MindStudio 5.0.RC1

Release Note

文档版本	01
发布日期	2022-05-17

华为技术有限公司

版权所有 © 华为技术有限公司 2022。 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。

目 录

1 认识 MindStudio.....1
1.1 关于 MindStudio.....1
1.2 关键概念.....3
2 版本配套关系.....6
3 新增特性.....7
4 Windows 特性支持清单.....9

版权所有 © 华为技术有限公司

1 认识 MindStudio

1.1 关于MindStudio
1.2 关键概念

1.1 关于 MindStudio

MindStudio提供您在AI开发所需的一站式开发环境，支持模型开发、算子开发以及应用开发三个主流程中的开发任务。

依靠模型可视化、算力测试、IDE本地仿真调试等功能，MindStudio能够帮助您在一个工具上就能高效便捷地完成AI应用开发。

MindStudio采用了插件化扩展机制，开发者可以通过开发插件来扩展已有功能。

说明

《MindStudio用户指南》需要根据产品选择其他相关手册适配使用，涉及安装以及开发相关辅助手册。

- Atlas 200 DK请参考《**Atlas 200 DK开发者套件**》。
- 其他产品请参考《**昇腾CANN**》。

功能简介

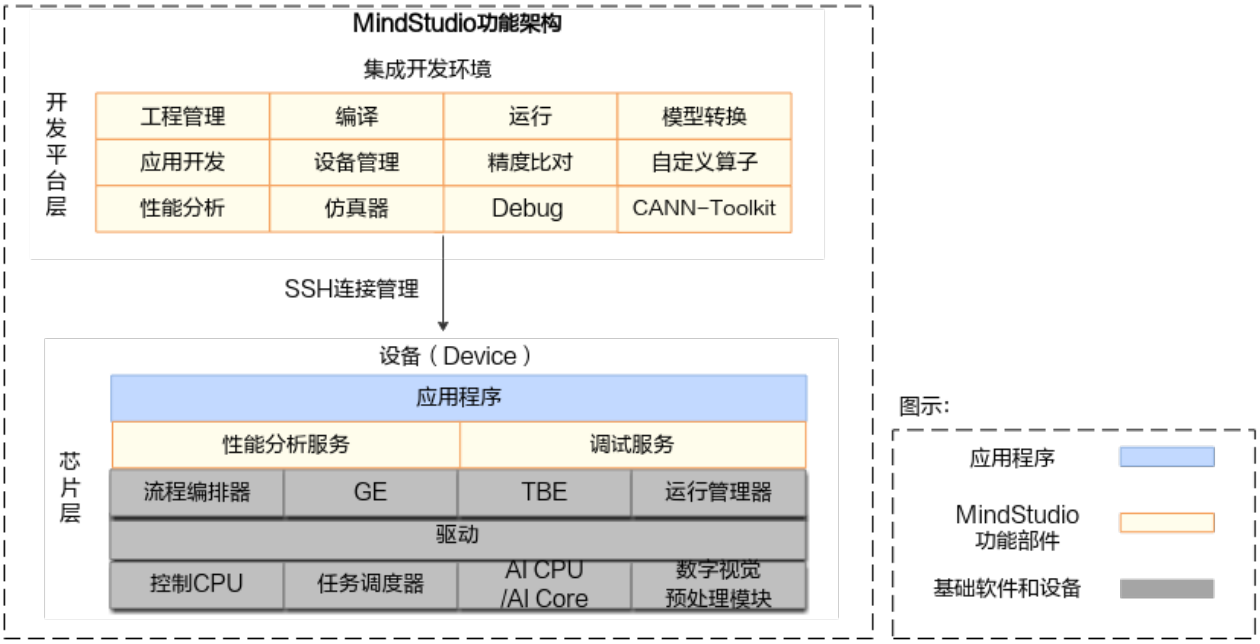
- 针对安装与部署，MindStudio提供多种部署方式，支持多种主流操作系统，为开发者提供最大便利。
- 针对网络模型的开发，MindStudio支持TensorFlow、Pytorch、MindSpore框架的模型训练，支持多种主流框架的模型转换。集成了训练可视化、脚本转换、模型转换、精度比对等工具，提升了网络模型移植、分析和优化的效率。
- 针对算子开发，MindStudio提供包含UT测试、ST测试、TIK算子调试等的全套算子开发流程。支持TensorFlow、PyTorch、MindSpore等多种主流框架的TBE和AI CPU自定义算子开发。
- 针对应用开发，MindStudio集成了Profiling性能调优、编译器、MindX SDK的应用开发、可视化pipeline业务流编排等工具，为开发者提供了图形化的集成开发环境，通过MindStudio能够进行工程管理、编译、调试、性能分析等全流程开发，能够很大程度提高开发效率。

版权所有 © 华为技术有限公司

功能框架

MindStudio功能框架如图1-1所示，目前含有的工具链包括：模型转换工具、模型训练工具、自定义算子开发工具、应用开发工具、工程管理工具、编译工具、流程编排工具、精度比对工具、日志管理工具、性能分析工具、设备管理工具等多种工具。

图 1-1 工具链功能架构



工具功能

MindStudio工具中的主要几个功能特性如下：

- 工程管理：为开发人员提供创建工程、打开工程、关闭工程、删除工程、新增工程文件目录和属性设置等功能。
- SSH管理：为开发人员提供新增SSH连接、删除SSH连接、修改SSH连接、加密SSH密码和修改SSH密码保存方式等功能。
- 应用开发：针对业务流程开发人员，MindStudio工具提供基于AscendCL（Ascend Computing Language）和集成MindX SDK的应用开发编程方式，编程后的编译、运行、结果显示等一站式服务让流程开发更加智能化，可以让开发者快速上手。
- 自定义算子开发：提供了基于TBE和AI CPU的算子编程开发的集成开发环境，让不同平台下的算子移植更加便捷，适配昇腾AI处理器的速度更快。
- 离线模型转换：训练好的第三方网络模型可以直接通过离线模型工具导入并转换成离线模型，并可一键式自动生成模型接口，方便开发者基于模型接口进行编程，同时也提供了离线模型的可视化功能。
- 日志管理：MindStudio为昇腾AI处理器提供了覆盖全系统的日志收集与日志分析解决方案，提升运行时算法问题的定位效率。提供了统一形式的跨平台日志可视化分析能力及运行时诊断能力，提升日志分析系统的易用性。
- 性能分析：MindStudio以图形界面呈现方式，实现针对主机和设备上多节点、多模块异构体系的高效、易用、可灵活扩展的系统化性能分析，以及针对昇腾AI处理器的性能和功耗的同步分析，满足算法优化对系统性能分析的需求。

版权所有 © 华为技术有限公司

- 设备管理：MindStudio提供设备管理工具，实现对连接到主机上的设备的管理功能。
- 精度比对：可以用来比对自有模型算子的运算结果与Caffe、TensorFlow、ONNX标准算子的运算结果，以便用来确认神经网络运算误差发生的原因。
- 开发工具包的安装与管理：为开发者提供基于昇腾AI处理器的相关算法开发套件包Ascend-cann-toolkit，旨在帮助开发者进行快速、高效的人工智能算法开发。开发者可以将开发套件包安装到MindStudio上，使用MindStudio进行快速开发。Ascend-cann-toolkit包含了基于昇腾AI处理器开发依赖的头文件和库文件、编译工具链、调优工具等。

1.2 关键概念

Host/Device

在异构计算架构中，昇腾AI处理器与CPU通过PCIe总线连接在一起来协同工作：

- Host：CPU所在位置称为主机端（Host），是指与昇腾AI处理器所在硬件设备相连接的X86服务器、ARM服务器或者WindowsPC，利用昇腾AI处理器提供的NN（Neural-Network）计算能力完成业务。如**图1-1**中的开发平台层。
- Device：是指安装了昇腾AI处理器的硬件设备，利用PCIe接口与服务器连接，为服务器提供NN计算能力。如**图1-1**中的芯片层。

对于Ascend RC产品形态，昇腾AI处理器所在硬件设备与与之相连接的ARM服务器合设，统称为Host。

模型转换

用户使用Caffe/TensorFlow等框架训练好的模型，可通过ATC工具将其转换为昇腾AI处理器支持的离线模型，模型转换过程中可以实现算子调度的优化、权重数据重排、内存使用优化等，可以脱离设备完成模型的预处理。

算子&模型速查

算子&模型速查工具帮助客户了解当前版本CANN支持的算子和模型信息。

MindConverter

原有的MindConverter功能由X2MindSpore代替，MindConverter功能不再提供。MindConverter为MindInsight的子模块，如需使用MindConverter功能，可参考**MindInsight安装指导**安装MindInsight，再使用MindConverter功能，MindConverter具体使用方法可参考**使用MindConverter迁移模型定义脚本**。

X2MindSpore

脚本转换工具X2MindSpore可将PyTorch和TensorFlow 2.x脚本转换成MindSpore代码。脚本转换工具根据适配规则，对用户脚本进行转换，大幅度提高了脚本迁移速度，降低了开发者的工作量。

版权所有 © 华为技术有限公司

TensorFlow/PyTorch GPU2Ascend

NPU是AI算力的发展趋势，但是目前训练和在线推理脚本大多还基于GPU。由于NPU与GPU的架构差异，基于GPU的训练和在线推理脚本不能直接在NPU上使用，需要转换为支持NPU的脚本后才能使用。

精度比对

在模型转换过程中对模型进行了优化，包括算子消除、算子融合、算子拆分，这些动作可能会造成自有实现的算子运算结果与业界标准算子（如Caffe、TensorFlow、ONNX）运算结果存在偏差。为了帮助开发人员快速解决算子精度问题，需要提供比对两者之间差距的工具。

精度比对工具提供自有模型算子的运算结果与Caffe、TensorFlow、ONNX标准算子的运算结果进行比对的功能，以便确认存在误差的算子。目前提供余弦相似度、最大绝对误差、累积相对误差、欧氏相对距离、KLD散度以及标准差的算法比对项目。

Profiling 性能分析

Profiling实现了Host+Device侧丰富的性能数据采集能力和全景Timeline交互分析能力，展示Host+Device侧各项性能指标，帮助用户快速发现和定位AI应用的性能瓶颈。包括资源瓶颈导致的AI算法短板，指导算法性能提升和系统资源利用率的优化。Profiling支持Host+Device侧的资源利用可视化统计分析，具体包括Host侧CPU、Memory、Disk、Network利用率和Device侧APP工程的硬件和软件性能数据。

AI Core Error 分析工具

在执行训练和推理发现AI Core Error问题时，使用AI Core Error Analyzer工具可以自动快速准确地收集定位AI Core Error问题所需的关键信息，提升开发者对AI Core Error的排查效率。

AutoML 工具（Beta）

AutoML（Auto Machine Learning）包括模型自动性能调优、模型自动生成和训练工程超参数自动调优，HPO（Hyper parameter optimization）和大模型微调四个模块。昇腾模型开发用户可以通过模型自动性能调优功能找到性能更好的模型。AI初学者可以通过AutoML工具结合数据集，自动生成满足需求的模型，也可以对训练超参进行自动调优。用户还可以通过大模型微调对紫东太初模型五个下游任务进行微调。

专家系统工具

专家系统（MindStudio Advisor）是用于聚焦模型和算子的性能调优TOP问题，识别性能瓶颈Pattern，重点构建瓶颈分析、优化推荐模型，支撑开发效率提升的工具。当前提供的功能有：基于Roofline模型的算子瓶颈识别与优化建议、算子优化分析、基于Timeline的AICPU算子优化和算子融合推荐。

MindX SDK 可视化流程编排

MindX SDK实现功能的最小粒度是插件，每一个插件实现特定的功能，如图片解码、图片缩放等。将这些插件按照合理的顺序编排，实现负责的功能。可视化流程编排是以可视化的方式，开发数据流图，生成pipeline文件供应用框架使用。

UT 测试

MindStudio提供了基于gtest框架的新的UT测试方案，简化了开发者开发UT测试用例的复杂度。

UT（Unit Test）即单元测试，是开发人员进行单算子运行验证的手段之一，主要目的是：

- 测试算子代码的正确性，验证输入输出结果与设计的一致性。
- UT侧重于保证算子程序能够跑通，选取的场景组合应能覆盖算子代码的所有分支（一般来说覆盖率要达到100%），从而降低不同场景下算子代码的编译失败率。

ST 测试

MindStudio提供了新的ST（System Test）测试框架，可以自动生成测试用例，在真实的硬件环境中，验证算子功能的正确性和计算结果准确性，并生成运行测试报告，包括：

- 基于算子信息库生成算子测试用例定义文件。
- 基于算子测试用例定义文件生成不同shape、dtype的测试数据和基于AscendCL的测试用例。
- 编译算子工程并将算子部署到算子库，最后在硬件环境中执行测试用例，验证算子运行的正确性。
- 自动生成运行报表（st_report.json）功能，报表记录了测试用例信息及各阶段运行情况。
- 根据用户定义并配置的算子期望数据生成函数，回显期望算子输出和实际算子输出的对比测试结果，验证计算结果的准确性。

版权所有 © 华为技术有限公司

2 版本配套关系

表 2-1 版本配套关系

MindStudio版本	CANN商用版	发布时间
5.0.RC1	昇腾社区版本：**5.1.RC1** 对应企业业务网站版本：**5.1.RC1**	2022.4
3.0.4	昇腾社区版本：**5.0.4** 对应企业业务网站版本：**5.0.4**	2022.1
3.0.3.6	昇腾社区版本：**5.0.3.6** 对应企业业务网站版本：**5.0.3.6**	2022.1
3.0.3	昇腾社区版本：**5.0.3** 对应企业业务网站版本：**5.0.3**	2021.10
3.0.2	昇腾社区版本：**5.0.2** 对应企业业务网站版本：**5.0.2**	2021.07
3.0.1	昇腾社区版本：**3.3.0** 对应企业业务网站版本：**5.0.1**	2021.04
2.0.0(release)	昇腾社区版本：**3.2.1** 对应企业业务网站版本：**V100R020C30**	2021.04

兼容性说明

旧版本AICPU算子工程在使用新版本的Ascend-cann-toolkit时，存在兼容性问题，具体解决方法请参见《**MindStudio用户指南**》“自定义算子开发 > AI CPU算子开发 > 算子工程编译”章节。

版权所有 © 华为技术有限公司

3 新增特性

安装管理

- MindStudio内置JBR以简化安装操作。
- 支持通过配置文件指定MindStudio的运行时环境。

基础开发

- CMakeLists文件可使用快捷键注释语句。

脚本转换

- X2MindSpore工具支持TensorFlow 2.x训练网络一键式整网迁移至MindSpore框架训练网络。
- X2MindSpore工具支持分布式迁移能力。
- X2MindSpore工具支持转换后的代码可以在ModelArts上运行。
- PyTorchGPU2NPU工具支持PyTorch1.8训练脚本的迁移。

模型转换

- 模型可视化功能支持Caffe模型的可视化。
- 新增离线模型改图调优工具。

应用开发

可视化编排功能优化。

- pipeline插件元件框可通过拉伸调整大小。
- 支持在设置时显示原名称。

算子开发

- 算子工程创建界面优化
  - 算子模板（Sample Template）提供算子分类，并新增配置提示。
  - 新增对IR Template及type等选项的配置提示。
- 算子ST测试

版权所有 © 华为技术有限公司

  - 运行配置界面的Enable System Profiler开关下新增配置选项。
  - 运行配置界面增加Toolchain配置项，提供系统预置Toolchain，并支持用户自定义Toolchain。
  - Windows远程模式支持运行Profiling。
- 算子工程编译
  - 编译配置界面增加Toolchain配置项，提供系统预置Toolchain，并支持用户自定义Toolchain。
- 算子UT测试
  - 支持从并行度视图跳转到TBE、CCE代码。

调试调优

- Profiling性能分析
  - Profiling性能分析“Profiling Options”配置界面增加MsprofTX参数。
  - Profiling性能分析支持Deployment。
  - Profiling性能分析支持HCCL采集。
  - Profiling性能分析支持训练任务采集。
- 精度比对
  - 精度比对工具重做，全新的界面和配置方式。
- 专家系统工具
  - 专家系统（专家系统入口）Roofline页面（基于Roofline模型的算子瓶颈识别与优化建议能输出结果）新增提供算子占比统计。
  - 专家系统（专家系统入口）支持Deployment。
  - 专家系统（专家系统入口）支持Windows。
  - 专家系统（算子工程入口）优化建议支持代码提示和跳转。
  - 专家系统（专家系统入口）SOC Version参数。
- Beta特性AutoML工具
  - AutoML支持模型量化调优。
  - AutoML支持紫东太初大模型微调工具。
  - AutoML新建可视化任务支持查看历史任务。

版权所有 © 华为技术有限公司

4 Windows 特性支持清单

当前在Windows系统上使用MindStudio，支持的特性清单如表4-1所示。

表 4-1 特性清单

特性	子特性
安装管理	MindStudio压缩包解压安装
	MindStudio安装包安装
	在线升级
	CANN版本管理
	CANN版本切换
	设备管理
	SSH配置管理
分析迁移	PyTorch GPU2Ascend训练脚本转换
	TensorFlow GPU2Ascend训练脚本转换
	X2MindSpore脚本转换工具
算子开发	创建算子工程
	TBE自定义算子开发UT验证
	算子开发ST测试用例自动生成与验证
	算子开发model仿真验证
	算子开发TMModel性能仿真
	算子远程编译与部署
	支持TIK debug可视化
模型开发	模型转换
	模型可视化

版权所有 © 华为技术有限公司

特性	子特性
	模型转换支持动态AIPP、动态ImageSize、动态Batch
	模型转换静态AIPP支持padding
	自定义输出节点标记
	模型评估报告（推理场景）
	历史配置文件加载
	离线模型改图调优工具
	算子模型速查工具
训练工程	模型训练执行
应用工程	ACL应用工程本地/远程编译运行
	ACL应用工程远程调试
	本地/远程运行生成dump文件
	MindX SDK安装与版本管理
	MindX应用编译调试运行
	MindX Pipeline流程编排
调试调优	精度比对
	Profiling性能分析
	AI Core Error分析工具
	专家系统工具
	AutoML工具

版权所有 © 华为技术有限公司