

DRP-AI Translator i8 V1.10

Release Note

Introduction

This release note describes the improvements of the DRP-AI Translator i8.

Key Features and Enhancements

- Update OS version to 22.04
- Update Python version to 3.10
- Support Convolution with asymmetric kernel
- Add YoloX-Pose model to Getting Started guide, and include sample onnx model

Contents

1. Improvement.....	2
1.1 Operator / Attribute updates	2
2. Fixed Issues	3
2.1 Operator: Convolution, MatMul, and Concat.....	3
2.2 Sparse option (--sparse).....	3
3. Known Issues	4
3.1 Error pattern in specific combinations of heigh/width/input channel/output channel	4
3.2 Softmax attribute	4
4. Getting Started Guide	5

1. Improvement

1.1 Operator / Attribute updates

- **Convolution**

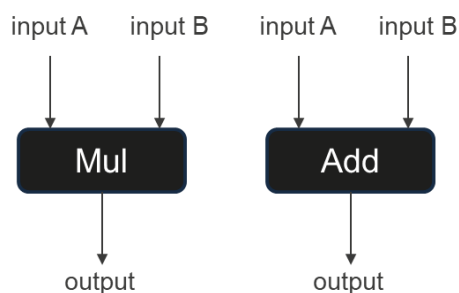
- Newly support ker1x3, stride 1, pad [l,r,t,b] = [1,1,0,0], group = 1, dilation = 1
- Newly support ker3x1, stride 1, pad [l,r,t,b] = [0,0,1,1], group = 1, dilation = 1
- Newly support ker1x7, stride 1, pad [l,r,t,b] = [3,3,0,0], group = 1, dilation = 1
- Newly support ker7x1, stride 1, pad [l,r,t,b] = [0,0,3,3], group = 1, dilation = 1

- **Mul**

- input A : Feature map (1,ch,H,W), input B : Feature map (1,ch,H,W)
 - input A : Feature map (1,ch,H,W), input B : Feature map (1,ch,1,1)
 - input A : Feature map (1,ch,H,W), input B Initializer(parameter) (1,ch,1,1)
 - input A : Feature map (1,ch,H,W), input B : Initializer(parameter) (1)
- * Inputs A and B also support reverse conditions.

- **Add**

- input A : Feature map (1,ch,H,W), input B : Feature map (1,ch,H,W)
 - input A : Feature map (1,ch,H,W), input B : Feature map (1,ch,1,1)
 - input A : Feature map (1,ch,H,W), input B Initializer(parameter) (1,ch,1,1)
 - input A : Feature map (1,ch,H,W), input B : Initializer(parameter) (1)
- * Inputs A and B also support reverse conditions.



- **Multi-person pose estimation model**

- Add new translation guide about multi-person pose estimation model. Please see [GettingStarted/how-to/pose_estimation/mmpose_yolox-pose](#) folder in [DRP-AI Translator i8](#) for the detail.



2. Fixed Issues

2.1 Operator: Convolution, MatMul, and Concat

Fixed the issues where DRP-AI object file was not generated correctly when certain combinations of heigh/width/input channel/output channel are used.

2.2 Sparse option (--sparse)

Fixed an issue where enabling the --sparse option caused errors in certain shape of Gemm/MatMul.

3. Known Issues

3.1 Error pattern in specific combinations of heigh/width/input channel/output channel

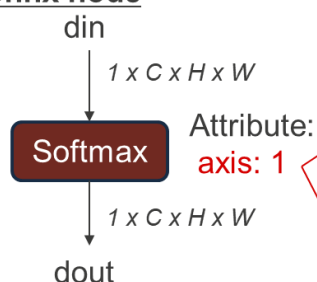
If the following conditions 1, 2, and 3 are true, there may be an error in the inference results.

1. operator: *Convolution* or *MaxPool* or *AveragePool*
2. $(Ker \% 2 == 0)$ or $(Ker \% 2 != 0 \ \& \ pad \neq (ker - 1) / 2)$
3. Feature map size is large
e.g. $ih = iw = 80, ich = 512, och = 512$

3.2 Softmax attribute

Due to a change in the interpretation of axis attribute, old opset softmax operation is not supported. There are error in output value.

onnx node



Supported Operation

```
# din shape is 1,c,h,w
tmp_din = din.transpose((0,2,3,1)) # transpose to 1,h,w,c
for _h in range(h):
    for _w in range(w):
        tmp_din[0][_h][_w] = softmax(tmp_din[0][_h][_w])
dout = tmp_din.transpose((0,1,2,3)) # transpose to 1,c,h,w
```

Not Supported Operation

```
# din shape is 1,c,h,w
tmp_din = din.reshape((1,c*h*w)) # reshape to 1,h*w*c
tmp_din[0] = softmax(tmp_din[0])
dout = tmp_din.reshape((1,c,h,w)) # reshape to 1,c,h,w
```

4. Getting Started Guide

After installing DRP-AI Translator i8, sample pruned/dense onnx models and the Getting Started guide are extracted along with the INT8 Quantizer & Translator. Getting Started helps you learn how to use DRP-AI Translator i8. If you use Translator i8 for the first time, please refer to Getting_Started/README.md. Below is a directory structure.

DRP-AI Translator i8(install directory)

```

├── Getting_Started ... Guide for DRP-AI Translator i8
│   ├── README.md ... Overview of Getting Started
├── onnx_models ... Sample pruned onnx models
├── drpAI_Quantizer ... Root directory of INT8 Quantizer
└── translator ... Root directory of Translator

```

The Getting Started guide describes how to translate the following AI models.

Category	AI model
Object Detection	Lightnet YOLOv2
	Megvii-BaseDetection YOLOX
Semantic Segmentation	torchvision DeepLabv3
Classification	torchvision ResNet50
Human Pose Estimation	MMPose HRNet (Single)
	MMPose YoloX-Pose (Multi)
Depth Estimation	PyTorch Hub MiDas(*1)

*1: Sample pruning model is not included in DRP-AI Translator i8. Please follow the guide to download the model.



Object Detection



Pose Estimation



Sematic Segmentation

