

Ascend 310

ReleaseNotes

文档版本

01

发布日期

2019-03-08



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。

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默示的声明或保证。

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。

华为技术有限公司

地址：深圳市龙岗区坂田华为总部办公楼 邮编：518129

网址：<http://www.huawei.com>

客户服务邮箱：support@huawei.com

客户服务电话：4008302118

目 录

1 简介.....	1
2 版本配套说明.....	2
2.1 软件信息说明.....	2
2.2 AI 平台说明.....	3
2.3 硬件平台说明.....	4
2.4 配套资料说明.....	4
3 特性.....	7
4 解决的问题.....	10
5 遗留问题.....	11
6 版本使用注意事项.....	12
A 安装 Mind Studio 后 msvp 界面无法导入结果.....	14
B 历史版本接口变更记录.....	15

1 简介

本文介绍了Ascend 310及Mind Studio相关版本信息和特性简介。开发者可以使用Mind Studio用于仿真形态或者ASIC等环境下模型运行流程编排开发和TE（Tensor Engine）(类TVM)算子开发。

2 版本配套说明

[2.1 软件信息说明](#)

[2.2 AI平台说明](#)

[2.3 硬件平台说明](#)

[2.4 配套资料说明](#)

2.1 软件信息说明

主要写软件配套的限制、版本说明等。

表 2-1 软件版本信息

软件名称	版本	说明
Mind Studio	1.1.T8.B750	AI 开发相关的IDE。
Ascend 310	B750	Ascend 310软件包。

表 2-2 软件包信息

包名	包说明
mini_mind_studio_Ubuntu.rar	Mind Studio在Ubuntu系统下安装的工具包。用于模型运行流程编排开发和TE（Tensor Engine）（类TVM）算子开发，提供API、库、工具链等开发组件。适用于x86_64 Ubuntu+PCIe以及3559场景。
mini_mind_studio_centOS.rar	Mind Studio在centOS系统下安装的工具包。用于模型运行流程编排开发和TE算子开发，提供API、库、工具链等开发组件。适用于x86_64 CentOS+PCIe场景。
mini_asic_pcie_Ubuntu.rar	PCIe板在Ubuntu系统下的软件包。
mini_asic_pcie_centOS.rar	PCIe板在centOS系统下的软件包。

包名	包说明
mini_asic_3559_it.rar	3559软件包。

表 2-3 周边配套软件版本说明

软件名称	版本	说明
Ubuntu版本	Ubuntu 16.04.3	linux版本。
centOS版本	centOS 7.4	linux版本。
浏览器	google chrome 67.0.3396.87 及以上	浏览器版本。

2.2 AI 平台说明

主要对本版本支持Ascend 310的能力进行说明。

表 2-4 仿真平台支持的网络模型

仿真平台支持的网络模型
ResNet-18
VGG16
VGG19
ResNet-50
ResNet-101
ResNet-152
ResNext-50
ResNext-101

表 2-5 ASIC 平台支持的网络模型

网络类型	ASIC平台支持的网络模型
分类网络	VGG16
	VGG19
	AlexNet
	ResNet-18

网络类型	ASIC平台支持的网络模型
	ResNet-50
	ResNet-101
	ResNet-152
	ResNext-50
	ResNext-101
	Inception-v3
	Inception-v4
	DenseNet
	Xception
	MobileNet V1/V2
检测网络	FasterRCNN
	YOLO V2
	SSD

2.3 硬件平台说明

表 2-6 硬件配套说明

硬件平台	配置	说明
Linux宿主机	单用户场景推荐：内存 ≥4G。	支持Linux服务器或者虚拟机。
PCIe	华为配套发布的PCIe单板。	无。

说明

- 若Linux宿主机内存为4G，在Mind Studio中进行模型转换和仿真运行时，建议Model文件大小不超过350M，如果超过此规格，操作系统可能会因为超过安全内存阈值而工作不稳定。
- 若Linux宿主机配置升级，比如8G内存，则相应支持的操作对象规格按比例提升。
例如，内存由4G升级到8G，则Model文件建议大小不超过700M。

2.4 配套资料说明

随版本发布的版本如表2-7所示。

表 2-7 配套资料说明

资料名称	说明	下载方法
Ascend 310 软件安装指南（run）	介绍Ascend 310 run软件包如何安装、配置、卸载和升级等流程，同时提供了常见的问题解答及故障处理方法。	随版本发布，同软件一起获取。
Ascend 310 软件升级指导（run）	介绍Ascend 310 run软件包如何升级，以及升级的流程，同时提供了常见的问题解答及故障处理方法。	
Ascend 310 Mind Studio工具安装指南（Ubuntu）	介绍Ubuntu方式下Mind Studio工具如何安装以及升级，同时提供了常见的问题解答及故障处理方法。	
Ascend 310 Mind Studio工具安装指南（CentOS）	介绍CentOS方式下Mind Studio工具如何安装以及升级，同时提供了常见的问题解答及故障处理方法。	
Ascend 310 Mind Studio基本操作	介绍Mind Studio基本的操作，包括工具管理DDK样例工程导入、数据集管理、模型管理。	
Ascend 310 Mind Studio快速入门	通过介绍工具的功能以及开发流程，再到深入的流程编排、离线模型转换，帮助开发人员快速入门。	
Ascend 310 HiAI Engine开发指导	介绍使用HiAI Engine编程框架的两种开发方式，基于图形化拖拽界面和基于用户自定义工程。	
Ascend 310 Mind Studio开发辅助工具	介绍Mind Studio常用工具，包括对比工具、日志工具、Profiling、黑匣子、DVPP调优工具。	
Ascend 310 Mind Studio常见问题处理	介绍使用Mind Studio工具过程中的常见问题。	
Ascend 310 TensorEngine API参考	Tensor Engine API接口参考。	
Ascend 310 DVPP API参考	DVPP API接口参考。	
Ascend 310 HiAI Engine API参考	HiAIEngine API接口参考。	
Ascend 310 AI CPU API参考	AI CPU对外提供的接口。	
Ascend 310 TE自定义算子开发指导	详细介绍如何开发TE自定义算子和插件，并通过加载算子插件完成模型转化，最终执行流程编排实现数字概率累加验证的功能的端到端流程。	
Ascend 310 C++算子开发指导	详细介绍如何开发C++自定义算子和插件，并通过加载算子插件完成模型转化，最终执行流程编排实现数字概率累加验证的功能的端到端流程。	

资料名称	说明	下载方法
Ascend 310 网络模型配置参考	介绍目前支持的网络模型。	
Ascend 310 算法移植指导	介绍如何将算法移植到Ascend 310。	
Ascend 310 源码编译安装指南	介绍3559和Atlas 300软件包二进制交付件的版本与用户环境不一致时（gcc/glibc/kernel/kernel配置文件有差异），如何进行源码编译和安装。	
Ascend 310 高性能应用编程用户手册	介绍在HiAI框架上实现高性能应用的约束和建议，帮助用户快速理解样例，使用户能够快速构建起自己的高性能应用。	
Ascend 310 Module User Guide	介绍3559RC如何使用。	
Ascend 310 Module Upgrade Guide	介绍3559RC如何升级。	
Ascend 310 DDK 样例使用指导	介绍DDK中的样例如何使用。	

3 特性

提供的特性如下：

- **DDK**

DDK是Mind解决方案提供的开发者套件包，Mind Studio通过安装DDK后获得Mind开发必需的API、库、工具链等开发组件。

- **工程管理**

用户可以新建、删除Mind工程，并可以更新工程属性。新建的工程根据选择的不同的DDK版本支持不同的产品形态。在工程中可以新建目录、新建源代码文件；编译、运行等功能均以工程为单位进行操作。

- **模型转换**

用户可以将主流的网络模型转换为Mind类型的模型，转换会将模型进行优化，转换好的模型可以直接转换到工程目录，或者保存在Model Zoo中供其他工程使用。

- **Mind网络图形化编排**

用户可以使用图形化拖拽的方式“零编码”完成模型网络编排开发工作。内置节点包含：数据集、模型、图形预处理、模型运行引擎、结果后处理等节点。另外，数据集和模型节点用户可以自定义添加。编排后的图形可以保存，代码生成、一键式编译运行、查看网络运作最终结果。

- **TE算子开发**

TE提供了基于TVM开发自定义算子的能力，在Mind Studio中使用TE模板化接口和TVM接口开发自定义算子，并且具备算子编译、自动生成单算子代码、单算子仿真运行能力，方便用户进行自定义算子开发。除TVM原生接口外，还提供简单易用的模板化接口，方便开发自定义算子，提升开发效率。模板化接口主要包括element-wise类和reduction类操作。

- **构建**

提供编译选项配置界面，可根据工程形态和编译选项自动生成makefile并执行编译过程，输出可执行程序。

- **运行**

提供运行配置界面，通过配置启动参数后可以启动运行。

- **支持多卡多芯特性**

支持选择device进行调试。

日志变更

- 支持通过如下命令设置TS、LPM3、TS-DUMP模块dlog的日志级别：
IDE-daemon-client --host 服务器的IP地址:端口号 --log 'SetLogLevel(0)[error]';

权限变更

- 禁止root用户通过ssh远程登陆device。
- 用户远程使用ssh登陆device需要使用HwHiAiUser用户，密码Huawei2012#。
- 非root用户可以使用scp。
- 删除sftp-server，不支持sftp。
- 删除useradd/groupadd/userdel/groupdel命令。

DVPP 变更

- VPC新接口当输出位置也即贴图区域左上角不在（0，0）时，1080P也进行4核加速；
- VPC新接口增加抠图个数限制，一个图最多抠256个；
- VPC新接口支持VDEC输出的压缩格式图片16倍放大功能；
- 修改vdec在内存不足时阻塞用户进程不能正常退出；
- VPC老接口修改scaler中的参数配置；
- DVPP除VDEC模块申请大页内存或4K内存时做内存建页；
- 去除VPC参数集路径的校验。
详情请参见《Ascend 310 DVPP API参考》。

HiAI Engine 变更

- B750版本新增GraphImpl::RegisterEventHandle接口，详细说明请参见《Ascend 310 HiAI Engine API参考》。

MindStudio 变更

- 在安装前通过root用户执行提升权限脚本。安装失败时，自动降低权限，再次安装时，需要再次用root用户执行提升权限脚本。
- 新增依赖安装：haveged python-skimage python3-skimage python-pip python3-pip。
详细请参见《Ascend 310 Mind Studio工具安装指南》。
- logout超时时间可以设置，默认30分钟，可以设置成永不超时。
- 重新登录速度变快，不会每次都重启workspace。

框架变更

- 支持全流程零“copy”，为了避免算法推理内部可能出现的内存拷贝，在调用模型管家Process接口时，建议输入数据（输入数据一般可直接使用框架传入的内存，该内存是由框架通过HIAI_DMAlloc申请得到）及输出数据都通过HIAI_DMAlloc接口申请，并通过CreateTensor接口将内存传给推理模块，这样就能够使能算法推理的零拷贝机制，优化Process时间。
- 新增模型管家接口：AITensorFactory::CreateTensor，详细说明请参见《Ascend 310 HiAI Engine API参考》。

- 删除AIStatus SetModelShareUserMem(const std::vector<AIModelDescription>& model_descs)接口。
- 数据类型TensorDimension变更：描述Tensor尺寸由NCHW四个变量改为1个vector变量，详细说明请参见《Ascend 310 HiAI Engine API参考》。

Driver 变更

- RUN包安装升级使用新增用户HwHiAiUser，详细请参见《Ascend 310 软件安装指南（run）》。

4 解决的问题

问题单号	1
涉及模块	框架
问题描述	灰度图模型转换失败
问题级别	一般。

5 遗留问题

问题单号	1
涉及模块	模型
问题描述	常稳压测下人脸识别业务有概率性异常。
问题级别	提示。
影响分析	偶现问题。
规避措施	后续版本解决。

问题单号	2
涉及模块	推理
问题描述	特定场景下概率性出现多batch进行推理申请内存失败。
问题级别	提示。
影响分析	偶现问题。
规避措施	后续版本解决。

问题单号	3
涉及模块	DVPP
问题描述	常稳压测下DVPP的vedc接口出现极小概率coredump。
问题级别	提示。
影响分析	偶现问题。
规避措施	后续版本解决。

6 版本使用注意事项

- B750之前版本暂不支持Mind Studio在线升级，如要升级Mind Studio，请参见《Ascend 310 Mind Studio工具安装指南（Ubuntu）》卸载后再进行新版本安装，卸载前请注意保存mind studio安装目录下的代码和数据。
- B750版本修改了root运行的权限，升级后如果有老版本工程中含有运行结果文件（out/result_files文件夹），请删除后重新执行。否则无法运行成功。
- B750版本su命令加固切换账号不继承上个账号的环境变量，使用su命令必须加-，例如 su - test，否则可能存在bash下面命令找不到的错误；
- B730版本msvp对数据库结构有所调整，新版本已不再向前兼容老版本的部分备份文件（arm.db dump.rdb），如果安装后出现msvp界面无法导入结果的问题，请参见[A 安装Mind Studio后msvp界面无法导入结果](#)解决。
- B650版本到B750版本对外接口变更请参见[B 历史版本接口变更记录](#)。
- 由于dvpp接口修改了对调用时入参的奇偶校验，在上层代码适配后，B720以前的旧工程可能会出现编译失败的情况（之前版本不校验）；如果要使用旧工程（不支持前向兼容），需要用户删除前处理节点，再重新拖拽新的前处理节点，配置相应的参数后，再编译并运行工程。
- B700之后版本不支持兼容之前版本创建的仿真推理工程（PVMModel，CAModel，Caffe），当前版本仅支持创建Caffe仿真推理工程。
- five2four算子在B670与B680模型不兼容：B700与B670的模型不会因为five2four算子的原因而不兼容，但B680、B690与B670、B700会因为five2four算子的原因而不兼容。若使用B670模型，则无影响。如果使用B680、B690版本的模型，则需要重新生成模型。
- run包升级说明：B650版本可以直接升级到当前版本，请参见《Ascend 310 软件升级指导（run）》文档升级；B630及之前的版本，无法直接用run包升级到B650或者B650之后的版本，参见《Ascend 310软件安装指南（run）》中“常用操作 > 卸载”，使用软件当前版本的run包将老版本卸载，然后重启，再使用需要升级的新版本run包进行安装。
- 离线模型转换可以将Caffe、Caffe2、Tensorflow、MxNet等业界开源的神经网络模型转换为华为NPU芯片支持的网络模型，当前版本仅支持Caffe，其他后续版本支持。
- 3559RC形态支持用户自己编好的程序，导入后通过运行配置实现3559RC上的运行，功能正常。但目前不支持直接编排、编译和运行。
- 3559RC形态下，当您尝试自己编译主机侧（3559侧）内核镜像时，请注意不要打开以下内核驱动的编译开关，包括：

```
"HiSilicon Hip05 and Hip06 SoCs PCIe controllers"  
"HiSilicon Hip05 and Hip06 and Hip07 SoCs PCIe controllers"  
"HiSilicon Hip05 and Hip06 and Hip07 SoCs ACPI PCIe controllers"
```

以上编译选项开关出现在内核4.9及以后的内核版本中，不同的内核版本表现不同，可能为其中一个。开关在menuconfig中的位置为：

“Bus options (PCI etc.) > PCI host controller drivers”。

- centOS中虚拟网口的使用说明

如果客户使用高级服务NetworkManager：Redhat中高级服务NetworkManager和基本服务Network的已知冲突问题，会造成虚拟网口的IP丢失。

- 模型拖拉组件兼容性说明：B620版本中模型拖拉组件新增batch_size属性，而B620之前版本无此属性导致B620版本模型拖拉组件无法向前兼容，即在B610或之前版本中生成的模型拖拉组件无法在B620及之后版本中使用。
- 后处理拖拉组件兼容性说明：B620版本中后处理引擎模块获取推理结果的目录结构有变动，导致B620之前版本生成的后处理拖拉组件在B620版本上运行无法可视化推理结果，即在B610或之前版本中生成的后处理拖拉组件无法在B620及之后版本中使用。

A 安装 Mind Studio 后 msvp 界面无法导入结果

问题描述

B730版本msvp对数据库结构有所调整，新版本已不再向前兼容老版本的部分备份文件（arm.db和dump.rdb），如果安装后出现msvp界面无法导入结果，是因为mind studio安装时默认调用profiling恢复备份接口，将老版本的两个数据库文件覆盖到了新版本中，导致msvp功能异常。如出现该问题，请参见如下步骤解决。

解决方案

- 如果未安装Mind Studio:
 - a. 安装B730版本前，请先将备份目录下profiler文件夹内的arm.db（默认路径：~/wsbackup/profiler/arm.db）以及dump.rdb（默认路径：~/wsbackup/profiler/dump.rdb）删除。
 - b. 如需要在msvp界面展示之前版本的数据，请在安装后把备份目录下profiler文件夹下的projects文件夹（默认路径：~/wsbackup/profiler/projects）拷贝到安装目录下（默认路径：~/tools），并从msvp界面选择import重新导入采集结果。
- 如果Mind Studio已经安装:
 - a. 请进入安装目录/profiler目录后执行如下命令（全部复制后执行，不要单行复制执行）：
sudo sqlite3 -init initDB.sql visualization/web/arm.db <<EOF
.quit
EOF
 - b. 按照[2.如需要在msvp界面展示之前版本的数据，请在...](#)中的方法重新导入数据。

B 历史版本接口变更记录

表 B-1 DVPP API 参考接口变更

接口名称	操作	备注
DvppCtl	删除	NEW VPC
	新增	所在章节：VPC 和 CMDLIST 统一接口 （增加注意项，如果参数涉及文件路径，调用者需要确保路径是正确路径）
	修改	对于VPC入参resize_param_in_msg、入参vpc_in_msg，增加注意说明，同时优化调用示例。
	修改	优化VPC下CMDLIST结构体的说明及调用示例。
	修改	修改JPEGE下的出参名称，同时优化调用示例。
	修改	修改JPEGD下的入参、出参，同时优化调用示例。
	修改	优化PNGD下的入参、出参描述，同时优化调用示例。
	修改	优化VENC下的调用示例。
VdecCtl	修改	对于VDEC入参vdec_in_msg和出参vdec_out_msg，增加注意说明，同时优化调用示例。

表 B-2 HiAI Engine API 参考接口变更

接口名称	操作	备注
Engine() virtual ~Engine()	新增	所在章节：Engine实现接口。
Engine::SetDataRecvFunctor	新增	所在章节：Engine实现接口。

接口名称	操作	备注
模型管家接口的数据类型说明	新增	所在章节：功能接口>数据类型。
模型管家编译依赖的接口	新增	所在章节：功能接口>其他用于编译依赖的接口。
uint32_t Graph::GetDeviceID();	新增	所在章节：辅助接口>数据获取接口。
tatic HIAI_StatusT Graph::CreateGraph(const std::string& configFile)	修改	增加参数的取值范围的说明。
static HIAI_StatusT Graph::CreateGraph(const GraphConfigList& graphConfig)	修改	增加参数的取值范围的说明。
Graph::SetPublicKeyForSignatur eEncryption	新增	所在章节：辅助接口>其他。
AIModelManager::Init	修改	增加model_name、model_path的限制说明。
HIAI_CreateGraph Graph::CreateGraph（根据配置 文件创建Graph） Graph::ParseConfigFile(const std::string& configFile, GraphConfigList& graphConfigList)	修改	增加注意事项，如果参数涉及文件路径，调用者需要确保路径是正确路径。
HIAI_C_SendData	修改	增加入参的注意事项。
HIAIMemory::HIAI_Dmalloc	修改	增加接口使用的注意事项。
HIAI_StatusT Graph:: GetNextGraphId(const uint32_t curGraphID,uint32_t * nextGraphID);	删除	删除接口。
AISatus SetModelShareUserMem(const std::vector<AIModelDescription >& model_descs)	删除	删除接口。
AITensorFactory::CreateTensor	新增	所在章节：功能接口>离线模型管家。
GraphImpl::RegisterEventHandl e	新增	所在章节：辅助接口>时间注册接口。