

# Performance Analysis of Machine Learning Software

Federico Bacci

# Acquire Data

- We have 2 dataframes
  - Test info
  - Test results
- Merging in one dataframe to analyze results

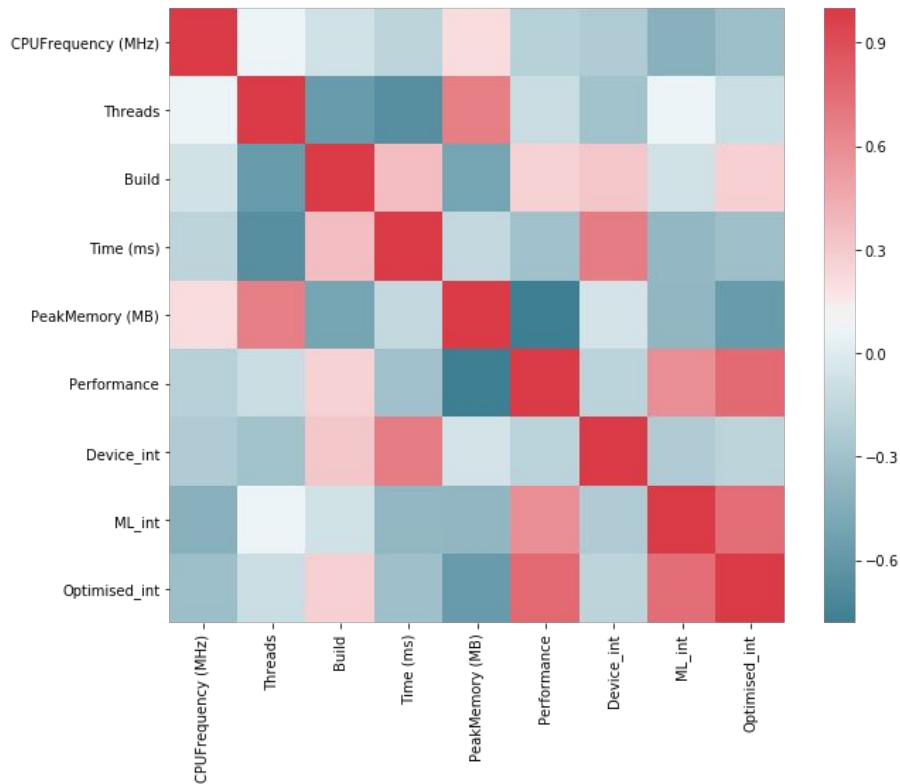
	TestId	Device	CPUFrequency (MHz)	Threads	MLNetwork	Build	Optimised	Time (ms)	PeakMemory (MB)
0	17	Device_0	1000	5	AlexNet	9	N	102.000000	449
1	16	Device_0	1000	5	AlexNet	8	N	104.000000	453
2	39	Device_1	1000	3	AlexNet	10	N	333.333333	302
3	31	Device_0	1000	5	AlexNet	10	N	100.000000	449
4	30	Device_0	1000	4	AlexNet	10	N	125.000000	450

# Correlation

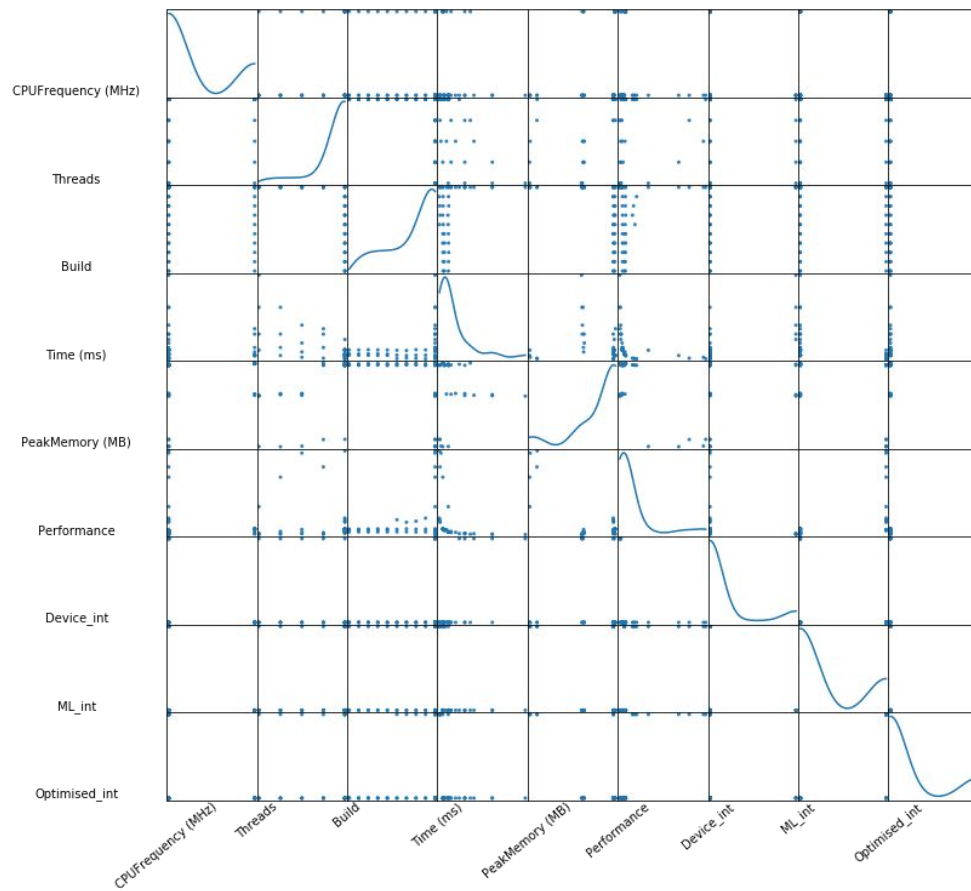
To help with Time and Memory we can create a new column, named **Performance that is higher when both time and Memory are lower**

# Correlation

Correlation  
between columns



# Correlation



# Findings

- 47 Tests
- 2 MLNetworks (AlexNet, MobileNet)
- 2 Devices (Device\_0, Device\_1)
- 2 CPU Frequencies (1000, 2000)
- 2 Optimised Status (Yes or No)
- 5 Use of Threads (from 1 to 5)
- Peak Memory in MB (from 50 to 460)
- Time in ms (from 19 to 800)

# Findings

- **High Correlation**

- Threads and Memory -> More Threads use more memory
- Time and Device -> Device\_0 uses more time
- ML and optimization -> Only MobileNet is tried optimized
- Performance and ML -> MobileNet is performing better
- Performance and Optimization -> Optimizing impact a lot on performance

- **Low Correlation**

- CPUFreq and Memory
- Build and Optimization
- Build and Device
- Build and Time
- Build and Performance

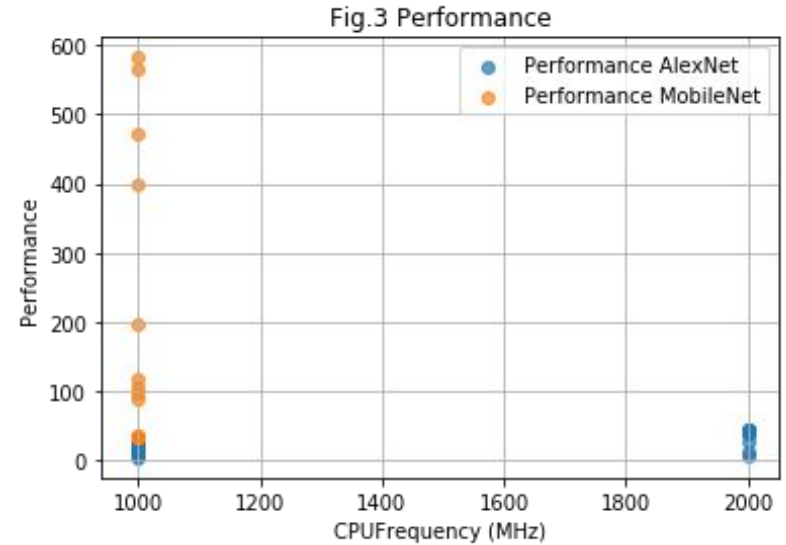
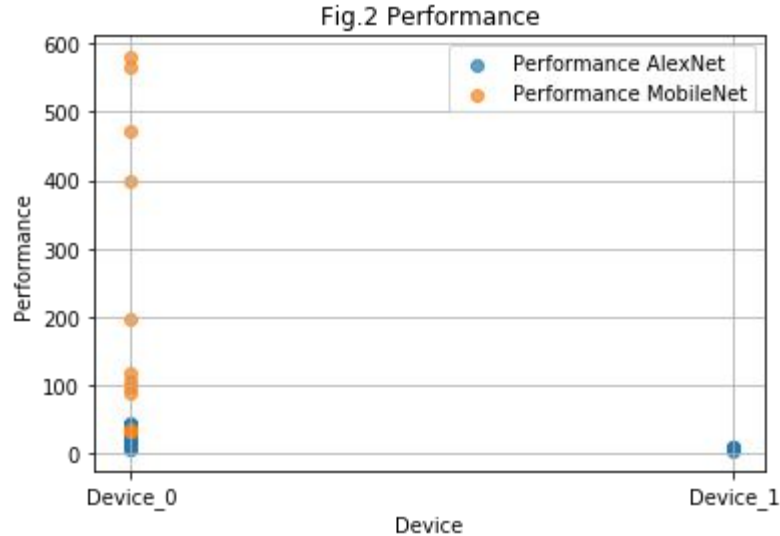
# Findings

## **Performance is maximized if:**

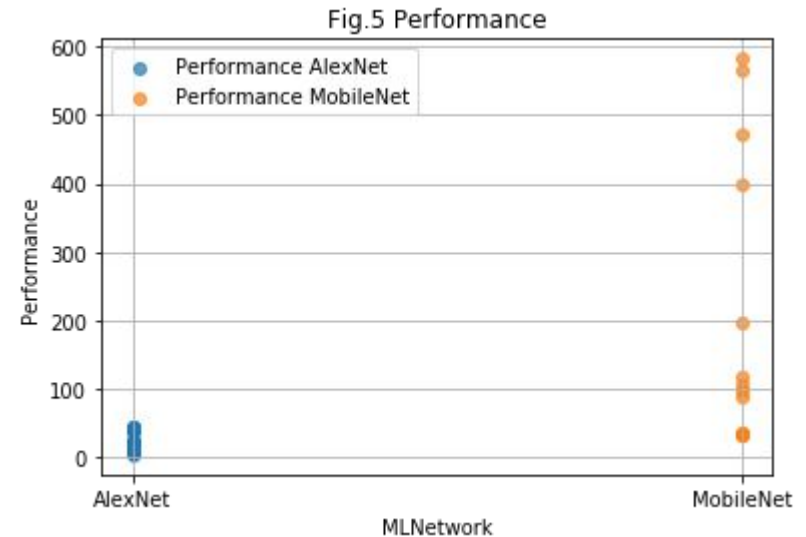
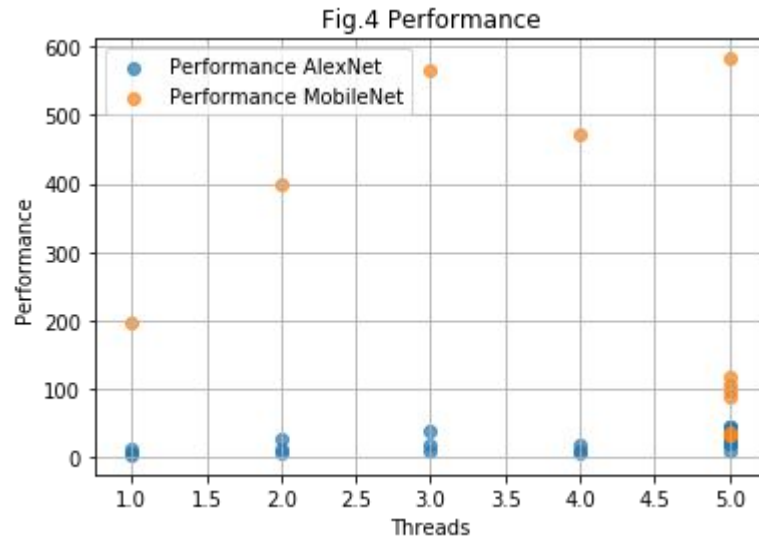
- More Thread are used
- CPU Freq is higher
- Build are higher
- Time and Memory are lower
- Mobile Net is used
- ML is optimized



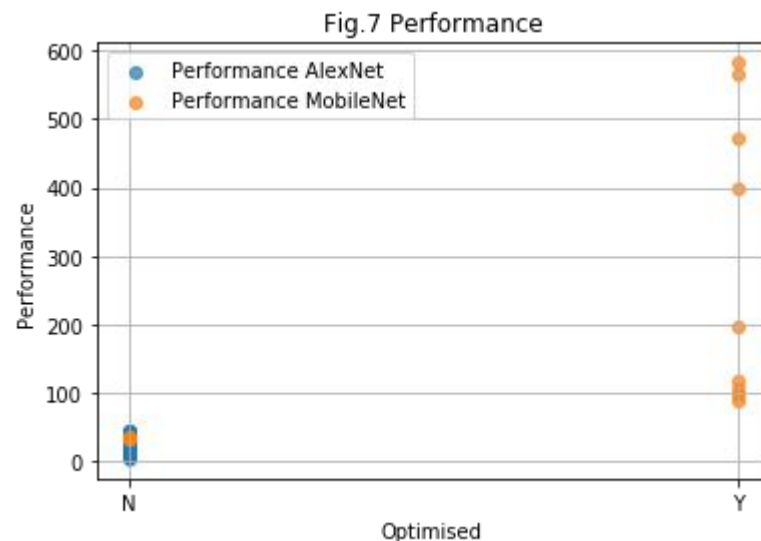
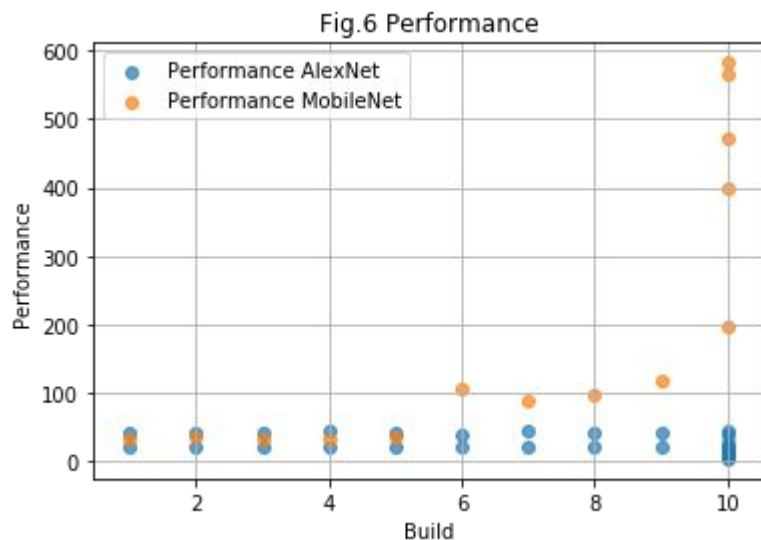
# ML Network Comparison



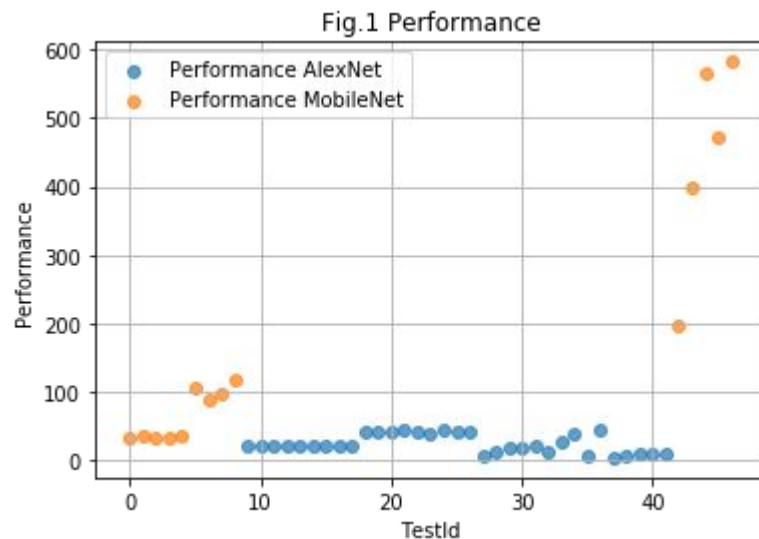
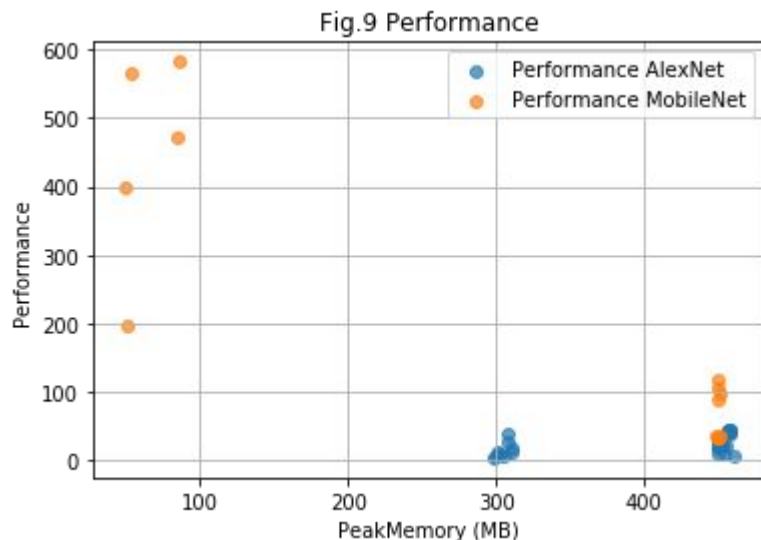
# ML Network Comparison



# ML Network Comparison



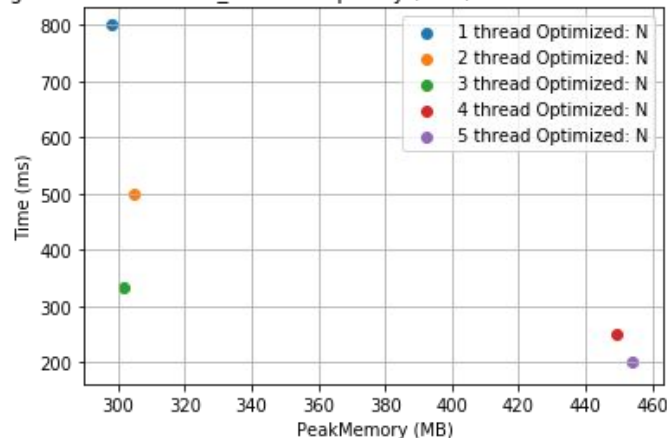
# ML Network Comparison



# AlexNet

- AlexNet is not optimized
- More Threads uses more memory but less time
- More CPU Frequency means less time
- To minimize time and space we can check which value is closer to 0 in figure 13,14,15 Overall the best option performance wise is to use 5 threads, if memory is a constraint we can lowe the number of threads

Fig.10 Device: Device\_1 CPU Frequency (MHz): 1000 MLNetwork: AlexNet



# AlexNet

Fig.11 Device: Device\_0 CPU Frequency (MHz): 1000 MLNetwork: AlexNet

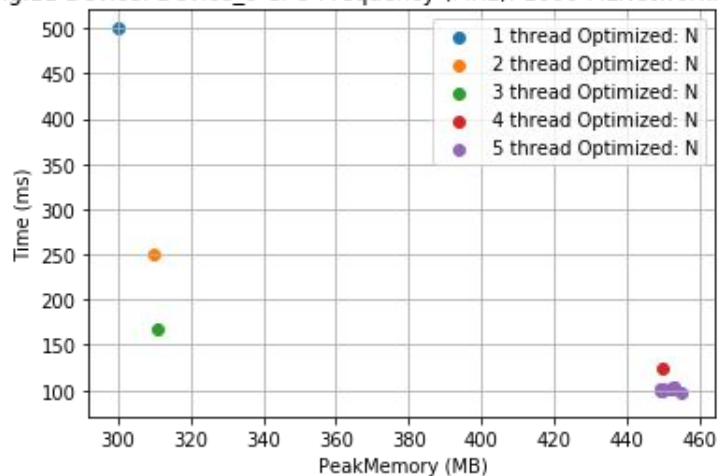
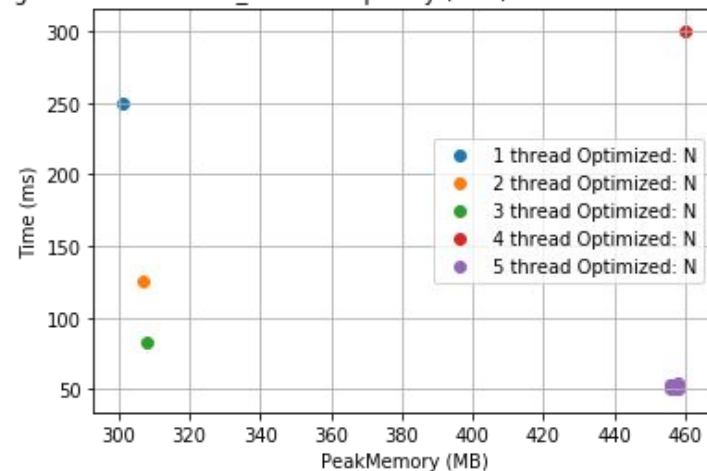


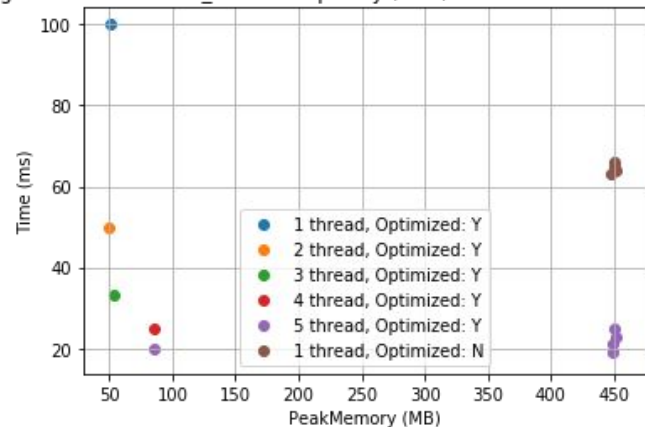
Fig.12 Device: Device\_0 CPU Frequency (MHz): 2000 MLNetwork: AlexNet



# MobileNet

- MobileNet is tested only on Device\_0
- MobileNet is tested only on CPUFreq: 1000MHz
- Optimization is more efficient than not optimised
- Best case: 5 Thread, Optimised

Fig.13 Device: Device\_0 CPU Frequency (MHz): 1000 MLNetwork: MobileNet



# Follow up actions

Talk with the engineers and show that

- Tests are promising, they should try to test:
  - CPU Freq: 1000 and 2000MHz on Device\_1 with MobileNet and 5 thread
  - AlexNet optimized
- MobileNet seems very promising as performance compared to AlexNet both for time and memory
- Optimization is fundamental, AlexNet should implement it
- Highest build generally perform better



# Performance Analysis of Machine Learning Software

Federico Bacci