



MindSpore

端侧AI引擎 MindSpore Lite 初体验

作者：邢朝龙

目录

contents

01

端侧AI开发工作流

02

端侧AI推理框架

03

MindSpore Lite初体验

04

谢谢观看



MindSpore

端侧AI开发工作流

机器学习工作流：数据处理→模型构建→模型训练→评估测试

- 数据处理
- 模型构建
- 模型训练
- 评估测试

端侧开发工作流：模型导出→模型转换（含模型量化）→模型测试→引擎开发→引擎测试

- 模型导出
- 模型转换（含模型量化）
- 模型测试（一致性）
 - 原始
 - 量化
- 引擎开发
 - 预处理
 - 模型推理
 - 后处理
- 引擎测试
 - 一致性
 - 稳定性
 - 性能



MindSpore

端侧AI推理框架

- Tensorflow Lite
- PyTorch Mobile
- Paddle Lite
- MindSpore Lite
- NCNN
- MACE
- MNN
- TNN

训练框架支持、算子支持、设备（架构）支持、第三方模型支持、端侧训练支持、模型量化算法、性能优化



MindSpore

MindSpore Lite初体验

- 项目编译
- 框架下载
- 模型转换
- 模型推理
- 模型可视化



MindSpore

项目编译

- 环境配置
 - GCC && Git
 - Cmake
 - Java
 - Android
 - sdk
 - ndk
 - Gradle
 - Python
- 编译源码



MindSpore

框架下载

```
.
├── runtime
│   ├── include
│   ├── lib
│   └── third_party
└── tools
    ├── benchmark
    ├── benchmark_train
    ├── codegen
    ├── converter
    └── cropper

10 directories, 0 files
```



模型转换

- 模型导出
- 模型转换
- 模型量化



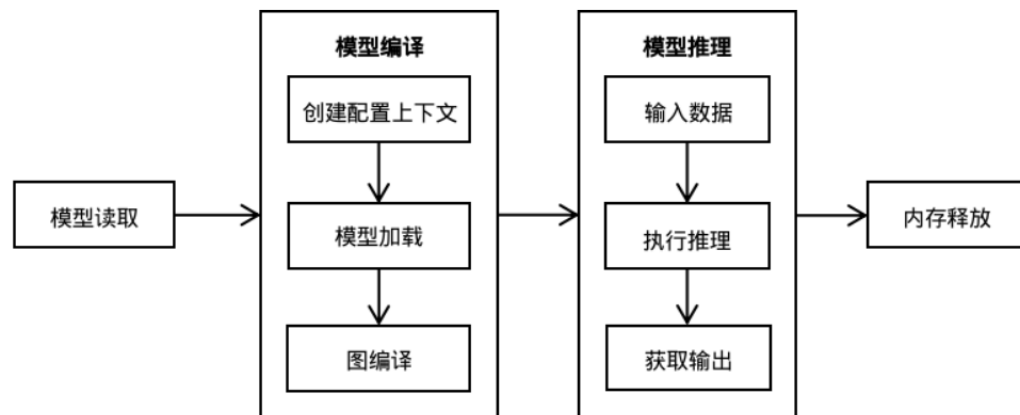
模型推理

- C++

- 推理步骤
- 推理代码
- 编译脚本
- 项目编译

- Python

- 环境配置
- 推理代码



模型可视化

Netron是一个基于Electron平台开发的神经网络模型可视化工具，支持许多主流AI框架模型的可视化，支持多种平台（Mac、Windows、Linux等）。

Netron支持MindSpore Lite模型，可以方便地查看模型信息。如下图所示，使用Netron加载.ms模型后，可以展示模型的拓扑结构和图、节点的信息等。

THANK YOU