Amlogic

Perlayer Visual Tool Introduction

Revision: 0.1

Release Date: 2022-02-10

Contact Information

- Website: www.amlogic.com
- Pre-sales consultation: contact@amlogic.com
- Technical support: support@amlogic.com

Revision History

Issue 0.1 (2022-02-10)

This is the Initial release.

Contents

1. Introduction

Perlayer visual is a visualization tool for each layer of the model based on hardware information printing developed by Amlogic. It is convenient for users to evaluate the running time and bandwidth infomations of the model, as well as find the Operators that take a long time and take up too much bandwidth.

Tool link:

https://github.com/Amlogic-NN/AML_NN_SDK/tree/master/Tools/get_layer_info

2. Perlayer visual usage

2.1 Get board inference results

1. A. Import share library files that support export.data model operation

Find the share library file corresponding to the specified platform under the Github path. Note that the DDK version and platform information must be consistent. It mainly includes four platforms: A311D_0x88, C305X_0xbe, C308X_0xa1, and S905D3_0x99. Common stores common share library files for each platform.

https://github.com/Amlogic-NN/AML_NN_SDK/tree/master/Linux/runtime_sharelib

Push the corresponding share library files to the board, such as /data/test/drivers, and set the environment variable:

export LD_LIBRARY_PATH=/data/test/drivers

2. B. Set environment variables before running the export.data model on the board

export VIV_VX_PROFILE=1

export VIV_VX_DEBUG_LEVEL=1

export CNN_PERF=1

export VIV_VX_SAVE_PATH=inceptionv1.txt #Set the path to save printing information

3. C. Run export.data case

./inceptionv1 inceptionv1.export.data input.tensor

Each layer of information printing will be saved in the previously set path inceptionv1.txt file.

2.2 Python visual script

1. The visualization script and inceptionv1.txt saved during the runtime of the board in the python3 environment.
2. Execute the script

python3 show_layer_info.py -i inceptionv1.txt -c 0xa1

Where -i is the input text file path, -c is the corresponding platform number

A311D --- 0x88

S905D3 --- 0x99

C305X --- 0xbe

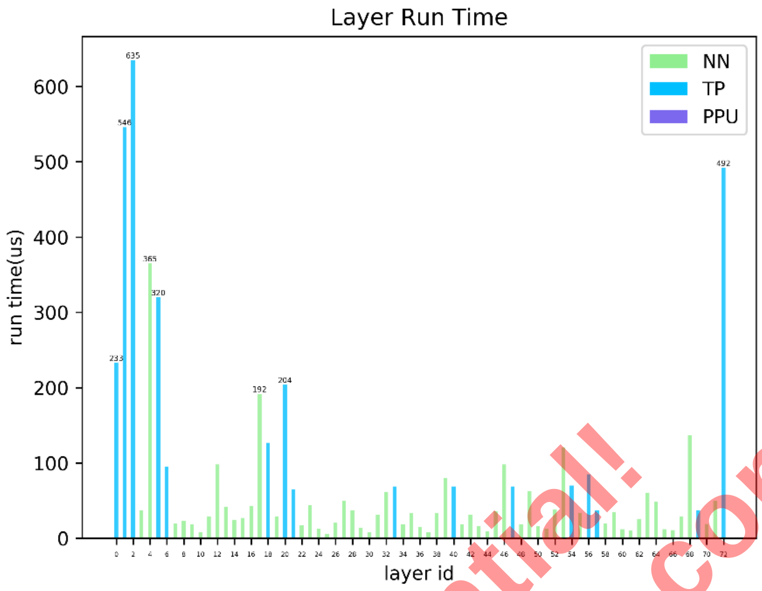
C308X --- 0xa1

3. Save the picture and the corresponding result file

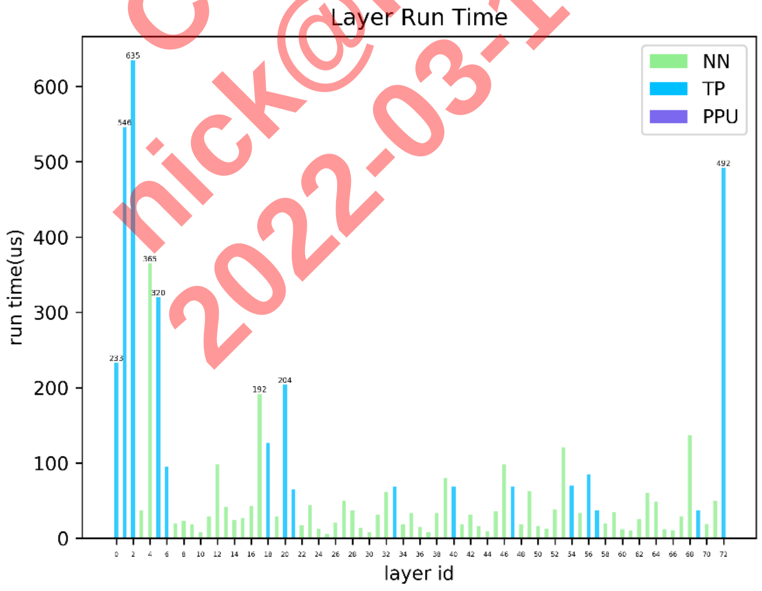
After executing the script, save DDR_BandWidth.png, Runtime.png describes the bandwidth and running time respectively, and will generate results.txt to save related information.

4. Examples

Execution results based on the DDK6487+C305X+inceptionv1 model



Runtime.png get each layer of the model to run on the specific hardware of NN/TP/PPU, and get the running time of each layer. DDR_BandWidth.png gets the DDK bandwidth infomations by each layer of the model. You can refer to the above-mentioned index-targeted optimization operator to improve the operating efficiency of the model. The specific operator information corresponding to each level of label can be found in results.txt.


Layer Run Time
NN
TP
PPU
run time(us)
layer id