MindSpore and ONNX

Zhipeng Huang
Contents

01 MindSpore Introduction

02 MindSpore and ONNX/MLIR

03 MindSpore and ONNX Training

04 Call For Participation
MindSpore Introduction

MindSpore is a new open source deep learning training/inference framework that could be used for mobile, edge and cloud scenarios. MindSpore is designed to provide development experience with friendly design and efficient execution for the data scientists and algorithmic engineers, native support for Ascend AI processor, and software hardware co-optimization. At the meantime MindSpore as a global AI open source community, aims to further advance the development and enrichment of the AI software/hardware application ecosystem.

Overview

MindSpore Deep Learning Framework

<table>
<thead>
<tr>
<th>MindSpore FrontEnd Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Python API</td>
</tr>
<tr>
<td>Training/Inference/Export</td>
</tr>
<tr>
<td>Data Processing</td>
</tr>
<tr>
<td>Data Format Transformation</td>
</tr>
<tr>
<td>MindSpore IR</td>
</tr>
<tr>
<td>GHLO</td>
</tr>
<tr>
<td>High Level Optimization</td>
</tr>
<tr>
<td>Auto Parallel</td>
</tr>
<tr>
<td>Auto Differentiation</td>
</tr>
</tbody>
</table>

MindSpore Graph Engine (Ascend/GPU/CPU Support)

| GLLO                          |
| Low Level Optimization        |
| Pipeline Parallel             |
| Grpah Execution               |
| On-Device Execution           |
| Distributed Libs (Comms/PS)   |

MindSpore Backend Runtime (Cloud/Edge/Mobile)

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
</tr>
<tr>
<td>GPU</td>
</tr>
<tr>
<td>Ascend 310</td>
</tr>
<tr>
<td>Ascend 910</td>
</tr>
<tr>
<td>Android/iOS</td>
</tr>
</tbody>
</table>
Technical Steering Committee

14 members from various universities/institutions/companies that forms an open and global technical governing body
SIGs
- FrontEnd
- Compiler
- Executor
- ModelZoo
- Data
- GraphEngine
- Visualization
- Security

WG
- Documentation
- Infrastructure

Open Developement
Community partners

Open Collaboration
Contents

01  MindSpore Introduction

02  MindSpore and ONNX/MLIR

03  MindSpore and ONNX Training

04  Call For Participation
MindSpore ONNX Exporter Introduction

1. Use MindSpore model train API to perform model training with saving checkpoint parameters
2. Load model parameters into the network to be exported (such like LeNet)
3. Call `train.export()` to convert MindSpore model to ONNX model
4. Perform model inference on ONNX Runtime
## MindSpore ONNX Exporter Support

<table>
<thead>
<tr>
<th>MindSpore Operator</th>
<th>ONNX Operator</th>
<th>MindSpore Operator</th>
<th>ONNX Operator</th>
</tr>
</thead>
<tbody>
<tr>
<td>TensorAdd</td>
<td>Add</td>
<td>BatchNorm</td>
<td>BatchNormalization</td>
</tr>
<tr>
<td>Mul</td>
<td>Mul</td>
<td>Reshape</td>
<td>Reshape</td>
</tr>
<tr>
<td>ReLU</td>
<td>Relu</td>
<td>ReduceMean</td>
<td>ReduceMean</td>
</tr>
<tr>
<td>Sigmoid</td>
<td>Sigmoid</td>
<td>Cast</td>
<td>Cast</td>
</tr>
<tr>
<td>Flatten</td>
<td>Flatten</td>
<td>PReLU</td>
<td>PReLU</td>
</tr>
<tr>
<td>Squeeze</td>
<td>Squeeze</td>
<td>Argmax</td>
<td>ArgMax</td>
</tr>
<tr>
<td>Conv2D</td>
<td>Conv</td>
<td>SimpleMean</td>
<td>AveragePool</td>
</tr>
<tr>
<td>BiasAdd</td>
<td>Add</td>
<td>MaxPool</td>
<td>MaxPool</td>
</tr>
<tr>
<td>MatMul</td>
<td>Gemm</td>
<td>AvgPool</td>
<td>AveragePool</td>
</tr>
</tbody>
</table>

- **Operator Support**: 18
- **Network Support**: LeNet, ResNet50, AlexNet
- **Roadmap**: More operators and networks (in ModelZoo) are WIP
Exploring MindSpore and ONNX-MLIR (Idea)

ONNX Frontend → ONNX Dialect → MLIR Dialect → LLVM Dialect

(Standard/Transformation/Affine Dialect etc.)

https://github.com/onnx/onnx-mlir

MindSpore

FrontEnd Expression
train.export()

ONNX Exporter

model.onnx

LLVM+
ONNX-MLIR

save as ONNX model
Contents

01
MindSpore Introduction

02
MindSpore and ONNX/MLIR

03
MindSpore and ONNX Training

04
Call For Participation
Exploring MindSpore and ONNX Training (Idea)
Contents

01. MindSpore Introduction

02. MindSpore and ONNX/MLIR

03. MindSpore and ONNX Training

04. Call For Participation
Call For Participation

• Checkout the Code
  - https://gitee.com/mindspore (main development)
  - https://github.com/mindspore-ai (mirror)

• Try with docker
  - docker pull mindspore/mindspore-cpu:0.1.0-alpha
  - docker pull mindspore/mindspore-cuda10.1:0.1.0-alpha
  - docker pull mindspore/mindspore-cuda9.2:0.1.0-alpha

• Discussion (you can initiate an ONNX WG !)
  - Slack: https://join.slack.com/t/mindspore/shared_invite/enQtOTcwMTIxMDI3NjM0LTNkMWM2MzI5NjlyZWU5ZWQ5M2EwMTQ5MWNiYzMxOGM4OWFhZjI4M2E5OGI2YTE3ODU1ODE2Ng1MTThiNWl3YmQ
  - Mailing list: https://mailweb.mindspore.cn/postorius/lists/mindspore-discuss.mindspore.cn/