



HiAI DDK V320

系统调试工具使用说明书

文档版本 02
发布日期 2020-06-09

华为技术有限公司



版权所有 © 华为技术有限公司 2019。 保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

法律声明

本文所描述内容可能包含但不限于对非华为或开源软件的介绍或引用，使用它们时请遵循对方的版权要求。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为 HiAI 申请方式

发送申请邮件到邮箱：developer@huawei.com

邮件名称：HUAWEI HiAI+公司名称+产品名称

邮件正文：合作公司+联系人+联系方式+联系邮箱

我们将在收到邮件的 5 个工作日内邮件给您反馈结果，请您注意查收。

官网地址 <https://developer.huawei.com/consumer/cn/>

前言

概述

本文提供对系统调试工具 model_run_tool 和 data_proc_tool 的功能和使用说明。

目前只支持 NPU 场景。支持平台：Kirin990/Kirin820/Kirin985。

修改记录

日期	修订版本	修改描述
2020-06-09	02	补充性能统计数据文件说明
2020-04-11	01	第一次正式发布

目 录

前言..... ii

1 概述.....1

2 使用说明.....2

2.1 操作流程 2

2.2 详细参数说明 2

2.3 示例 5

1 概述

系统调试工具可以对模型推理进行性能统计，并获取性能数据。

工具主要包括两个可执行程序：`model_run_tool` 和 `data_proc_tool`。

`model_run_tool` 用于开启底层的性能统计(Profiling)功能，并执行模型推理，生成性能统计数据。

`data_proc_tool` 将 `model_run_tool` 运行后生成的性能统计数据转换成 csv 文件。

2 使用说明

2.1 操作流程

须知

model_run_tool 开启 Profiling 并进行推理，只在当次推理生效。

- 步骤 1 解压 ddk。将 ddk/tools/tools_sysdbg 目录下的所有文件，PUSH 到手机环境中，放在 /data/local/tmp 目录下（注意：必须放到/data/local/tmp 或其子目录下）。
- 步骤 2 添加动态链接库路径，与步骤 1 中存放路径保持一致。执行命令：export LD_LIBRARY_PATH={步骤 1 的 so 存放路径}。
- 步骤 3 添加 model_run_tool 和 data_proc_tool 的执行权限。执行命令：chmod +x model_run_tool data_proc_tool。
- 步骤 4 执行 model_run_tool（请参考表 2-1）完成推理并生成中间数据。
- 步骤 5 执行 data_proc_tool（请参考表 2-2）生成分析数据。
- 步骤 6 分析数据，请参考表 2-3 和表 2-4。

----结束

2.2 详细参数说明

表2-1 model_run_tool 参数说明

参数名称	参数描述	是否必选	默认值
--model	离线模型路径(注意：需在/data/local/tmp 或其子目录下)。例如： --model=/data/local/tmp/hiai_model.om	是	不涉及

参数名称	参数描述	是否必选	默认值
--input	推理的输入数据文件路径(注意：目录需在/data/local/tmp或其子目录下) 输入文件可以有多个，以逗号“,”分隔，逗号前后可以有空格。例如： <pre>--input=/data/local/tmp/input1.dat -- input="./data/local/tmp/input1.dat, ./data/local/tmp/input2.dat, ./data/local/tmp/input2.dat"</pre>	否	生成随机输入数据
--input_type	指定输入 tensor 的数据类型。 支持以下值： FP32, FP16, INT32, UINT8, INT8, INT16, BOOL, INT64, UINT32, DOUBLE 多输入时，以逗号“,”分隔。 <i>注：当输入 tensor 的数据类型为 FP32/INT32/UINT32 时，可以不指定此参数。</i>	否	不涉及
--output_type	指定输出 tensor 的数据类型。 支持以下值： FP32, FP16, INT32, UINT8, INT8, INT16, BOOL, INT64, UINT32, DOUBLE 多输出时，以逗号“,”分隔。 <i>注：当输出 tensor 的数据类型为 FP32/INT32/UINT32 时，可以不指定此参数。</i>	否	不涉及
--output_dir	推理输出结果存放的目录(注意：需在/data/local/tmp或其子目录下)。 <i>注：目录必须已存在，不支持动态创建。</i> 推理结果存储为二进制文件，自动命名为 output_\${index}。	否	当前目录
--enable_item	开启或关闭 Profiling： --enable_item=0 或--enable_item=off，不开启 Profiling； --enable_item=1 或--enable_item= profiling，开启 Profiling； 其它值无效，返回错误。	否	0/off
--dev_perf	模型运行的算力要求等级(模型推理使用) --dev_perf=1 要求等级为 LOW； --dev_perf=2 要求等级为 NORMAL； --dev_perf=3 要求等级为 HIGH； --dev_perf=4 要求等级为 EXTREME；	否	3

参数名称	参数描述	是否必选	默认值
--aipp_format	AIPP 模型输入图片的格式。 <i>注：使用兼容格式（旧版AIPP 格式）生成的模型，需要显示指定此参数。</i> 可选值为：YUV420SP_U8, XRGB8888_U8, YUV400_U8, ARGB8888_U8, YUYV_U8, YUV422SP_U8, AYUV444_U8。 当有多输入文件时，会对应有多种格式，格式之间以“,” 分隔，如--aipp_format="YUV420SP_U8, XRGB8888_U8, YUV400_U8"。	否	不涉及
--version	查询版本号	否	不涉及
--help	打印帮助和使用信息	否	不涉及

表2-2 data_proc_tool 参数说明

参数名称	参数描述	是否必选	默认值
--result_path	性能统计文件输出目录(注意：目录需在/data/local/tmp或其子目录下)。保存经转换后的 csv 文件。 说明 <ul style="list-style-type: none"> 目录必须已存在，不支持动态创建。 生成的 csv 文件分别为：模型执行流程表 (*_model.csv) 和算子执行记录表 (*_op.csv)。 	否	当前目录
--version	查询版本号	否	不涉及
--help	打印帮助和使用信息	否	不涉及

表2-3 模型执行流程表说明

项目	描述
item event	事件类型，有如下定义： <ul style="list-style-type: none"> load begin: 模型加载开始 load end: 模型加载结束 inference begin: 模型推理开始

项目	描述
	<ul style="list-style-type: none">inference end: 模型推理结束unload begin: 模型卸载开始unload end: 模型卸载结束 <p>说明</p> <p>工具执行过程中，模型推理会执行 2 次。inference begin 和 inference end 会对应生成 2 条记录。</p>
enter time	事件的上报时间。
total time	各个事件 begin->end 的执行时间。单位：us。 仅在 end 处记录，begin 处固定为 0。

表2-4 算子执行记录表说明

项目	描述
op name	算子名称。 由执行的模型结构决定。例如：transdata_for_nd_0。
start time	算子执行的开始时间。
total time	算子执行耗时。单位：us。 注意：同一次推理中，如遇到算子名重复的 NPU 算子，需将重复记录的时间加总得到总时间。
device type	执行设备类型。 <ul style="list-style-type: none">npucpu

2.3 示例

查询工具版本号

```
./model_run_tool --version
./data_proc_tool --version
```

打印帮助信息

```
./model_run_tool --help
./data_proc_tool --help
```

普通模型（多输入）推理

```
./model_run_tool --model=./offline.om --input
```

```
= "./input1.dat,./input2.dat,./input3.dat" --output_dir=./output
```

AIPP模型(单输入)推理

```
./model_run_tool --model=./aipp_offline.om --input=image.bin --  
output_dir=./output --aipp_format=YUV420SP_U8
```

AIPP模型(多输入)推理

```
./model_run_tool --model=./aipp_offline.om --  
input="./image1.bin,./image2.bin,./image3.bin" --output_dir=./output --  
aipp_format="YUV420SP_U8,XRGB8888_U8,YUV400_U8"
```

开启 Profiling 推理。最终在 result_path 指定的目录下生成性能数据(csv 格式)

```
./model_run_tool --model=./offline.om --input_files="./input1.dat" --  
output_dir=./output --enable_item=1  
./data_proc_tool --result_path=./prof_data
```