

主讲人 :正点原子团队

硬件平台:正点原子ATK-DLRV1126开发板

版权所有:广州市星翼电子科技有限公司

资料下载: www.openedv.com/docs/index.html

教学平台: www.yuanzige.com

天猫店铺: zhengdianyuanzi.tmall.com

技术论坛: www.openedv.com/forum.php

公众平台:正点原子





■ 模型的预编译

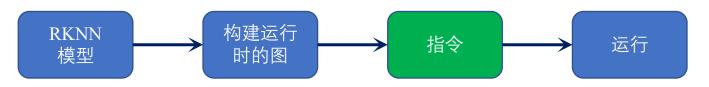


- 1,什么是模型预编译
- 2,注意事项
- 3,模型预编译流程
- 4,模型预编译实战



1、什么是模型预编译

RKNPU 在加载 RKNN 模型时,先根据网络结构构建一个运行时的图,接着转成能在 NPU上运行的指令。对于网络结构比较复杂的模型,这一步会消耗比较多的时间,即模型初始化(init_runtime 或 rknn_init)时间太长。



为了解决这个问题,rknn-toolkit提供了模型预编译功能,预编译功能会提前将RKNN模型转成相应硬件上的运行指令,减少RKNN模型在板端的初始化(init_runtime 或 rknn_init)时间,可以让模型初始化时间降到 1 秒以内。

参考《TensorFlow自训练模型部署到RV1126》视频课程,模型在未进行预编译时,当部署到开发板上跑时,发现模型在加载阶段消耗比较多的时间。

"原子哥"在线教学平台: www.yuanzige.com

技术支持论坛: www.openedv.com/forum.php

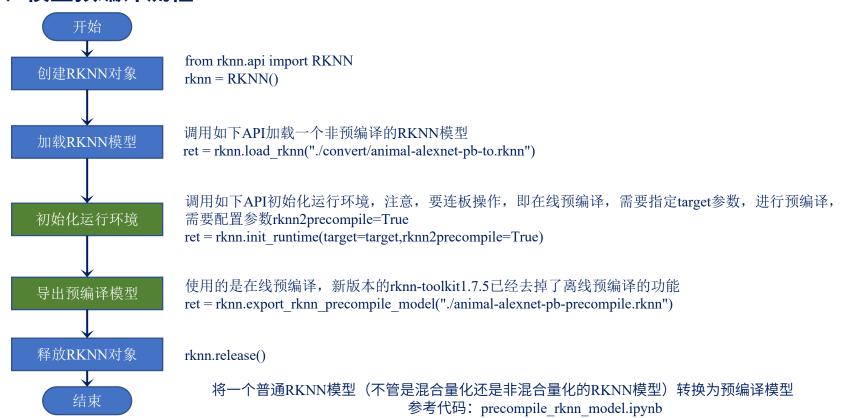


2、注意事项

- ① 模型预编译功能,在rknn-toolkit版本里有,所以适用于RV1126、RV1109、RK1808、RK1806平台(但不适用于RK3399pro平台);
- ② 自rknn-toolkit1.7.3版本之后,去掉了离线预编译功能,仅支持在线预编译功能(即需要ADB连板操作);
 - rknn.build()接口的pre_compile参数已经作废了(即离线预编译功能作废了),请不要配置pre_compile=True
 - rknn.init_runtime()接口的target要指定(即ADB连板操作),且设置参数rknn2precompile=True
- ③ rknn-toolkit2没有模型预编译功能,所以RK3566、RK3568、RK3562、RK3588和RK3588S平台就没有该功能;
- ④ 若模型文件大小比较小,使用ADB连接开发板即可,若模型文件大小比较大,需要开启ntb设备来连接开发板(本人测试经验总结);
- ⑤ 预编译的模型只能在NPU设备上运行,不能在模拟器运行;
- ⑥ NPU mini版本的驱动只支持预编译的RKNN模型,不支持非预编译的RKNN模型;
- ① 如果 NPU 驱动有更新,预编译模型通常也需要重新构建;
- ⑧ RKNN Toolkit 从 1.7.5 版本开始支持量化参数导出功能(导出 RKNN 模型中每一层的量化参数),对于预编译模型,只能获取输入、输出层的量化参数;



3、模型预编译流程



"原子哥"在线教学平台: www.yuanzige.com

技术支持论坛: www.openedv.com/forum.php



3、模型预编译流程



from rknn.api import RKNN rknn = RKNN()

调用rknn.config()配置模型,主要进行:输入的预处理(如归一化操作、通道顺序的调整)、量化类型选择、量化参数优化算法选择、模型优化等级配置、指定 RKNN 模型目标运行平台等

调用rknn.load tensorflow()、rknn.load tflite()、rknn.load pytorch()、rknn.load onnx()...... 等等接口,加载对应深度学习框架下的模型

调用rknn.build (),根据加载的模型结构及权重参数,构建对应的RKNN模型,在这一部分会进行:是否对模型进行 量化、确定量化的数据集、要推理的批量大小

调用如下API初始化运行环境,注意,要连板操作,即在线预编译,需要指定target参数,进行预编译,需要配置参数 rknn2precompile=True ret = rknn.init runtime(target=target,rknn2precompile=True)

使用的是在线预编译,新版本的rknn-toolkit1.7.5已经去掉了离线预编译的功能 ret = rknn.export rknn precompile model("./animal-alexnet-pb-precompile.rknn")

rknn.release()

将一个其它深度学习框架下的模型(非RKNN模型)转换为预编译模型 参考代码: precompile no rknn model.ipynb

"原子哥"在线教学平台:www.yuanzige.com

技术支持论坛: www.openedv.com/forum.php



4、模型预编译实战

参考demo:

将一个普通RKNN模型(不管是混合量化还是非混合量化的RKNN模型)转换为预编译模型

参考代码: precompile rknn model.ipynb

将一个其它深度学习框架下的模型(非RKNN模型)转换为预编译模型

参考代码: precompile_no_rknn_model.ipynb

若要在开发板端查看日志,请在开发板端执行如下命令:

export RKNN LOG LEVEL=5 && export RKNN SERVER LOGLEVEL=5 && restart rknn.sh

开启ntb设备的方法: 在开发板端运行命令start usb.sh ntb即可





版权所有: 广州市星翼电子科技有限公司

天猫店铺: https://zhengdianyuanzi.tmall.com