



Benchmarking vs. Benchmarketing

Why should you care?

Felix Baum

AI Software, Director of Product Management

Qualcomm Technologies, Inc.

September, 2020



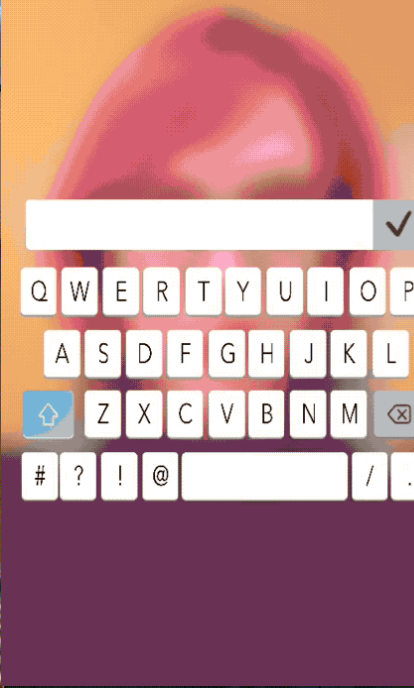
*AI will drive transformation
across industries*



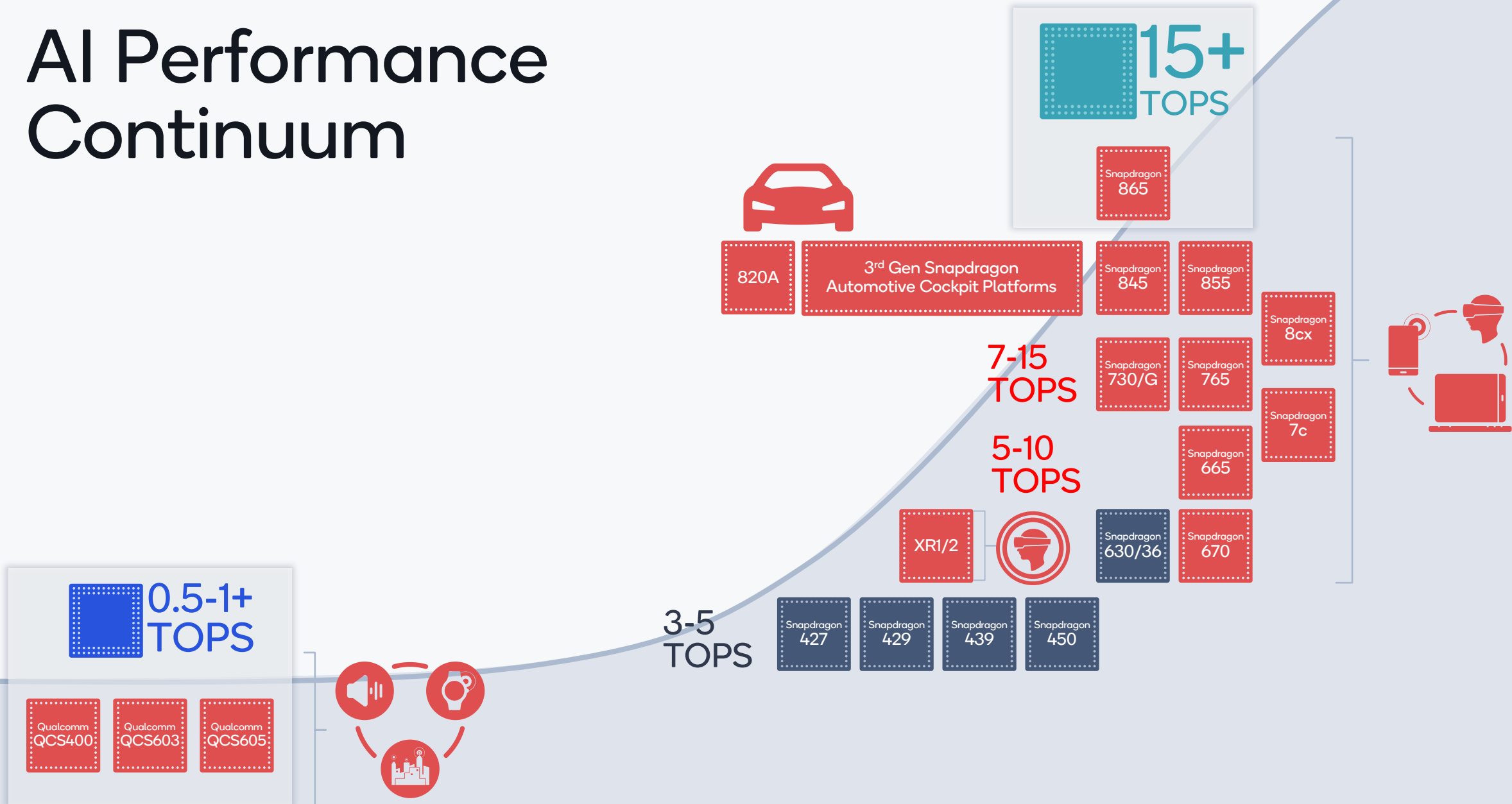
*Powering the factory
of the future*



*Shaping the future
of transportation*



AI Performance Continuum



AI Performance Continuum

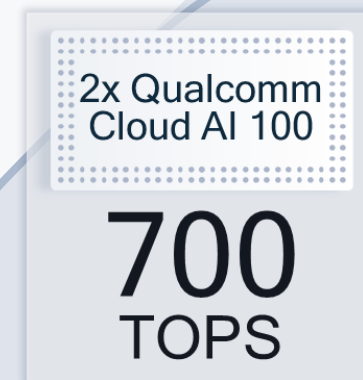
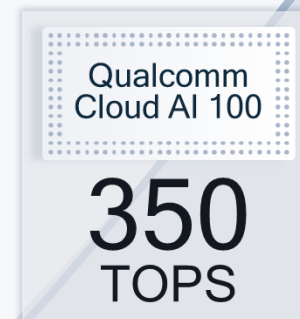
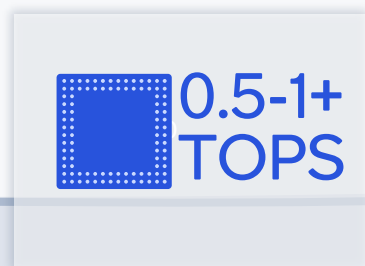


Photo Capture with AI



Still

FPS

~5

TOPs required

~4

Neural Nets

Segmentation

De-noise

Classification

Detection

4K HDR video capture with AI



60
FPS

>10
TOPs required

~8
Neural Nets

Segmentation

De-noise

Audio noise suppression

Audio echo cancellation

Classification

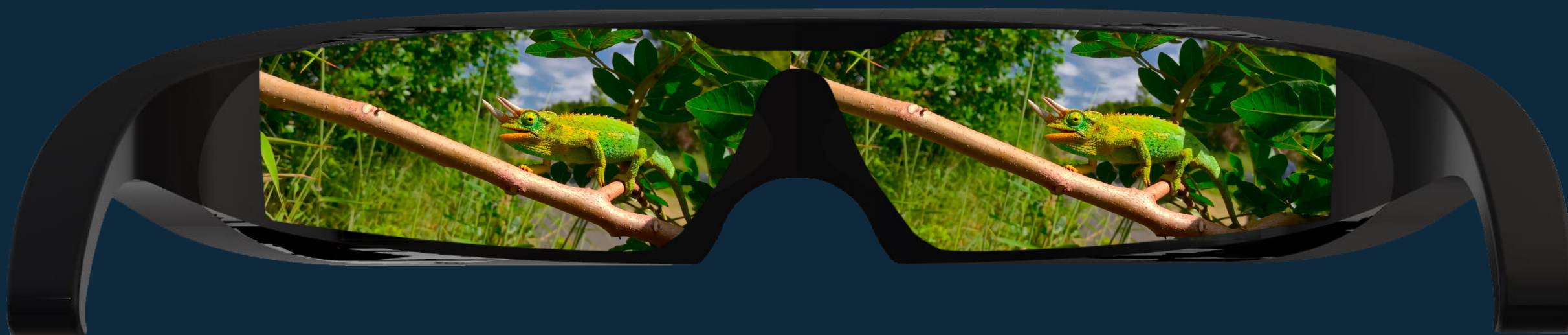
Detection

Enhancement

Semantic Filtering



XR



60

FPS

2

Displays

>15

TOPs required

~8+

Neural Nets

Head tracking

Plane detection

Audio noise suppression

Context Awareness

Hand tracking

Classification

Controllers

Super Resolution

Semantic Segmentation

Facial expression

How to compare AI capabilities of hardware: apples vs. oranges

AI is changing fast and there are factors to consider when using AI benchmarks

Not all TOPS are created equal

Arbitrary

Not actual performance

Incentives vaporware

Intel's new AI chip performs

10 trillion

Operations per second

Intel introduced its new AI-embedded chipset, code-named Keem Bay – *The burn-in*

MediaTek Dimensity 1000 5G marks its comeback into the high-end chip market

...the new chip boasts of

4.5 TOPS (trillion

Operations per second) – Gizmochina

MediaTek Dimensity 800 vs Snapdragon 765G: What's the Difference?

This allows the processor to perform a total of approximately

2.5 TOPS (trillion operations

per second) – DealNTech

How to compare AI capabilities of hardware: apples vs. apples (slightly better)

AI is changing fast and there are factors to consider when using AI benchmarks



Caveats

Not indicative to the best performance that hardware can deliver

Some models may be irrelevant to commercial use cases

Hard to use this to compare hardware from various vendors (some vendors don't support the model)

How to compare AI capabilities of hardware: better apples (fragmented)

AI is changing fast and there are factors to consider when using AI benchmarks

Commercial AI Benchmarks Challenges

- ▶ AI Benchmarks using legacy networks vs, state of the art
- ▶ New networks and operators continue to emerge so older AI benchmarks may not reflect these changes
- ▶ New networks & operators require new software to accelerate processing that AI benchmarks don't support

Factors	Considerations	What is the impact?
Data Type	▶ Integer vs Floating Point	▶ Affects AI use case performance & power
Runtime Framework	▶ Vendor SDK vs NNAPI	▶ OEMs mostly using Vendor SDKs
Score Weighting	▶ Bias scoring toward a test case or category	▶ Can be misleading AI score

AI benchmark Score Bias: (real example)

Benchmark weightage can skew results and wrongfully determine an AI performance “winner”

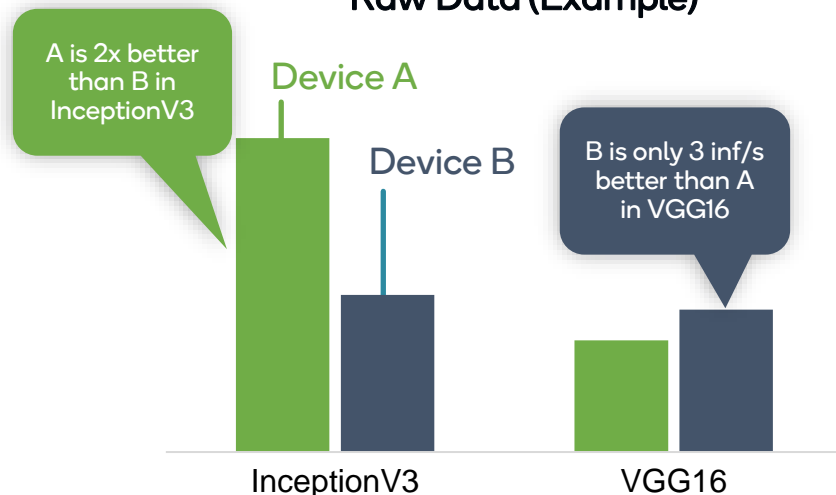
How to check AI benchmark score weightage?

Benchmarks should be checked for fair weighting to reflect real-world usage (e.g., between networks, floating point versus integer data types, CPU/GPU/accelerator cores)

Scores composition must be transparent and using the latest networks & use cases (demonstrate no biases for chipsets, nor obsolete networks/use cases).

Be sure to look at multiple AI workloads/benchmarks

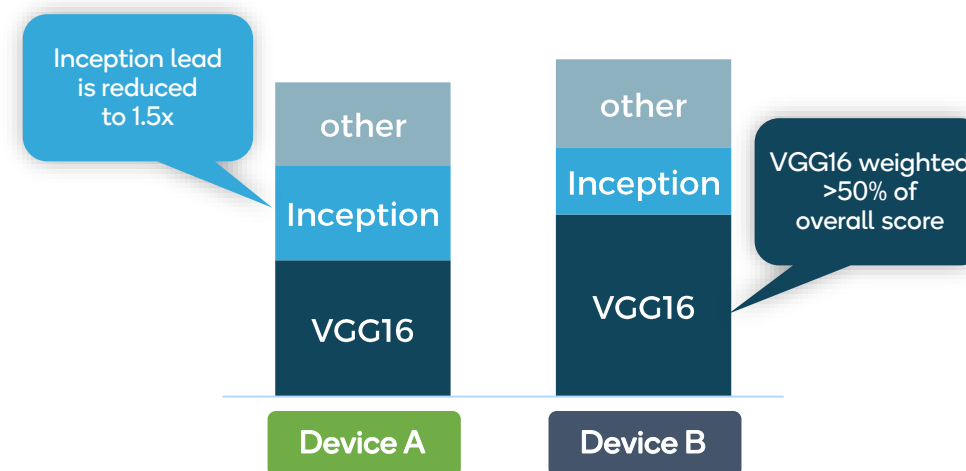
Raw Data (Example)



Who should win?



Overweighting a subtest biases benchmark to skew results



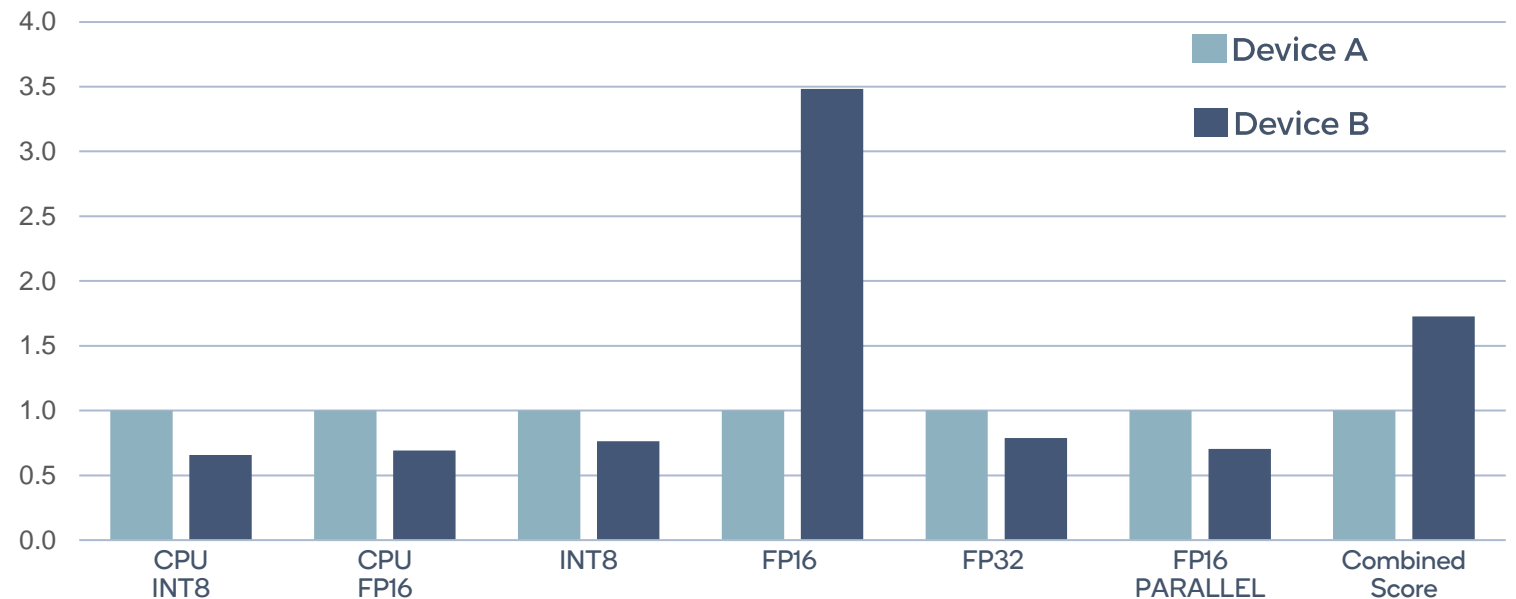
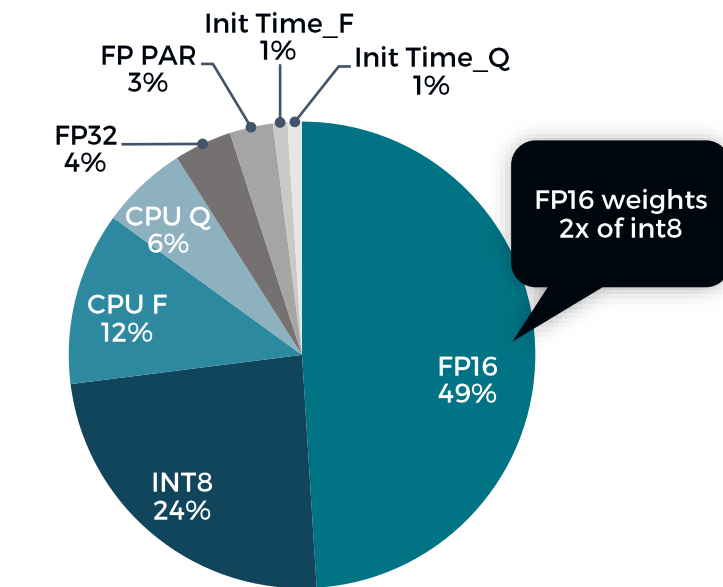
AI benchmark Score Bias: (another example)

Benchmark over weighing FP16 results in a misleading AI performance score

Almost every FP16 test weights 2x than its INT8 counter part

Device A leads Device B on every category except FP16 (Device A leads on FP32)

Normalized score break down for each section



AI benchmarks overview: Almark



Almark has a variety of common AI models and use cases

Description:

Developed by Ludashi Holdings software company

- Ludashi offers variety of software products for PCs and smartphones

Almark considerations:

- Common models which most chipset vendors use for acceleration
- Using Qualcomm® Neural Processing SDK (Qualcomm NPE) allows concurrent use of Qualcomm's Hexagon Vector eXtensions +Hexagon Tensor Accelerator that shows Qualcomm's AI performance capability
- Most vendors and ISVs use each chipset vendor's SDK, so Aimark results are more representative
- Uses INT8 models

Use Cases	Models
Classification	InceptionV3
Classification	Resnet34
Object Detection	MobilenetV1_SSD
Segmentation	DeeplabV3



Access to ranking chart

Aimark

benchmark ranking can be accessed within the app

By clicking "Benchmark Chart" benchmark ranking will be displayed

AI benchmarks overview: Altutu



Altutu benchmark focuses on Classification & Object Detection use cases

Description:

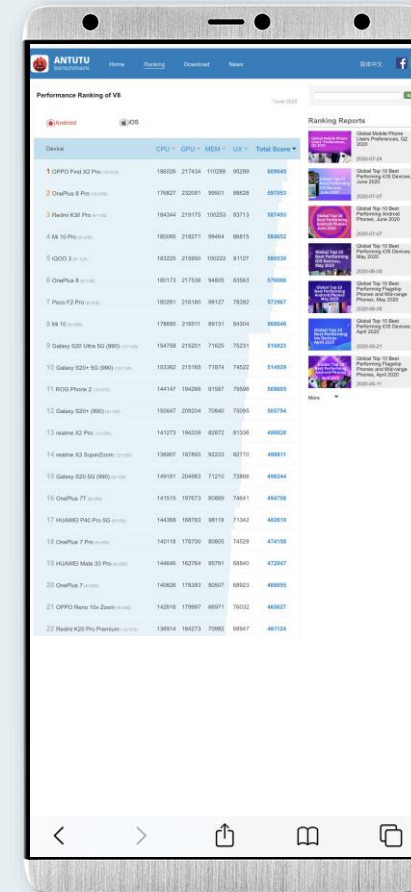
Developed by Antutu

- Antutu benchmark developer well known for Antutu system benchmark
- Altutu utilizes vendor's SDK

Altutu considerations:

- Includes a breakdown of scores for accuracy
- Leverage both of Qualcomm's Hexagon Vector eXtensions +Hexagon Tensor Accelerator
- Uses INT8 precision

Use Cases	Models
Classification	InceptionV3
Object Detection	MobilenetV2_SSD



Access to
ranking
chart

Altutu
(Antutu AI
Benchmark)
app available
at Google Play
and Antutu
Website

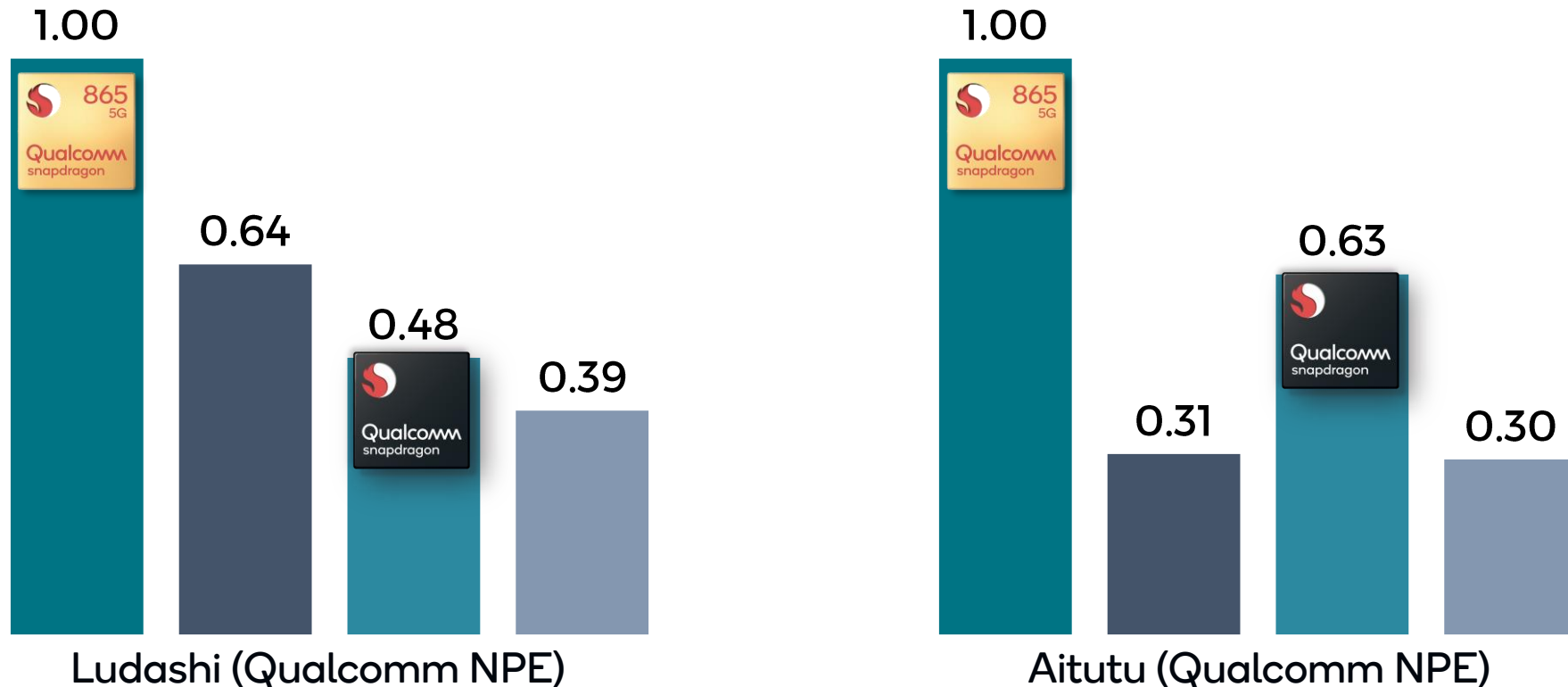
Aitutu benchmark
rankings available
in English and
Chinese

Reference SDK based AI benchmark comparisons

Snapdragon 865 and Snapdragon 765 significantly lead on AI performance

- Reflective of actual users experience on Snapdragon 865 and 765 smartphones

AI benchmarks summary



AI benchmarks overview: ETH Zurich AI Benchmark

ETH Zurich AI Benchmark evaluates variety of use cases on the various processors

Description:

Developed by graduate student in Computer Vision Lab at ETH Zurich University

- First benchmark using TFlite
- Latest version v4 had just been released

ETH AI Benchmark considerations:

- Uses a range of AI models & use cases
- Includes a breakdown of scores
- **14** out of 43 tests utilize Hexagon Vector eXtension
 - Relevant to actual user's experience on smartphone
- **29** out of 43 tests are FP16 or utilize only CPU
 - Helpful reference to developers who are evaluating the performance of not yet commercialized and unoptimized models during development

ETH AI Benchmark v4

Use Cases	Models	Notes
Classification	MobilenetV2	
Classification	InceptionV3	
Face Recognition	MobilenetV3	
Classification	MobilenetV2	
OCR	CRNN	FP only
Deblur	PyNet-Mini	
Super Resolution	VGG19	
Super Resolution	SRGAN	
Bokeh	U-Net	
Segmentation	DeeplabV3+	
Enhancement	DPED-Resnet	
Text Completion	LSTM	FP only
Memory Limits	SRCNN	Memory test

MLPerf benchmark

Description:

Open industry forum group establishing benchmark standard for ML hardware and software

- Represents the consensus of the industry due to large list of the participating companies (see <https://mlperf.org/>)
- New rounds of result submission and first version of the mobile AI benchmark app is expected to be available in Oct 2020

MLPerf considerations:

- Active participation and feedback from vendors such as Google, Intel, Nvidia, Qualcomm, etc.
- Benchmark implementation and decision process is transparent



There are good and bad benchmarks – Here is why

AI is changing fast and there are factors to consider when using AI benchmarks

TOPS	Public well-known repeatable workflows	Well Known Benchmarks
Arbitrary	Not indicative to the best performance that hardware can deliver	Can be biased on certain types of models
Not actual performance	Some models may be irrelevant to commercial use cases	Tend to use legacy models
Incentives vaporware	Hard to use this to compare hardware from various vendors	Vendor results might not be accelerated

**Bottom
Line:**

There is no perfect solution to comparing AI hardware but certain benchmarks are more representative of actual performance



Thank you

- Altutu benchmark ranking at:
<http://www.antutu.com/ranking/rank3.htm>
- AIMark benchmarking ranking at:
https://play.google.com/store/apps/details?id=com.ludashi.aibench&hl=en_US
- ML Perf benchmark at:
<https://mlperf.org/get-involved#join-the-forum>

2020 Embedded Vision Summit

- “Advancing Embedded Vision for an Autonomous World”
 - Tuesday, September 15, 11:30 – 12:00 AM PT
- Qualcomm AI: Leading the Way with Distributed Intelligence
 - Thursday, September 24 11:00 – 11:30 AM PT