CNN-IA 2022-06-29

### **Analyzer Report (Memory Capcity)**

This report generated by Convolutional Neural Network Inference Analyzer (CNN-IA) to summarize the analysis needed to reach the optimal memory capacity for mlp\_fc3\_batch16 by varying architecture parameters for the optimal loop blocking in a restricted schedule space.

#### **Memory Architecture:**

	L0	L1	L2	L3	
Capacity	64	256	65536	536870912	
Access cost	0.2	0.8	6.0	200.0	
Static cost	0.0	0.0	0.0	0.0	
Parallel count	1	256	1	1	
Parallel mode	0	1	0	0	
Parallel cost	0.0	2.0	0.0	0.0	

Precision : 16

Minimum utilization : 0.0%

Outputs can be buffered by MAC : 0

Replication to improve utilization : True

### **Glossary:**

- Memory Levels: (L0, L1, L2, L3)
The smallest index the nearest to CPU.

- Loop Notations: (FX, FY, OX, OY, OC, IC, ON)

FX : FILTER WIDTH

FY : FILTER HEIGHT

OX : OUTPUT WIDTH

OY : OUTPUT HEIGHT

OC : OUTPUT CHANNEL

IC : INPUT CHANNEL

ON : BATCH

CNN-IA 2022-06-29

## **Exploration Parameters:**

	L0	L1
Scaling count	5.0	5.0
Size scaling	2.0	2.0
Cost scaling	20	20

# **Exploration Table:**

L0-SIZE	L0-COST	L1-SIZE	L1-COST	L2-SIZE	L2-COST	L3-SIZE	L3-COST	TOTAL
4.0	0.0125	16.0	0.05	65536.0	6.0	536870912.0	200.0	7115750.0
8.0	0.025	16.0	0.05	65536.0	6.0	536870912.0	200.0	7165700.0
16.0	0.05	16.0	0.05	65536.0	6.0	536870912.0	200.0	inf
32.0	0.1	16.0	0.05	65536.0	6.0	536870912.0	200.0	inf
64.0	0.2	16.0	0.05	65536.0	6.0	536870912.0	200.0	inf
4.0	0.0125	32.0	0.1	65536.0	6.0	536870912.0	200.0	4193550.0
8.0	0.025	32.0	0.1	65536.0	6.0	536870912.0	200.0	4065500.0
16.0	0.05	32.0	0.1	65536.0	6.0	536870912.0	200.0	4215400.0
32.0	0.1	32.0	0.1	65536.0	6.0	536870912.0	200.0	9152800.0
64.0	0.2	32.0	0.1	65536.0	6.0	536870912.0	200.0	inf
4.0	0.0125	64.0	0.2	65536.0	6.0	536870912.0	200.0	4026750.0
8.0	0.025	64.0	0.2	65536.0	6.0	536870912.0	200.0	3926700.0
16.0	0.05	64.0	0.2	65536.0	6.0	536870912.0	200.0	4022600.0
32.0	0.1	64.0	0.2	65536.0	6.0	536870912.0	200.0	4168000.0
64.0	0.2	64.0	0.2	65536.0	6.0	536870912.0	200.0	4867600.0
4.0	0.0125	128.0	0.4	65536.0	6.0	536870912.0	200.0	4239550.0
8.0	0.025	128.0	0.4	65536.0	6.0	536870912.0	200.0	3939500.0
16.0	0.05	128.0	0.4	65536.0	6.0	536870912.0	200.0	3827400.0
32.0	0.1	128.0	0.4	65536.0	6.0	536870912.0	200.0	3822400.0
64.0	0.2	128.0	0.4	65536.0	6.0	536870912.0	200.0	4422000.0
4.0	0.0125	256.0	0.8	65536.0	6.0	536870912.0	200.0	5761150.0
8.0	0.025	256.0	0.8	65536.0	6.0	536870912.0	200.0	4837900.0
16.0	0.05	256.0	0.8	65536.0	6.0	536870912.0	200.0	4437000.0
32.0	0.1	256.0	0.8	65536.0	6.0	536870912.0	200.0	4227200.0
64.0	0.2	256.0	0.8	65536.0	6.0	536870912.0	200.0	4626800.0