

DRP-AI Translator V1.60

Release Note

Introduction

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

Key Features and Enhancements

- Simplification of pre and postprocessing definition format
- New supported preprocessing type: Crop

Contents

1.	Improvements	2
1.1	Simplification of Pre and Postprocessing Definition Format	2
1.2	Preprocessing Update	2
1.3	New Supported Operators	2
1.4	Operator Attribute Update	3
1.5	Data Size Restriction Update	3
1.6	New Supported Graph Structure	3
1.7	DRP Library Update	4
2.	Fixed Issues.....	4
2.1	Bug Fix	4
3.	Known Issues	4
3.1	Graph Structures that Require -PrePostFULL Option to be Specified.....	4
3.2	Graph Structure that Causes Error during Inference on the device.....	5

1. Improvements

1.1 Simplification of Pre and Postprocessing Definition Format

The format of the pre and postprocessing definition file has been simplified. Compared to the previous format, the amount of description required is reduced by half. Please refer to Section 5 of User's Manual for more details.

Pre and post definition file sample
for DRP-AI Translator V1.5

```
...
#####
# Postprocess
#####
postprocess:
-
  src: ["resnetv17_dense0_fwd"]
  shape_in : [[1000]]
  dtype_in : ["fp16"]
  dorder_in: ["C"]
  dest: ["post_out"]
  shape_out: [[1000]]
  dtype_out: ["fp32"]
  dorder_out: ["C"]
  operations:
  -
    op : softmax
    shape_in : [[1000]]
    dtype_in : ["fp16"]
    dorder_in: ["C"]
    shape_out: [[1000]]
    dtype_out: ["fp32"]
    dorder_out: ["C"]
    param:
      DOUT_FORMAT: 1 # FP32
```

Pre and post definition file sample
for DRP-AI Translator V1.60

```
...
#####
# Postprocess
#####
postprocess:
-
  src: ["resnetv17_dense0_fwd"]
  dest: ["post_out"]
  operations:
  -
    op : softmax
    param:
      DOUT_FORMAT: 1 # FP32
```



1.2 Preprocessing Update

- crop

Crop is now supported as a preprocessing type. Users can now crop a specified area of the input image.

- conv_yuv2rgb

New supported input formats have been added. The following formats can now be converted to RGB by conv_yuv2rgb preprocessing.

Y₀UY₁V(YUY₂), Y₀VY₁U, UY₀VY₁, VUY₁Y₀

1.3 New Supported Operators

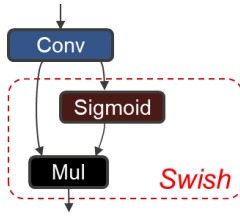
The following operators are now supported.

- SeLU
- Softmax

Note : Softmax is supported only when used as the last operator in the onnx model, and its shape is 1-dimension only.

- Swish

Swish activation function is represented in ONNX format by the following graph structure



1.4 Operator Attribute Update

New kernel and padding shapes are now supported for the operators below. The improvements from previous version are highlighted in red. Please refer to Section 4 of User's Manual for more details.

- Depthwise Convolution kernel: [3,3], stride: 1, pad [1,1,1,1], pad [0,0,0,0], pad [1,0,1,0], pad [0,1,0,1]
- Depthwise Convolution kernel: [3,3], stride: 2, pad [1,1,1,1], pad [0,0,0,0], pad [1,0,1,0], pad [0,1,0,1]
- MaxPool, kernel: [2,2], stride :1, pad [1,1,1,1], pad [0,0,0,0], pad [1,0,1,0], pad [0,1,0,1]
- MaxPool, kernel: [2,2], stride :2, pad [1,1,1,1], pad [0,0,0,0], pad [1,0,1,0], pad [0,1,0,1]
- MaxPool, kernel: [3,3], stride :2, pad [1,1,1,1], pad [0,0,0,0], pad [1,0,1,0], pad [0,1,0,1]
- MaxPool, kernel: [5,5], stride :1, pad [2,2,2,2], pad [0,0,0,0], pad [2,0,2,0], pad [0,2,0,2]

1.5 Data Size Restriction Update

The following pre/postprocessing data size constraints have been removed.

- Memcopy : minimum size constraint related to input data shape is now removed
- cast_fp16_fp32 : minimum size constraint related to input data shape is now removed
- resize_hwc : maximum size constraint related to input data shape is now removed

The following operator data size constraints have been improved.

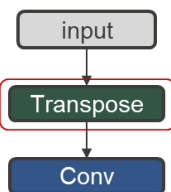
- resize_onnx: updated constraint related to minimum input data shape.

1.6 New Supported Graph Structure

The graph structure analysis function has been updated to support the following structures.

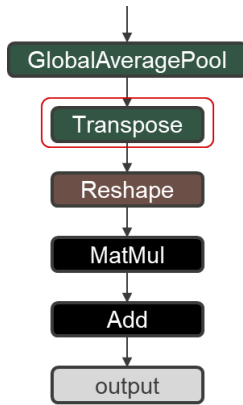
- Input > Transpose

Below is a graph structure that may be found in ONNX generated by Keras. Transpose node will be skipped and DRP-AI will process the input data as HWC format.



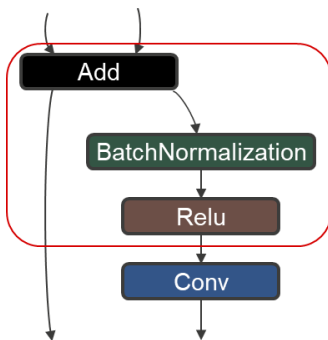
- Transpose > Reshape > MatMul

Below is a graph structure that may be found in ONNX generated by Keras. Transpose node will be skipped and DRP-AI will process the input data as HWC format.



- Add > BatchNormalization > Relu

The following graph structure which can be found in various models (e.g. ResNetV2) is now supported.



1.7 DRP Library Update

Updated DRP libraries and function. The supported attribute or inference time are also improved.

2. Fixed Issues

2.1 Bug Fix

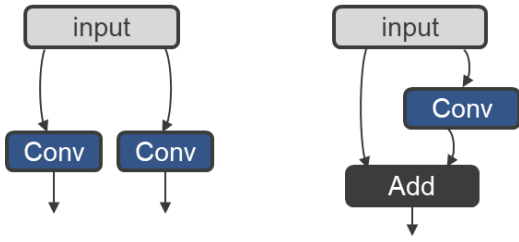
- Fixed the issue where DRP-AI object file was not generated correctly when certain combinations of heigh/width/input channel/output channel are used.
- Fixed the issue related to Softmax calculation when the input data size is not a multiple of 4. In the older version, the output values of softmax contained an extremely small error.

3. Known Issues

3.1 Graph Structures that Require -PrePostFULL Option to be Specified

When using ONNX with the graph structure described below, please specify the -PrePostFULL option of DRP-AI Translator. In this case, all parameters must be described in the pre/post processing definition file. Refer to the older format of the pre/post processing definition file described in the Appendix A of User's Manual.

- Graph structure where input node is directly connected to two or more operators



3.2 Graph Structure that Causes Error during Inference on the device

The onnx model with the following graph structure cannot be translated correctly. Inference on the device fails. If expressing Batch Normalization by a combination of Mul and Add operator, models with Mul constants containing “0” cannot be supported. In that issue case, the following Warning will be output.

```
drpai_common/numpy_util.py:82: RuntimeWarning: divide by zero encountered in true_divide
mean = minus_bias / vec1 if otype1 == 'Mul' else minus_bias * vec1
```

