



# NVIDIA TRANSFER LEARNING TOOLKIT & Jetbot

NVIDIA 开发者社区 何琨

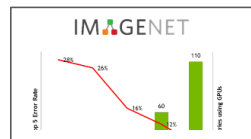
# POWERING THE DEEP LEARNING ECOSYSTEM

## COMPUTER VISION

OBJECT DETECTION



IMAGE CLASSIFICATION



## SPEECH & AUDIO

VOICE RECOGNITION

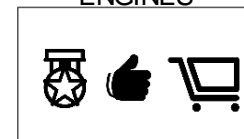


LANGUAGE TRANSLATION



## NATURAL LANGUAGE PROCESSING

RECOMMENDATION ENGINES



SENTIMENT ANALYSIS



## DEEP LEARNING FRAMEWORKS

Caffe



DL4J  
Deeplearning4j

Mocha.jl



K  
KERAS

MatConvNet

Microsoft  
CNTK

MINERVA

mxnet

OpenDeep



Pylearn2



theano



## NVIDIA DEEP LEARNING SDK

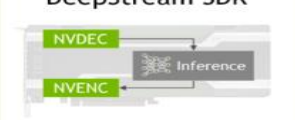
cuDNN



TensorRT



DeepStream SDK



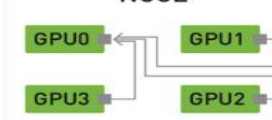
cuBLAS



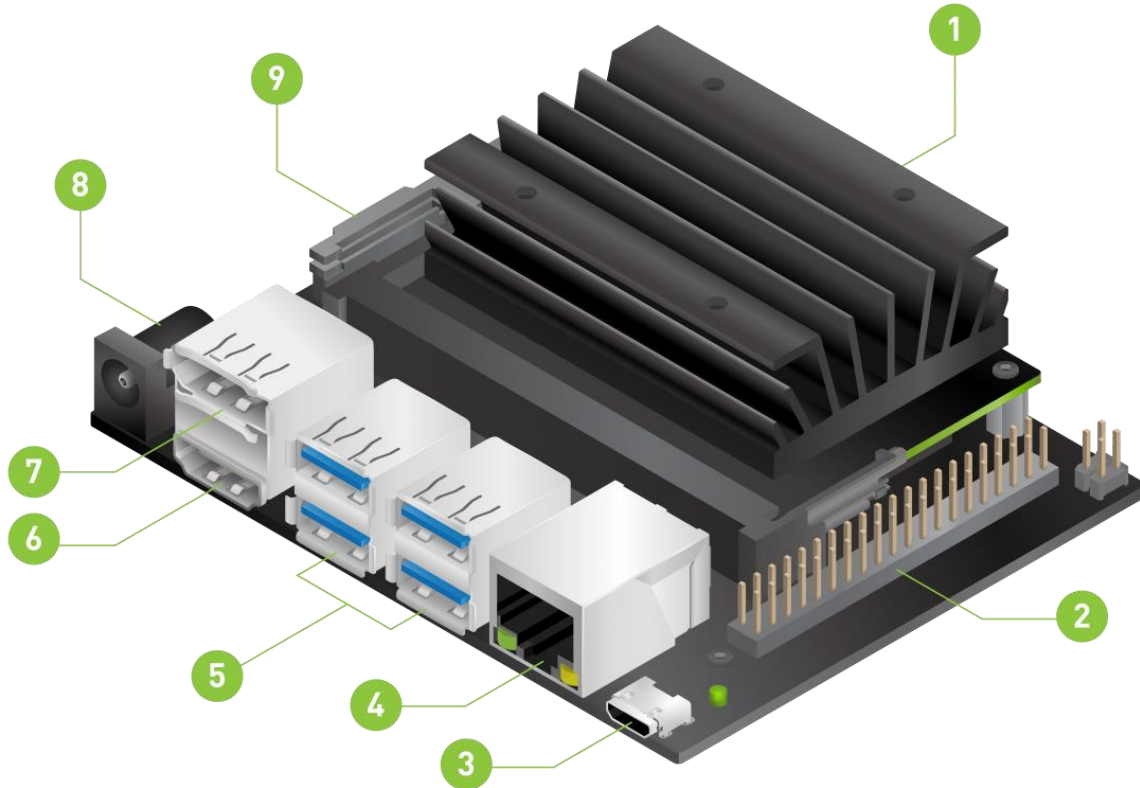
cuSPARSE



NCCL



# 英伟达 NANO开发板

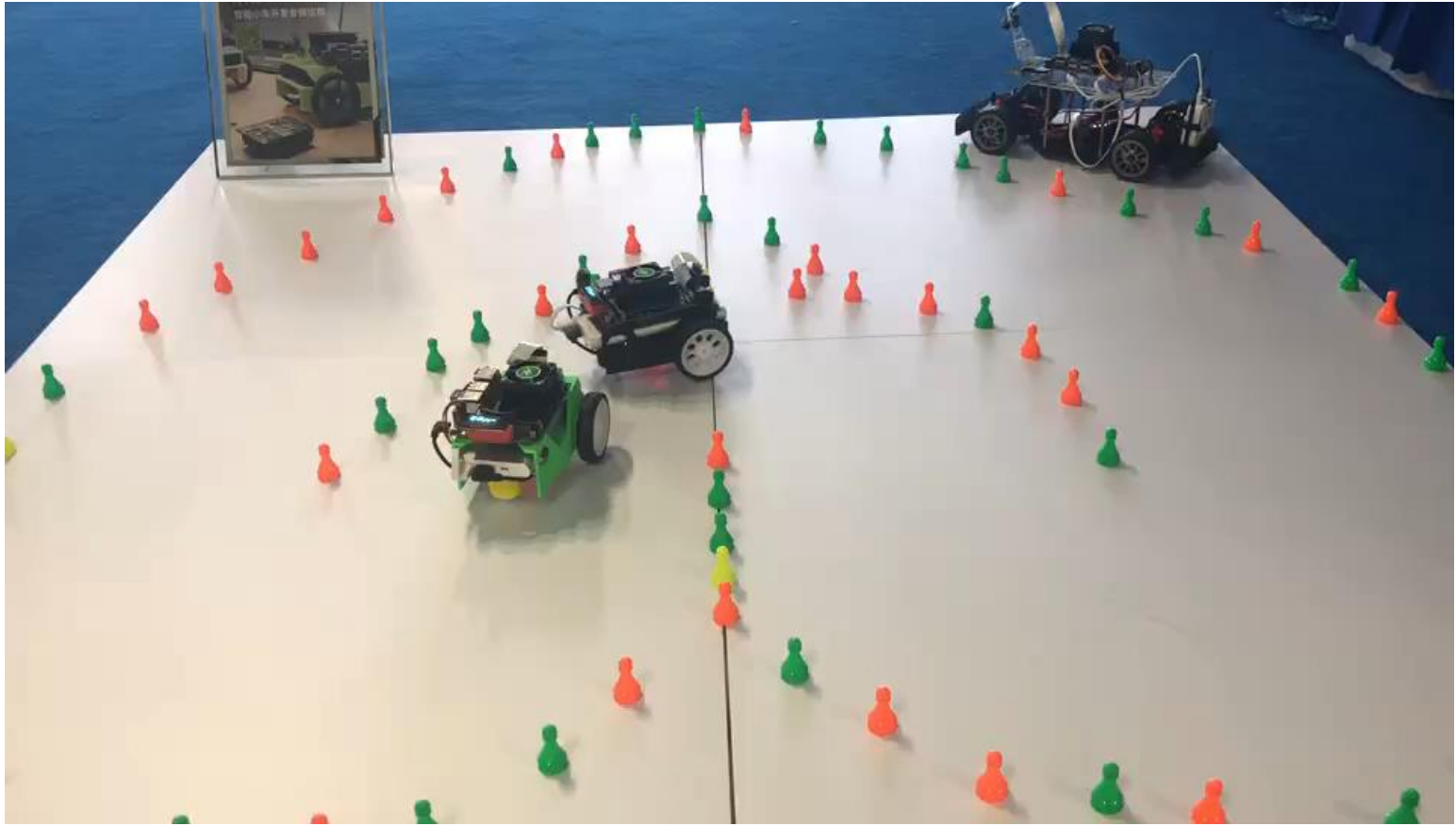


1. microSD card slot for main storage
2. 40-pin expansion header
3. Micro-USB port for 5V power input or for data
4. Gigabit Ethernet port
5. USB 3.0 ports (x4)
6. HDMI output port
7. DisplayPort connector
8. DC Barrel jack for 5V power input
9. MIPI CSI camera connector

1. microSD card (16GB UHS-1 minimum)
2. USB keyboard and mouse
3. Computer display (either HDMI or DP)
4. Micro-USB power supply (5V=2A)







# Deep Learning Workflow Management

## Deep Learning Challenges

### Third Party Pre-Trained Models

- Lack accuracy
- Use case limitations
- Model size limitations
- Unoptimized for GPUs

### Deep Learning Training

- Compute Resources
- Time spent training from scratch
- Learning DL frameworks

### Deep Learning Inference

- Unclear workflows for production ready models
- Complex application pipeline

## NVIDIA Deep Learning Solution

### Transfer Learning Toolkit

- GPU accelerated pre trained models
- Incremental Training
- Pruning
- Easy to use
- Abstraction from learning DL frameworks

### DeepStream SDK

- Faster intelligent insights
- Track inference
- End to end- easy AI deployment

# 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:

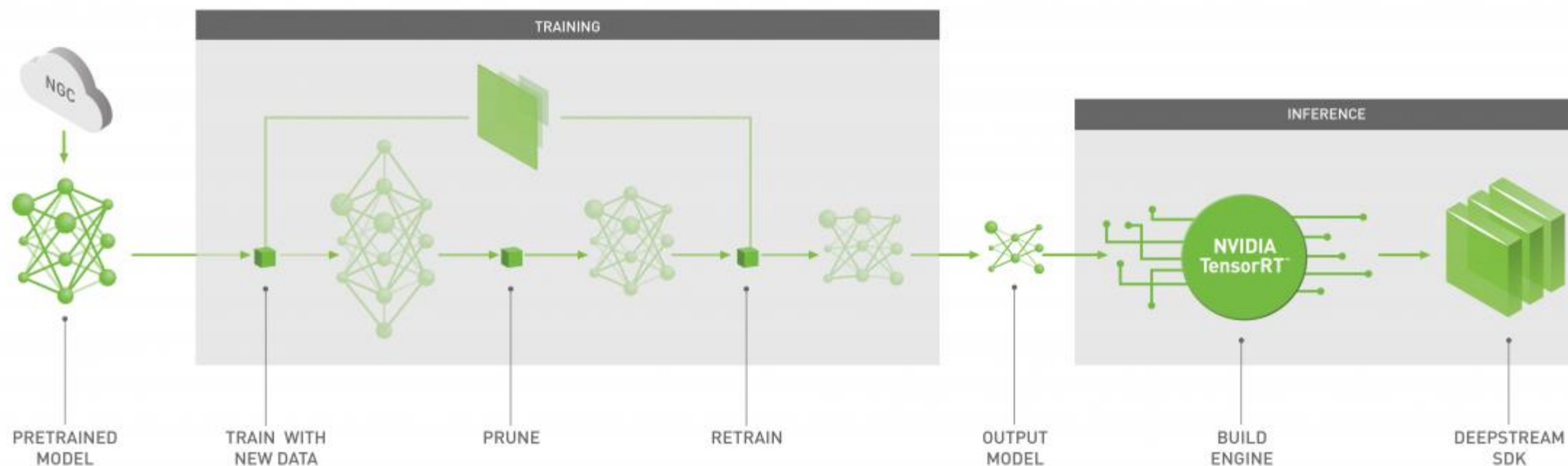
- ResNet10/18/50
- VGG 16/19
- GoogLeNet
- MobileNet V1/V2

### SSD:

- ResNet10/18

# NVIDIA TRANSFER LEARNING TOOLKIT

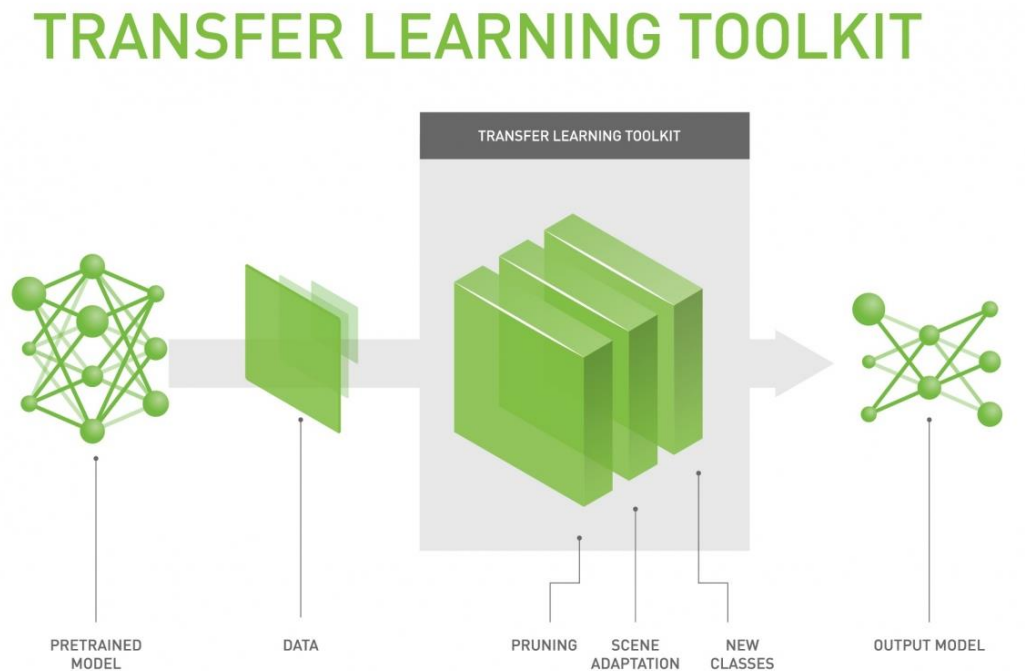
为应用在计算机视觉领域的深度学习工作流程，提供了全方位的便利工具



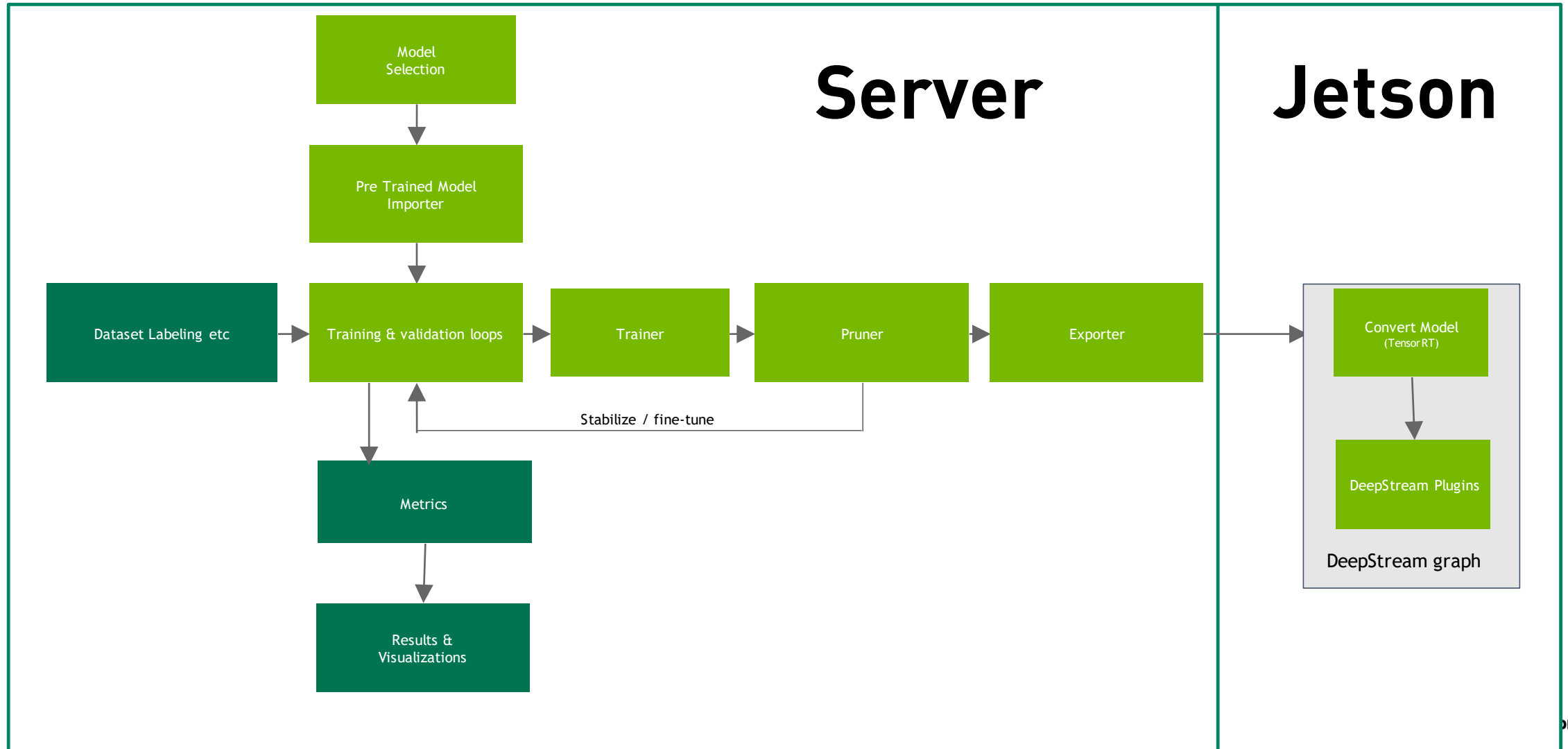
# NVIDIA TRANSFER LEARNING TOOLKIT

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优化的预训练砒码，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



# TRANSFER LEARNING TOOLKIT的工作流程



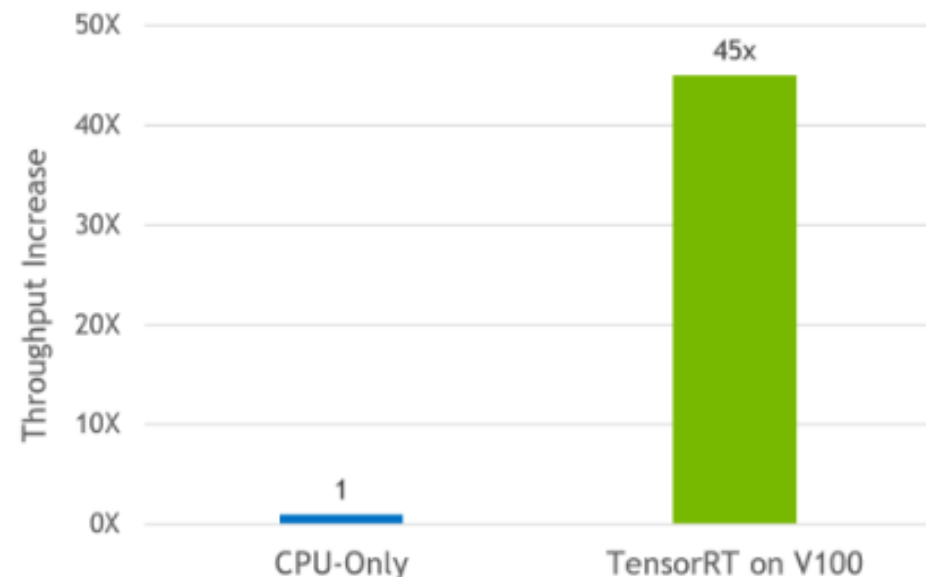
# TensorRT

NVIDIA TensorRT™是一种高性能深度学习推理优化器和运行时加速库，可为深度学习推理应用程序提供低延迟和高吞吐量。使用TensorRT，您可以优化神经网络模型，以高精度校准低精度，最后将模型部署到超大规模数据中心，嵌入式或汽车产品平台。

- 新编译器加速递归神经网络常用的语音和异常检测
- 支持超过20个新的ONNX操作，加速关键语音模型，如BERT，TacoTron 2和WaveRNN
- 扩展了对动态形状的支持，以支持关键的人工智能会话模型

ONNX: Added ConstantOfShape, DequantizeLinear, Equal, Erf, Expand, Greater, GRU, Less, Loop, LRN, LSTM, Not, PRelu, QuantizeLinear, RandomUniform, RandomUniformLike, Range, RNN, Scan, Sqrt, Tile, and Where

## 45x Higher Recommender System Throughput With TensorRT Than CPU



SAP recommender system performance using TensorRT on NVIDIA Tesla V100 GPUs compared with TensorFlow running on Intel Xeon E5-2698 v4 CPU at 2.20GHz.

(Click to Zoom)

# TENSORRT WORKFLOW

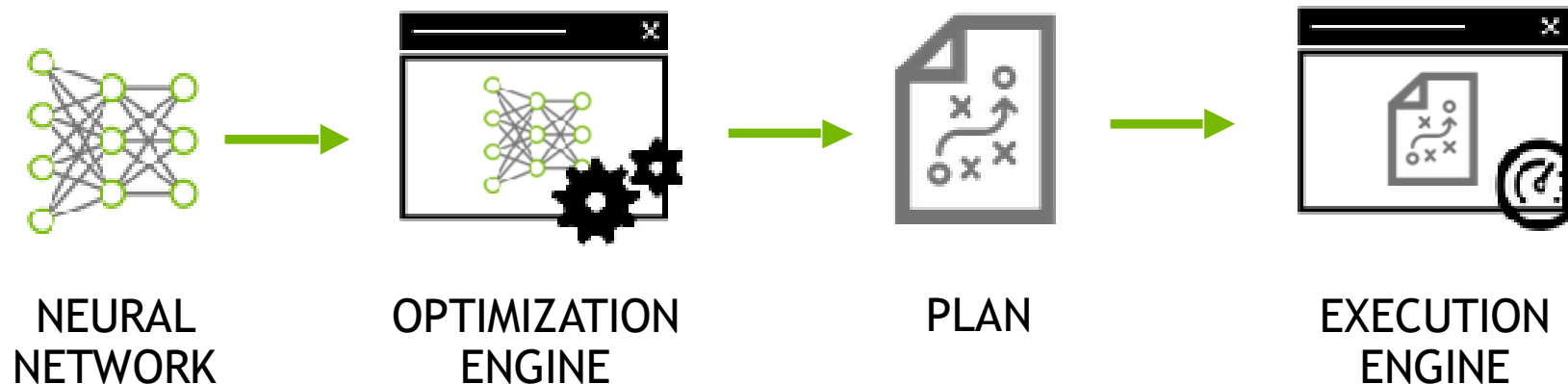
## Input

- Pre-trained FP32 model and network

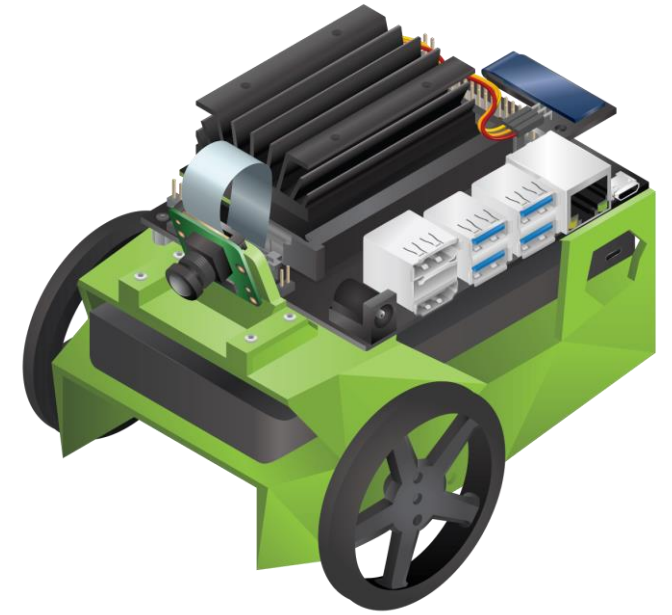
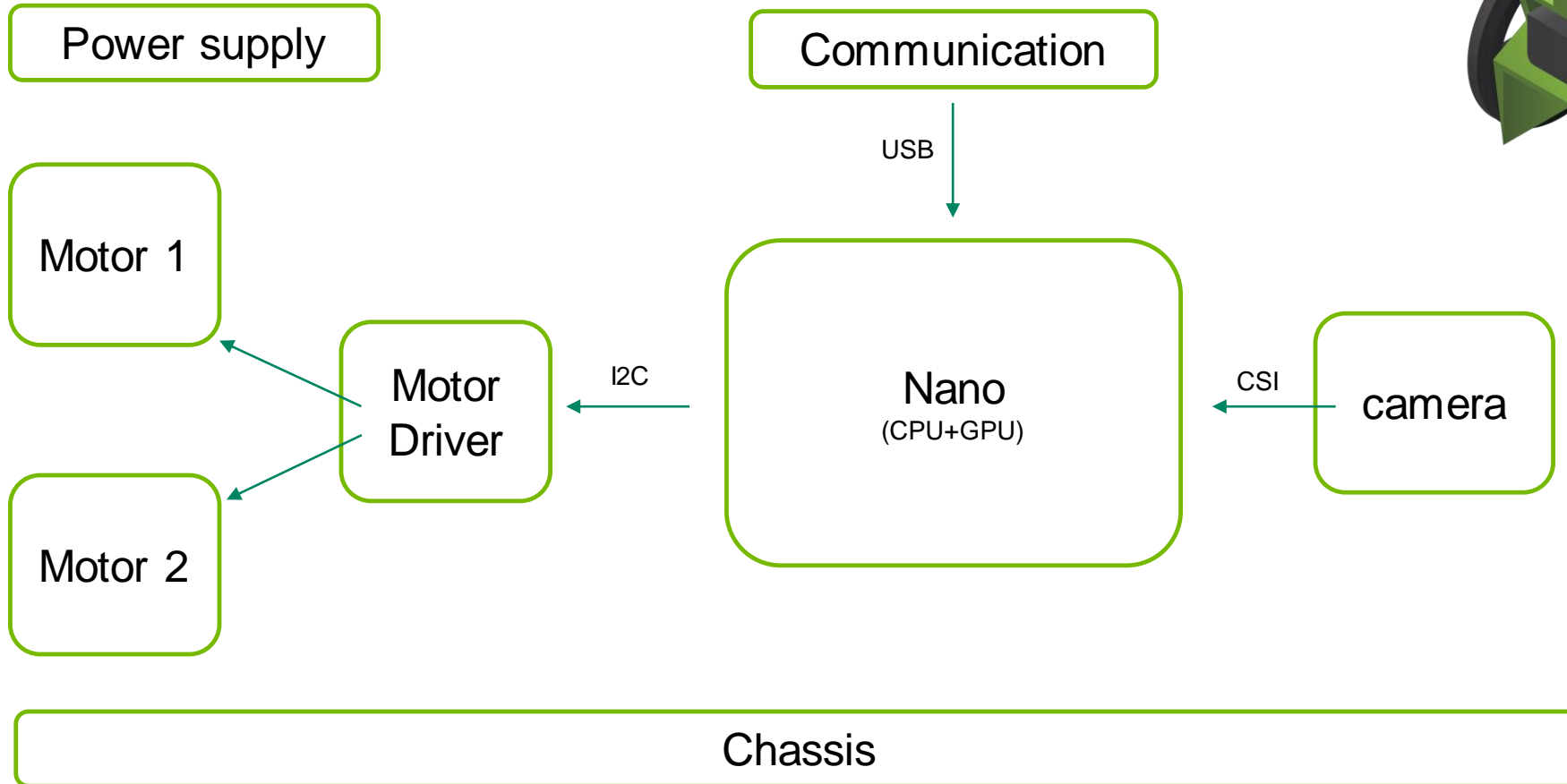
## Output

- Optimized execution engine on GPU for deployment

Serialized a PLAN can be reloaded from the disk into the TensorRT runtime. There is no need to perform the optimization step again.



# 智能小车框架图



# JetBot 系统体系结构

- 控制接口
  - `class Robot(SingletonConfigurable):`
  - `def set_motors(self, left_speed, right_speed):`
  - `def forward(self, speed=1.0, duration=None):`
  - `def backward(self, speed=1.0):`
  - `def left(self, speed=1.0):`
  - `def right(self, speed=1.0):`
  - `def stop(self):`

# JetBot 系统体系结构

- 视频接口

- `class Camera(SingletonConfigurable):`

```
class Camera(SingletonConfigurable):

    value = traitlets.Any()

    # config
    width = traitlets.Integer(default_value=224).tag(config=True)
    height = traitlets.Integer(default_value=224).tag(config=True)
    fps = traitlets.Integer(default_value=21).tag(config=True)
    capture_width = traitlets.Integer(default_value=3280).tag(config=True)
    capture_height = traitlets.Integer(default_value=2464).tag(config=True)

    def __init__(self, *args, **kwargs):
        self.value = np.empty((self.height, self.width, 3), dtype=np.uint8)
        super(Camera, self).__init__(*args, **kwargs)

    try:
        self.cap = cv2.VideoCapture(self._gst_str(), cv2.CAP_GSTREAMER)

        re, image = self.cap.read()

        if not re:
            raise RuntimeError('Could not read image from camera.')

        self.value = image
        self.start()
    except:
        self.stop()
        raise RuntimeError(
            'Could not initialize camera. Please see error trace.')
```

# JetBot 系统体系结构

- 深度学习模型推理接口

- class ObjectDetector(object):
- 
- def \_\_init\_\_(self, engine\_path, preprocess\_fn=bgr8\_to\_ssd\_input):
- logger = trt.Logger()
- trt.init\_libnvinfer\_plugins(logger, '')
- load\_plugins()
- self.trt\_model = TRTModel(engine\_path, input\_names=[TRT\_INPUT\_NAME],
- output\_names=[TRT\_OUTPUT\_NAME, TRT\_OUTPUT\_NAME + '\_1'])
- self.preprocess\_fn = preprocess\_fn
- 
- def execute(self, \*inputs):
- trt\_outputs = self.trt\_model(self.preprocess\_fn(\*inputs))
- return parse\_boxes(trt\_outputs)
- 
- def \_\_call\_\_(self, \*inputs):
- return self.execute(\*inputs)

## 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
- 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
- 使用TLT导出模型时要注意参数的设置，特别是数据精度以及Tensor维度

链接：<https://pan.baidu.com/s/1TWT5PMYn-VkYoxbVCWkoWA>  
提取码：cdaq



# NVIDIA 深度学习学院 (DLI)

## 人工智能和加速计算实践培训

- 面向开发者、数据科学家和研究人员
- 权威机构和专家强强联合打造专业培训
- 运用前沿技术的端到端、多行业应用开发课程
- 真实经验案例分享，获取现实可用的专业知识
- 完全配置的GPU实时开发环境
- 由具有学科专业知识的DLI认证讲师授课
- 全球开发者培训证书

查看课程 [www.nvidia.cn/dli](http://www.nvidia.cn/dli)

# 停课不停学 高校师生免费领

## 第一项 《在线自主培训课程》

- ▶ 高校学生和教师可领取
- ▶ 任意一门30美元“在线自主培训”课程

[www.nvidia.cn/dli](http://www.nvidia.cn/dli)

- ▶ 此活动每人仅限参加一次



## 第二项 《人工智能教学教材》

- ▶ 高校教师可领取
- ▶ 含讲义文档和视频, 实验室方案, 编程项目, 电子书和更多免费 DLI 课程

## 第三项 《DLI校园大使及培训平台》

- ▶ 高校教师可参加
- ▶ 认证后可免费用DLI平台给学生授课学习



# 更多资源：

# <https://developer.nvidia-china.com>



何琨-Ken

北京 密云



扫一扫上面的二维码图案，加我微信



THANK YOU

