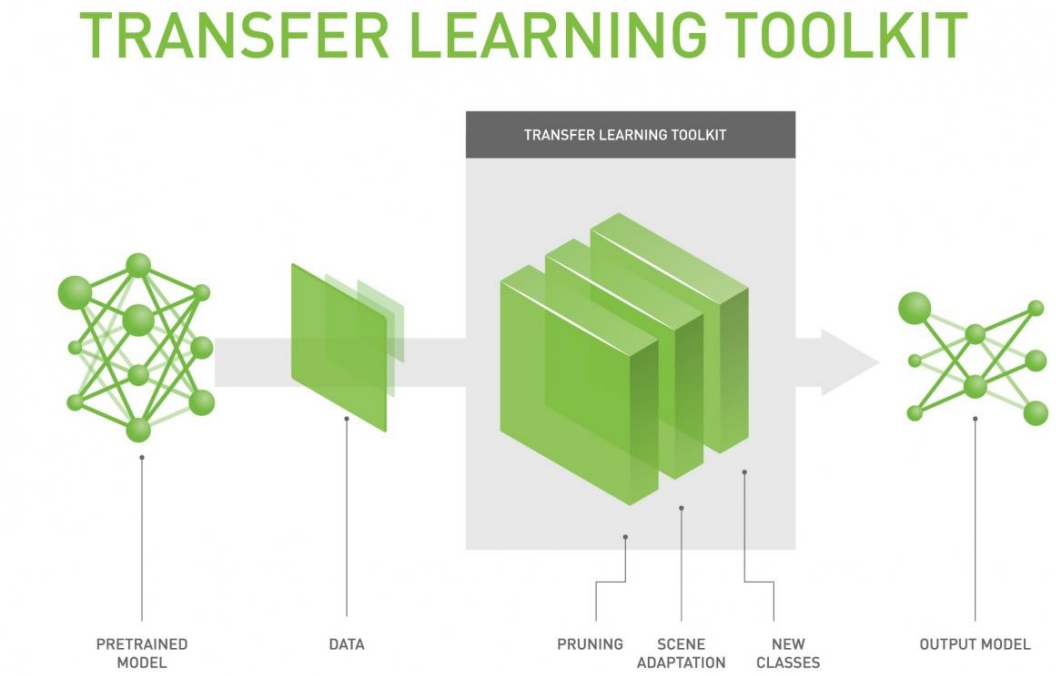
NVIDIA DeepStream SDK是为大规模创建和部署基于AI的视频分析应用程序解决方案而设计的，它提供完整的框架和所有基本构建模块。

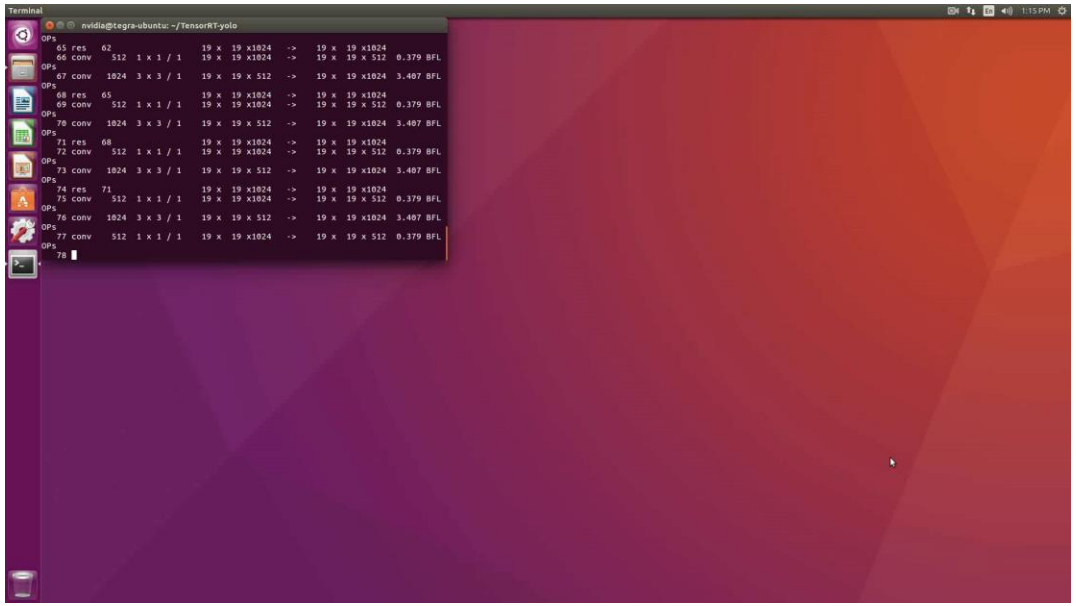
- Gstreamer
- TensorRT

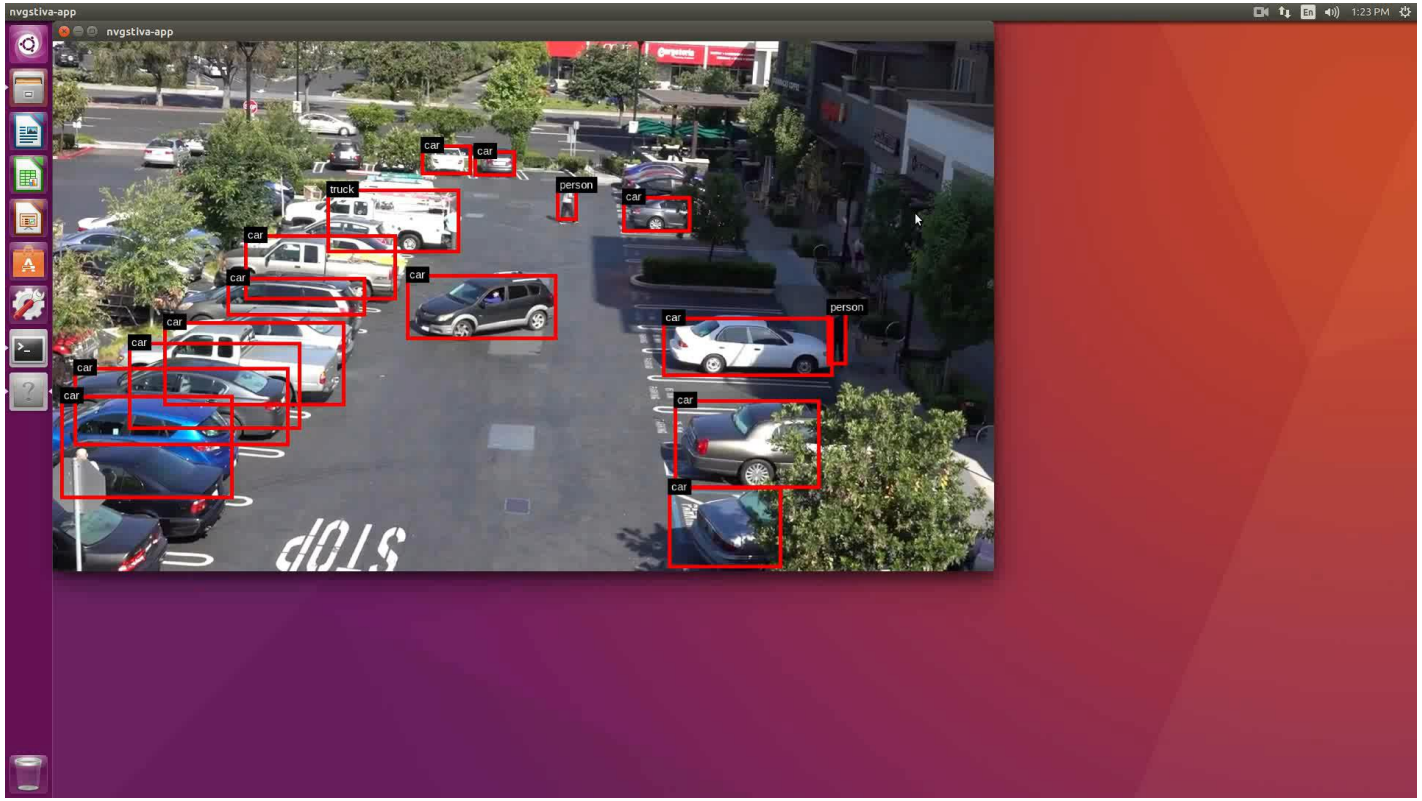


NVIDIA DEEPSTREAM

- A unified release across all our platforms will allow application portability. This includes support for Jetson Nano, our newest AI platform.
- 50% reduction in memory footprint resulting in exceptional stream processing density.
- Turnkey integration with Azure Edge IoT to build applications and services, leveraging the power of Azure cloud.
- Containerized deployment for Jetson platforms. This enhances the ability to deploy applications to Dockers, dramatically enhancing the delivery and maintenance of applications at large scale.
- Plugin sources for inference, message schema converter, and message broker plugins.
- A new reference tracker for robust object tracking.
- Added support for heterogeneous cameras, segmentation networks, monochrome images and hardware accelerated JPEG decode and encode.



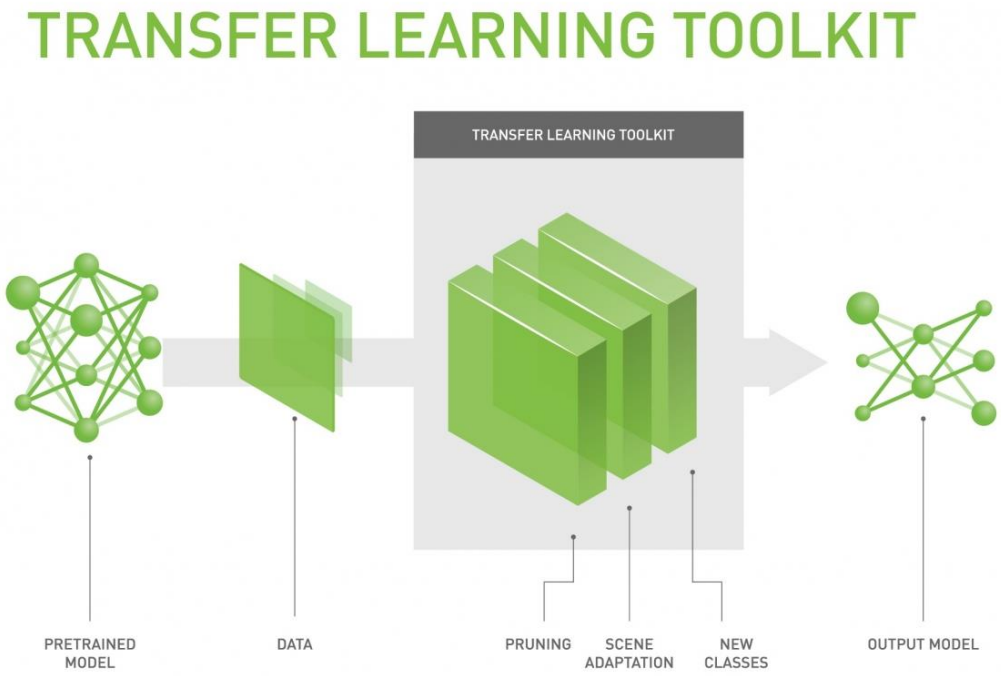

Terminal
1:15 PM
nvidia@tegra-ubuntu: ~/TensorRT-yolo
OPs
65 res 62 19 x 19 x1024 -> 19 x 19 x1024
66 conv 512 1 x 1 / 1 19 x 19 x1024 -> 19 x 19 x 512 0.379 BFL
OPs
67 conv 1024 3 x 3 / 1 19 x 19 x 512 -> 19 x 19 x1024 3.407 BFL
OPs
68 res 65 19 x 19 x1024 -> 19 x 19 x1024
69 conv 512 1 x 1 / 1 19 x 19 x1024 -> 19 x 19 x 512 0.379 BFL
OPs
70 conv 1024 3 x 3 / 1 19 x 19 x 512 -> 19 x 19 x1024 3.407 BFL
OPs
71 res 68 19 x 19 x1024 -> 19 x 19 x1024
72 conv 512 1 x 1 / 1 19 x 19 x1024 -> 19 x 19 x 512 0.379 BFL
OPs
73 conv 1024 3 x 3 / 1 19 x 19 x 512 -> 19 x 19 x1024 3.407 BFL
OPs
74 res 71 19 x 19 x1024 -> 19 x 19 x1024
75 conv 512 1 x 1 / 1 19 x 19 x1024 -> 19 x 19 x 512 0.379 BFL
OPs
76 conv 1024 3 x 3 / 1 19 x 19 x 512 -> 19 x 19 x1024 3.407 BFL
OPs
77 conv 512 1 x 1 / 1 19 x 19 x1024 -> 19 x 19 x 512 0.379 BFL
OPs
78


nvgstiva-app
1:23 PM
nvgstiva-app
car
car
truck
person
car
car
car
car
person
car
car
car
car
car
car
STOP

NVIDIA TRANSFER LEARNING TOOLKIT

Transfer Learning Toolkit是一个基于python的工具包，它使开发人员能够利用NVIDIA预先训练的模型，并为开发人员提供一系列的工具，使流行的网络架构适应他们自己的数据，并且能够训练、调整、修剪和导出模型以进行部署。它还拥有简单的接口和抽象API，提高了深度学习训练工作流的效率。

- GPU优化的预训练砝码，可用于计算机视觉任务
- 轻松修改配置文件以添加新类并使用自定义数据重新训练模型
- 在异构的多GPU环境中执行模型调整和重新训练
- 使用修剪功能缩小模型尺寸
- 模型导出API，可在具有NVIDIA Tesla和Jetson产品的NVIDIA DeepStream SDK上部署



NVIDIA TRANSFER LEARNING TOOLKIT

Efficient Pre-trained Models

GPU-accelerated high performance models trained on large scale datasets.

Faster Inference with Model Pruning

Model pruning reduces size of the model resulting in faster inference

Training with Multiple GPUs

Re-training models, adding custom data for multi GPU training using an easy to use tool

Abstraction

Abstraction from having deep knowledge of frameworks, simple intuitive interface to the features

Containerization

Packaged in a container easily accessible from NVIDIA GPU Cloud website. All code dependencies are managed automatically

Integration

Models exported using TLT are easily consumable for inference with **Deep Stream SDK**

NVIDIA TRANSFER LEARNING TOOLKIT

在指定的公共数据集上训练的图像分类和目标检测模型，可与Transfer Learning Toolkit一起使用。

Image Classification

- ResNet10/18/50
- VGG16/19
- MobileNet V1/V2
- AlexNet
- SqueezeNet
- GoogLeNet

Faster RCNN supporting backbones:

- ResNet10/18/50
- VGG16/19
- GoogLeNet
- MobileNet V1/V2

Object Detection

DetectNet_v2 supporting backbones:

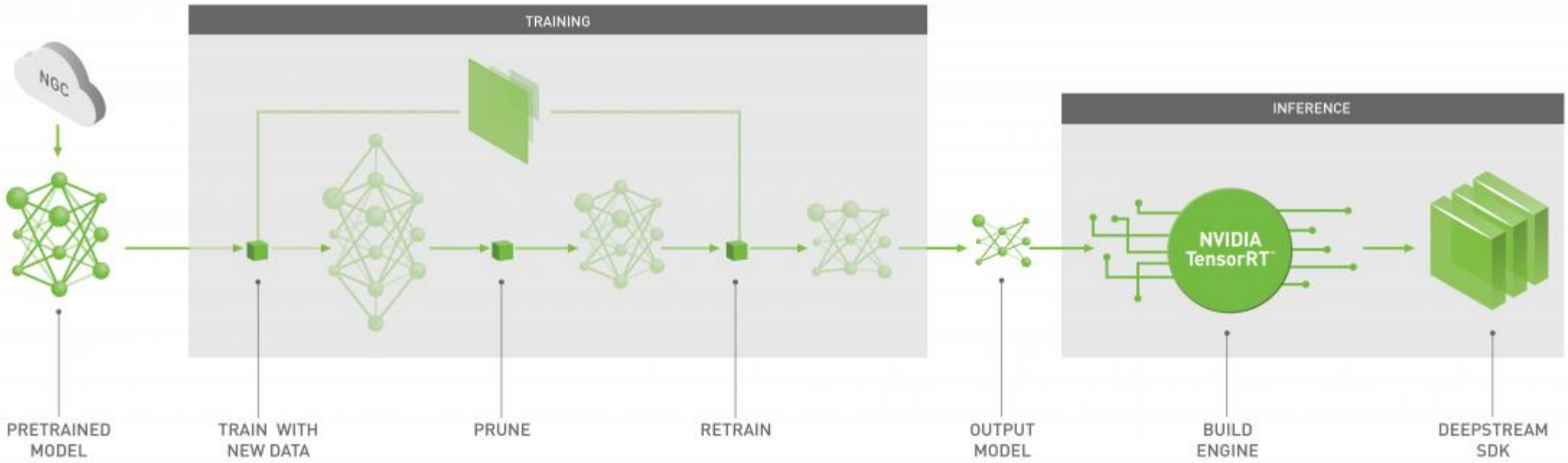
- ResNet10/18/50
- VGG 16/19
- GoogLeNet
- MobileNet V1/V2

SSD:

- ResNet10/18

NVIDIA TRANSFER LEARNING TOOLKIT

为应用在计算机视觉领域的深度学习工作流程，提供了全方位的便利工具



在服务器上部署TRANSFER LEARNING TOOLKIT

Hardware Requirements

Minimum

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

Recommended

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

Software Requirements

- Ubuntu 18.04 LTS/16.04 LTS
- NVIDIA GPU Cloud account and API key - https://ngc.nvidia.com/
- docker-ce installed, https://docs.docker.com/install/linux/docker-ce/ubuntu/
- nvidia-docker2 installed, instructions: https://github.com/nvidia/nvidia-docker/wiki/Installation-(version-2.0)
- NVIDIA GPU driver v410.xx or above

Note: DeepStream 4.0 - NVIDIA SDK inference https://developer.nvidia.com/deepstream-sdk is recommended.

在服务器上部署TRANSFER LEARNING TOOLKIT

Installation Prerequisites

•Install Docker. See: https://www.docker.com/.
•NVIDIA GPU driver v410.xx or above. Download from https://www.nvidia.com/Download/index.aspx?lang=en-us.
•Install NVIDIA Docker 2 from: https://github.com/NVIDIA/nvidia-docker.

Get an NGC API key

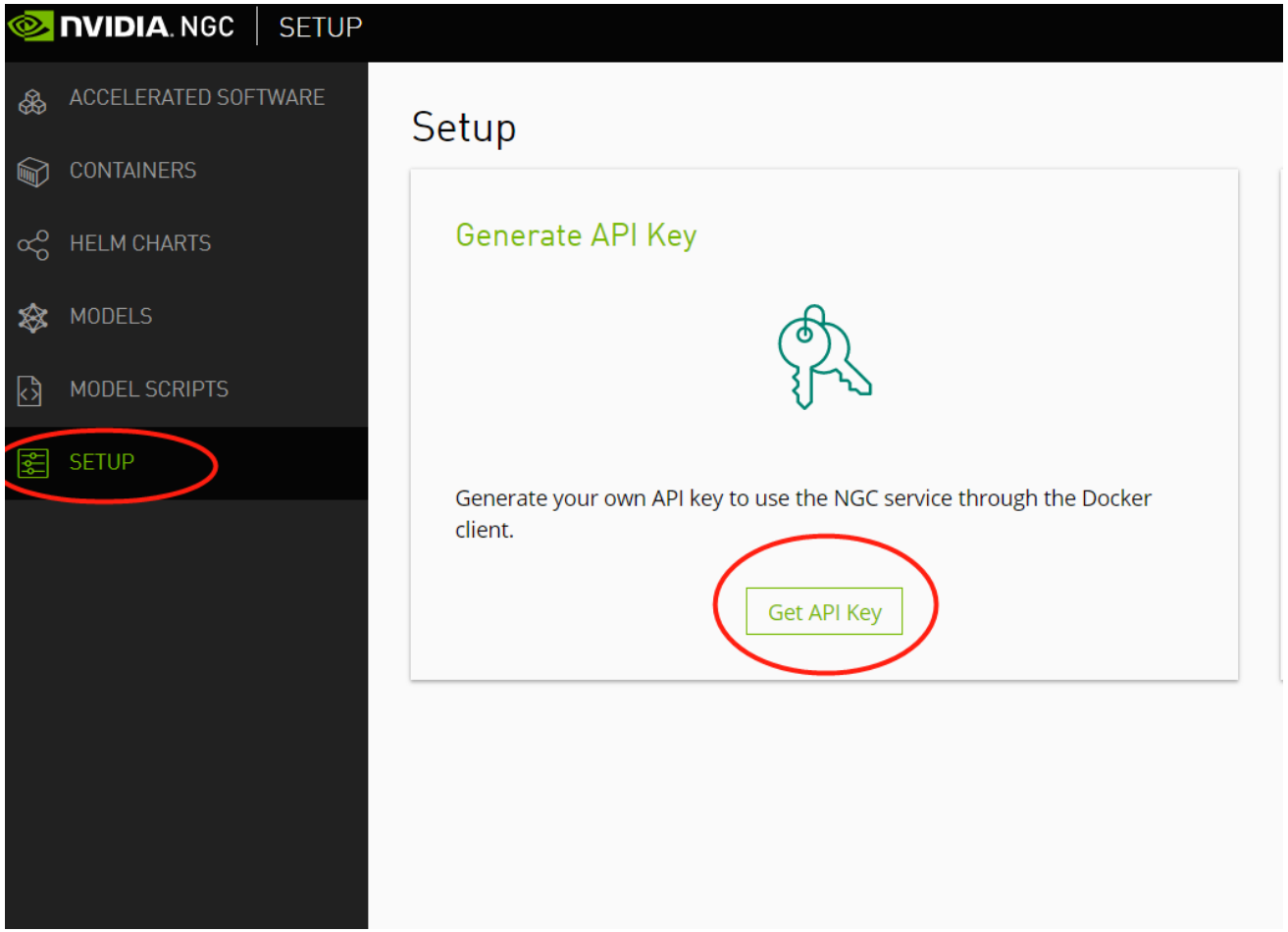
•NVIDIA GPU Cloud account and API key - https://ngc.nvidia.com/

1. Go to NGC and click the **Transfer Learning Toolkit** container in the **Catalog** tab. This message is displayed, **Sign in to access the PULL feature of this repository**.
2. Enter your email address and click **Next** or click **Create an Account**.
3. Choose your **organization** when prompted for Organization/Team.
4. Click **Sign In**.
5. Select the **Containers** tab on the left navigation pane and click the **Transfer Learning Toolkit** tile.

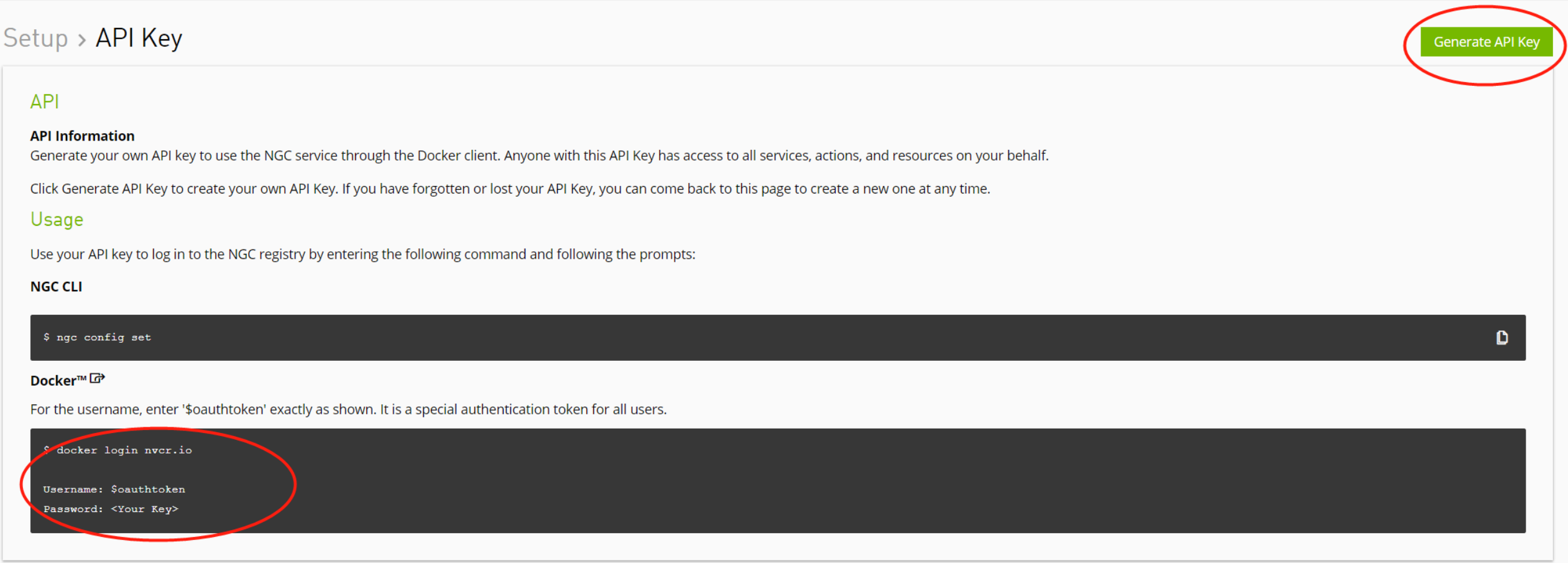
Download the docker container

•Execute docker login nvcr.io from the command line and enter your username and password.
  •Username: $oauthtoken
  •Password: API_KEY
•Execute docker pull nvcr.io/nvidia/tlt-streamanalytics:<version>

在服务器上部署TRANSFER LEARNING TOOLKIT



在服务器上部署TRANSFER LEARNING TOOLKIT



在服务器上部署TRANSFER LEARNING TOOLKIT

在服务器上运行TLT的镜像

1.Run the image using this command.

```
docker run --runtime=nvidia -it nvcr.io/nvidia/tlt-
streamanalytics:v1.0.1_py2 /bin/bash
```

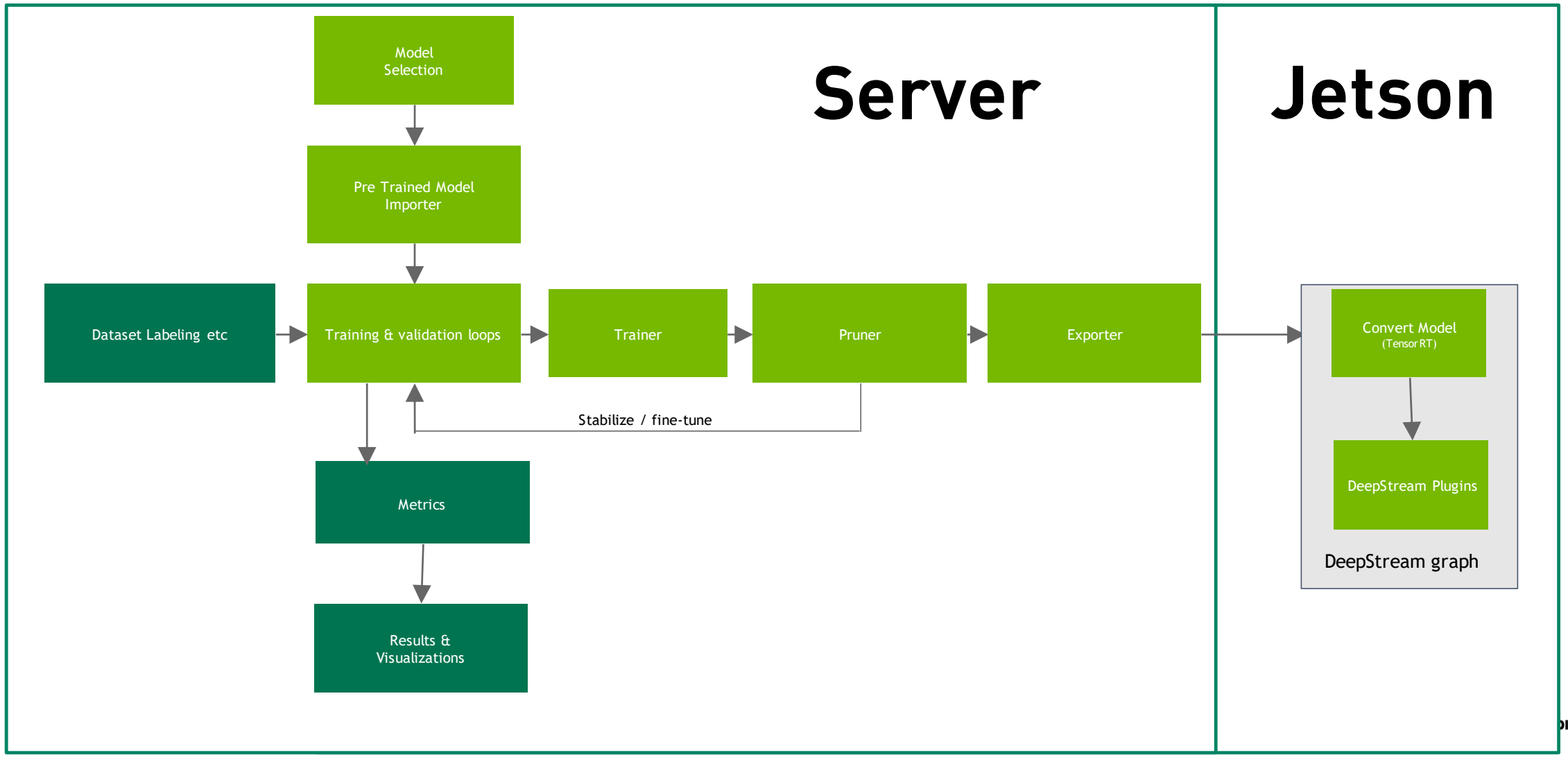
2.Mount local directories using -v and expose the docker ports to the host using -p

```
docker run --runtime=nvidia -it \
-v "/path/to/dir/on/host":"/path/to/dir/in/docker" \ -p 8888:8888 \
nvcr.io/nvidia/tlt-streamanalytics:v1.0_py2 /bin/bash
```

3.Invoke the jupyter notebook using the following command

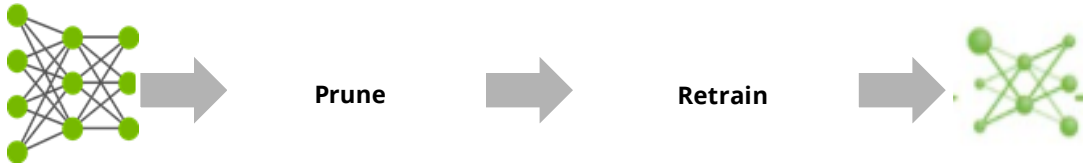
```
jupyter notebook --ip 0.0.0.0 --port 8888 --allow-root
```

TRANSFER LEARNING TOOLKIT的工作流程



TRANSFER LEARNING TOOLKIT的工作流程

Pruning



Network compression

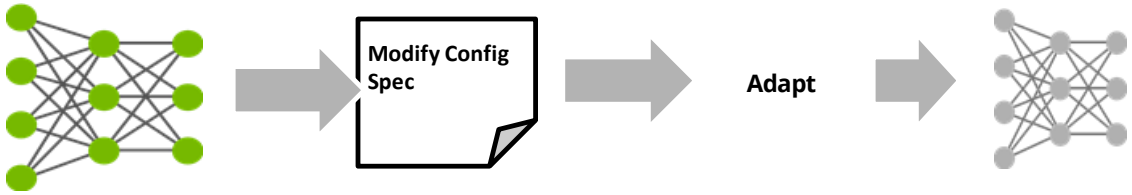
Scene Adaptation

Same network adapting to different angles and vantage points

Same network adapting to new data

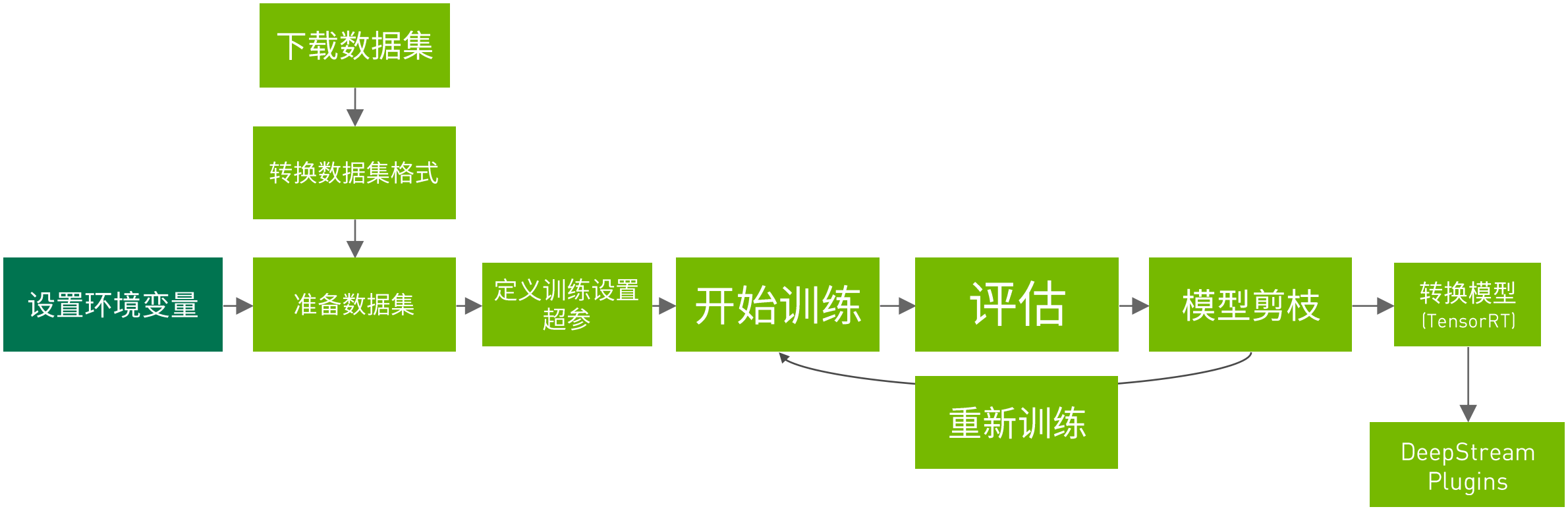
Train with new data from another vantage point, camera location, or added attribute

New Classes



Edited network detecting new class defined in config spec

TRANSFER LEARNING TOOLKIT的工作流程



总结

- NVIDIA Transfer Learning Toolkit为深度学习训练部署流程提供了完整的工具链
- Transfer Learning Toolkit 的安装部署需要使用NGC
- 把训练和剪裁好的模型部署在边缘设备(Jetson 平台)上时，需要在边缘设备上转换成TRT的格式

链接：https://pan.baidu.com/s/1TWT5PMYn-VkYoxbVCWKoWA
提取码：cdaq

更多资源：

https://developer.nvidia-china.com


THANK YOU
NVIDIA®