AIBench Scenario: Scenario-distilling AI Benchmarking

Wanling Gao

Benchmarking in the Data Center: Expanding to the Cloud (BID'21)

2021.2.28

Bench Council

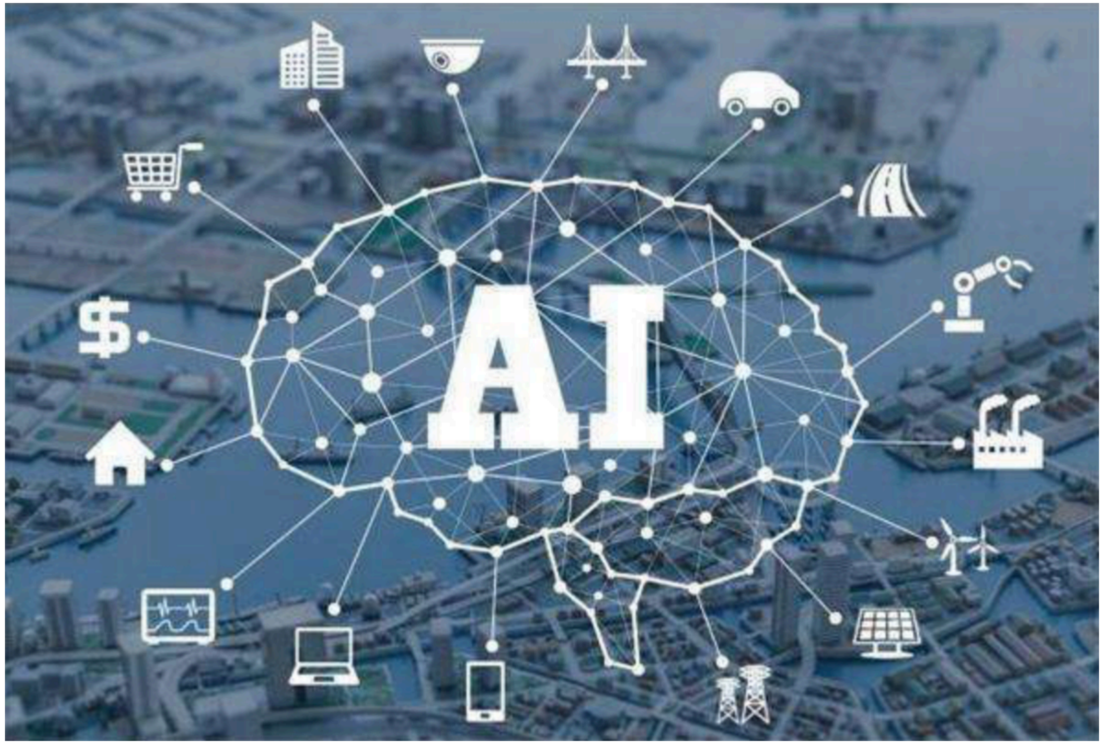
Acknowledgement

- Thanks for the invitation of Prof. Juan (Jenny) Chen

- Thanks for the workshop organizaters

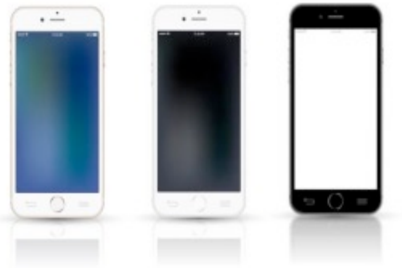
AI: One of the Most Important Workloads

Datacenter



Supercomputing

Edge Computing



AIoT

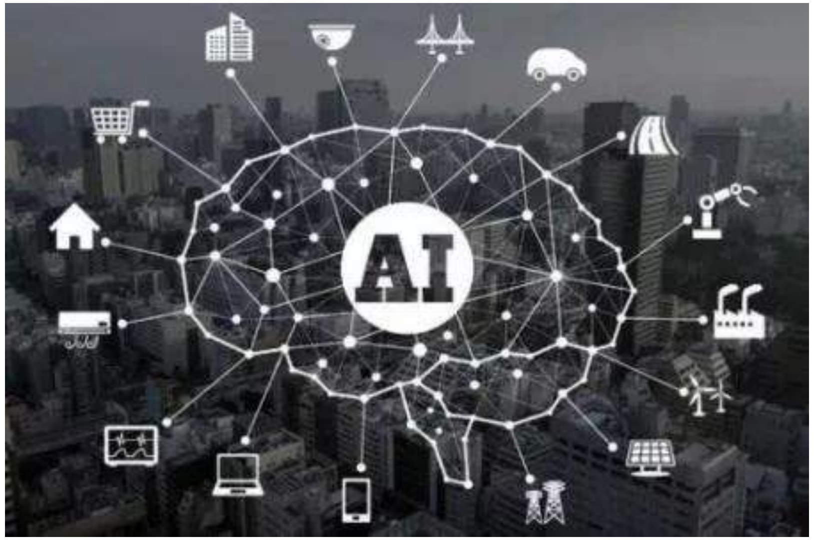
Bench Council

The Challenges of AI Benchmarking

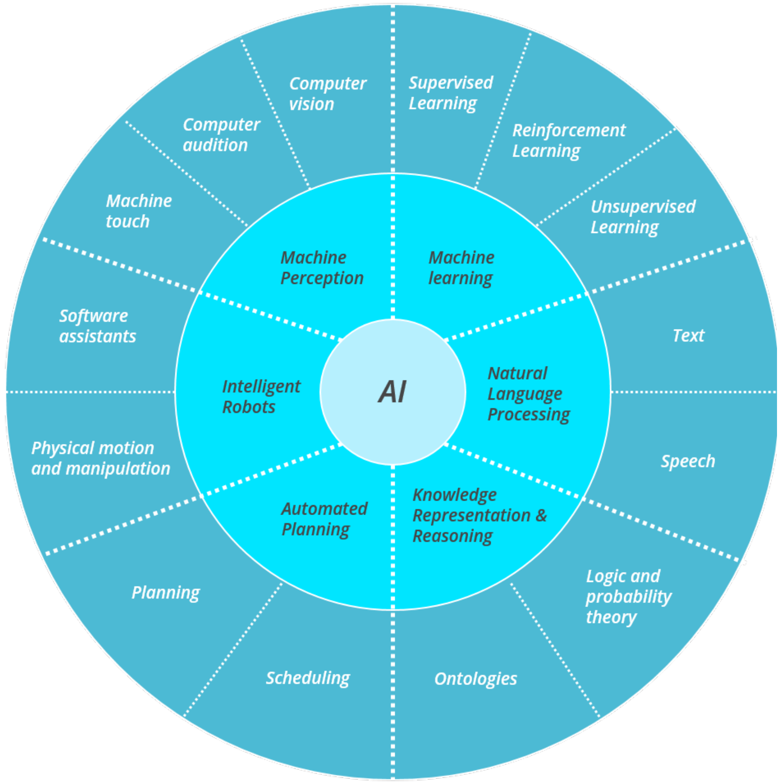
- FIDSS
  - Fragmented
  - Isolated
  - Dynamic
  - Service-based
  - Stochastic

Jianfeng Zhan, Lei Wang, Wanling Gao, and Rui Ren. BenchCouncil's View On Benchmarking AI and Other Emerging Workloads. Technical Report 2019. https://arxiv.org/abs/1912.00572

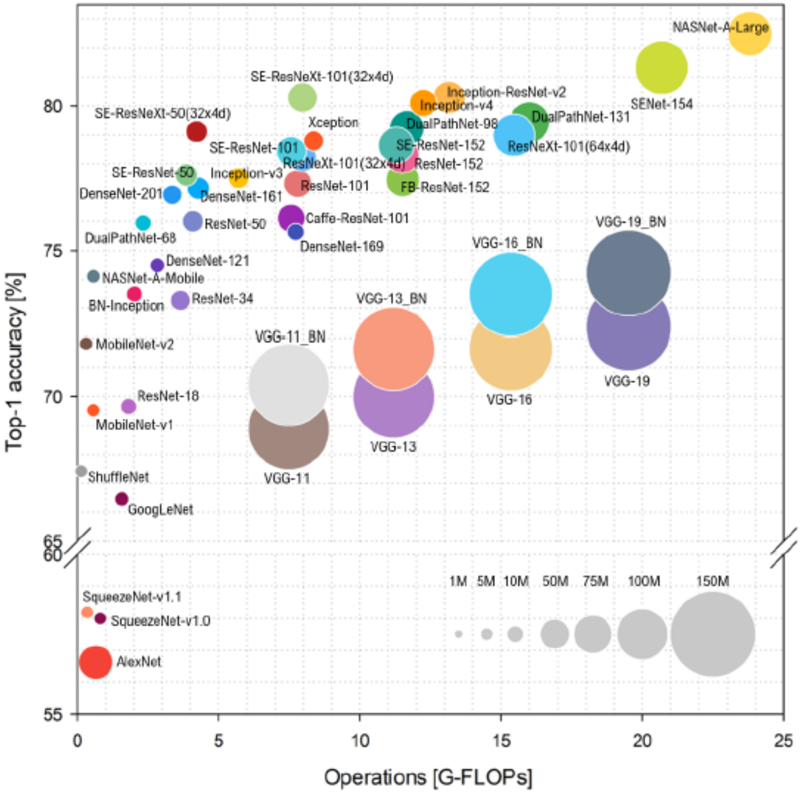
Fragmented: A Huge Variety of Application Scenarios and Models Scenarios



Domains



Applications

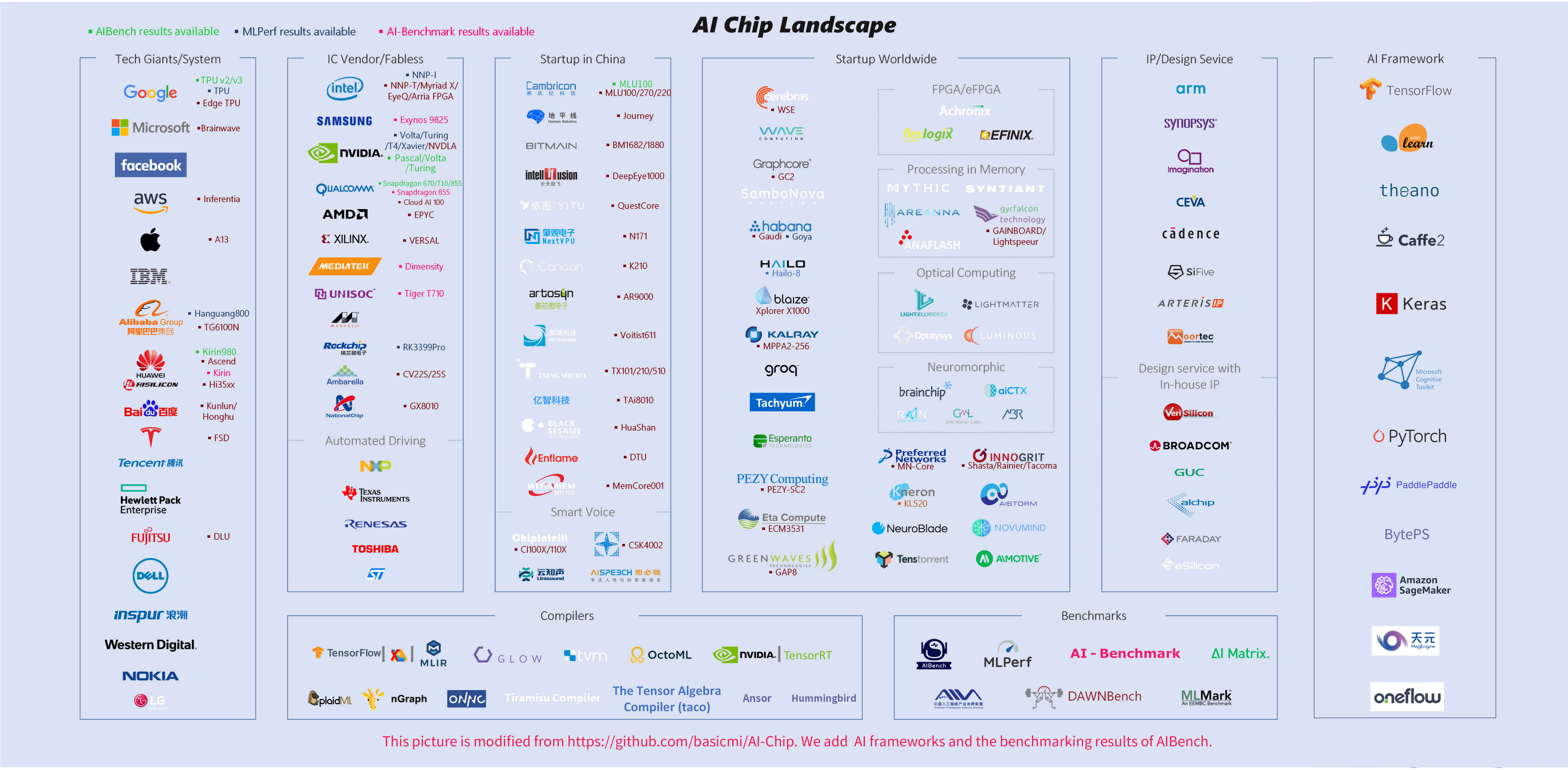


Models

Picture from:
[1]:http://www.hatdot.com/keji/2552842.html
[2]:https://medium.com/appanion/a-five-minute-guide-to-artificial-intelligence-c4262be85fd3
[3]:Bianco, S. et al. "Benchmark Analysis of Representative Deep Neural Network Architectures." IEEE Access 6 (2018)

Bench Council

Landscape of AI Chips



Adapted from the Source: github.com/basicmi/AI-Chip

Council

Isolated: Hidden within Giant's Datacenters

- Real-world datasets and workloads or even AI models are treated as first-class confidential issues

- Isolated between academia and industry, or even among different providers.

- Poses a huge obstacle for our communities towards developing an open and mature research field.

Dynamic Complexity

- Common requirements are handled collaboratively by datacenters, edge, and IoT devices.

- Different distributions of data sets, workloads, ML models may substantially affect the system's behaviors.

- System architectures are undergoing fast evolutions in terms of the interactions among IoT, edge, and datacenters.

Service-based Architecture: The Side Effect

- SaaS model changes workloads fast
  - workload churn
  - not scalable or even impossible to create a new benchmark or proxy for every possible workload.

- Microservice-based architecture
  - distributed across different datacenters
  - consist of diversity of various modules with very long and complex execution paths.
  - tail latency matters

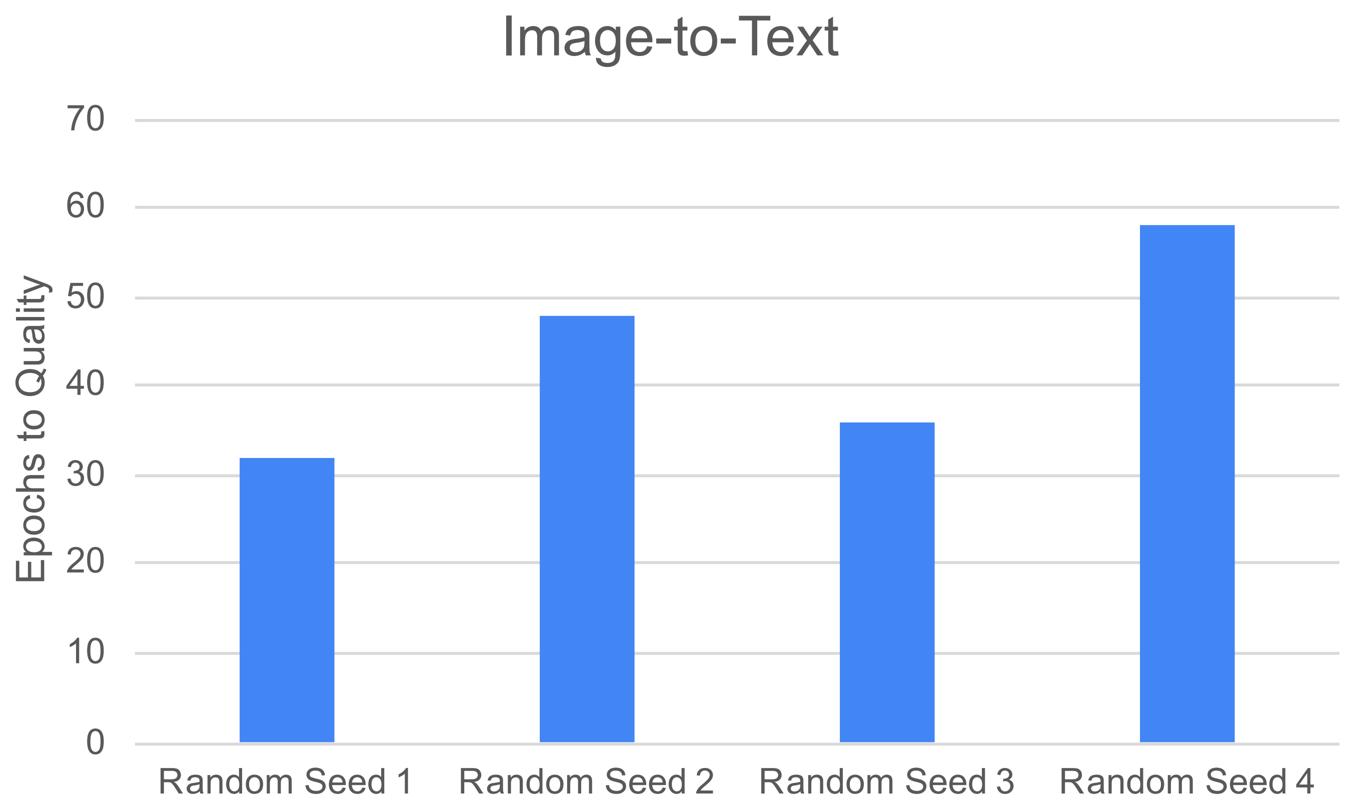
Stochastic nature of AI

- Random seeds affect model initialization, data traversal order, etc.
- Non-idempotence of floating-point operations
- Huge hyper-parameters

BenchCouncil

Example: Randomness

- The epochs to achieve target quality vary significantly under different random seeds



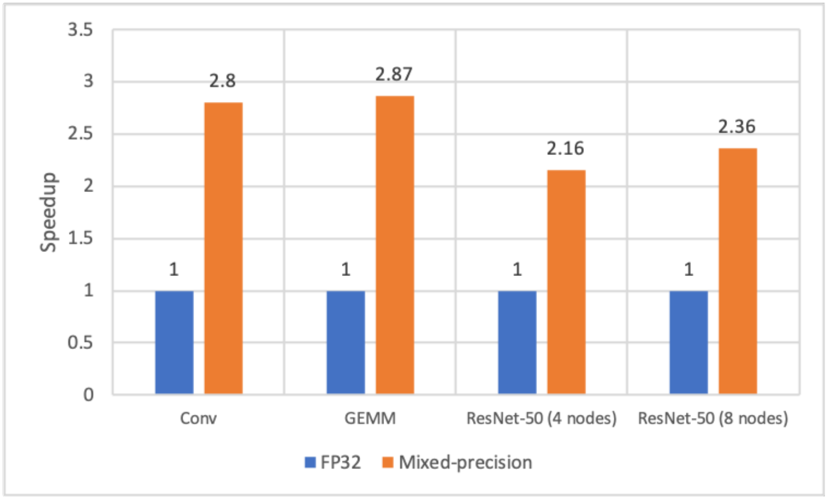
Run Image-to-Text from AIBench four times using different random seeds

[1]F. Tang *et al.*, "AIBench Training: Balanced Industry-Standard AI Training Benchmarking"

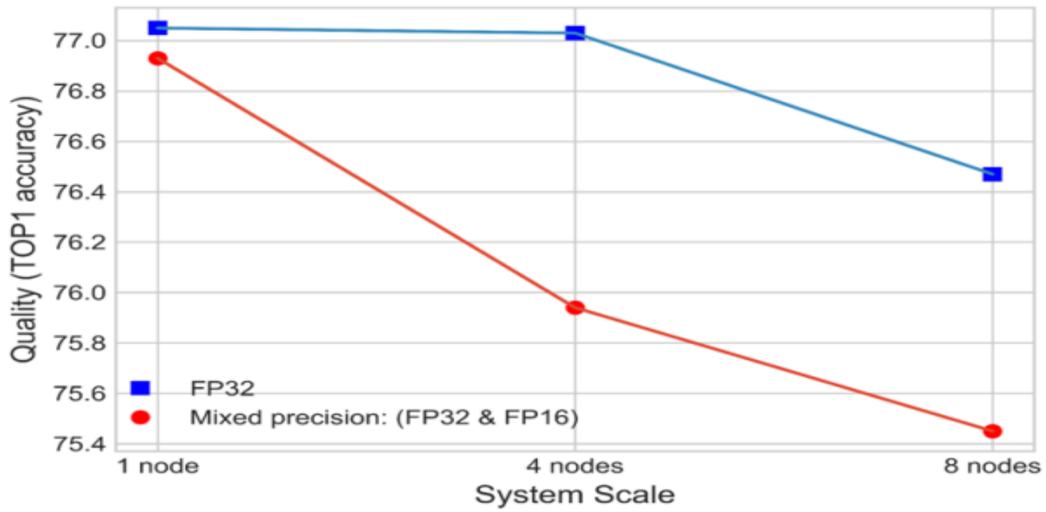
- Challenges to Traditional Benchmarking Methodology

Is Micro Benchmark Sufficient ?

- AI workloads need to consider both computational efficiency and model quality
  - FLOPS is no longer the only metric
- Mixed-precision training significantly improve FLOPS, however, it may deteriorate the model quality



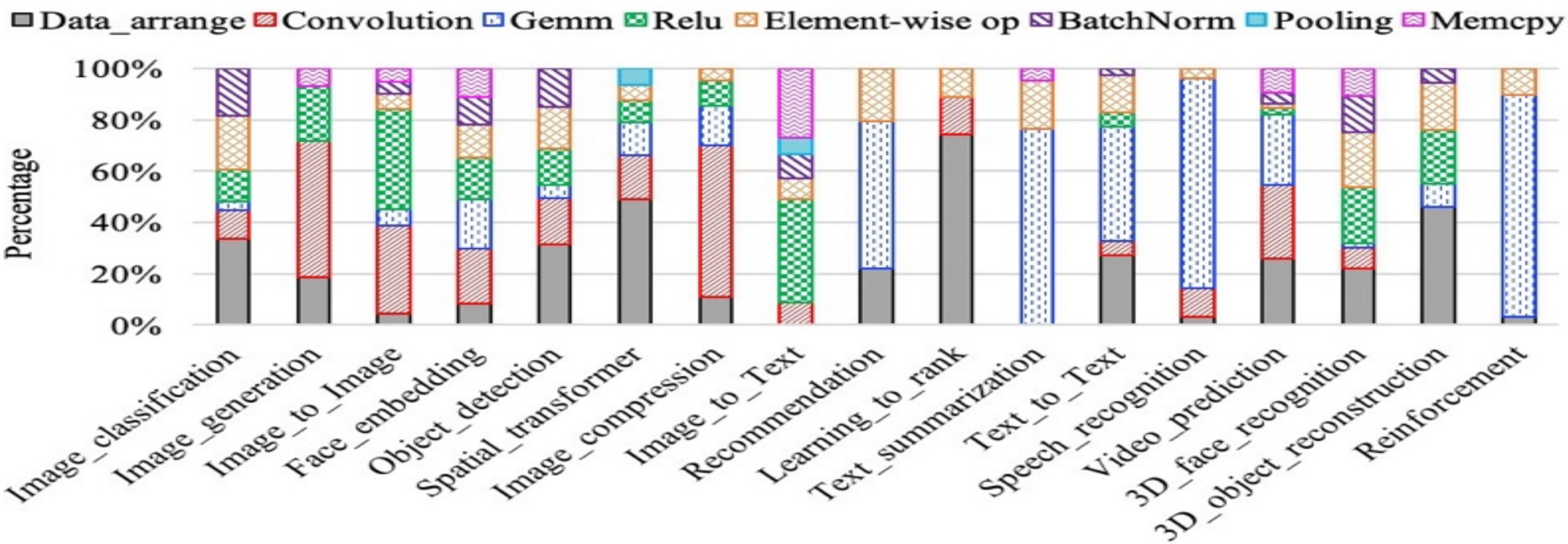
FLOPS comparison of ResNet50 model and operators



The ResNet50 quality comparison

No Single Kernel

- The kernels' runtime breakdown of 17 AI workloads
  - Some micro benchmarks may occupy a little percentage



Bench Council

Is Component Benchmark Sufficient ?

Benchmarking with a Single Component

Image-to-Text

OCR

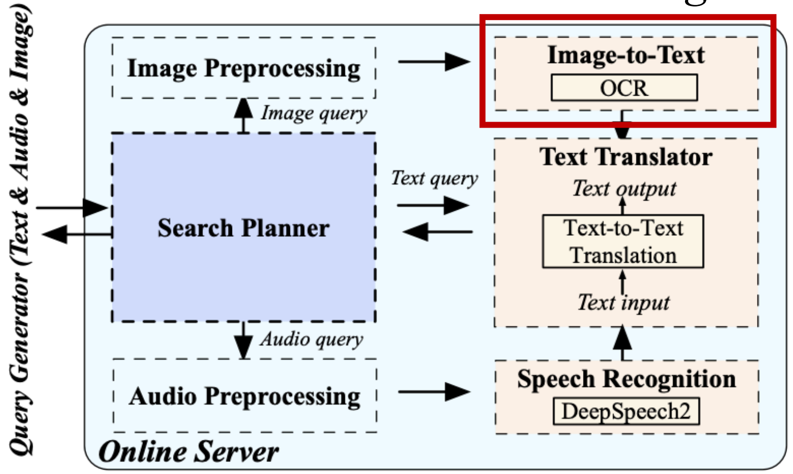
OCR Model Pruning

Tail latency shortens by **51%**

Accuracy decreases **1%**

Optimization Technique

Putting the Component into a realistic scenario: Online Translation Intelligence



OCR Model Pruning

Tail latency has no improvement

Accuracy decreases **1%**

Inconsistent conclusion

Only a single component may need to error-prone conclusions

Wanling Gao, Fei Tang, Jianfeng Zhan, Xu Wen, Lei Wang, Zheng Cao, Chuanxin Lan, Chunjie Luo and Zihan Jiang. AIBench: Scenario-distilling AI Benchmarking. arXiv preprint arXiv:2005.03459.

Single Component vs. Realistic Application

- ***E-commerce Search Intelligence***
- The overall system tail latency deteriorates even 100X comparing to a single component tail latency
  - **2.2X** comparing to recommendation component
  - **180X** comparing to text classification component

Benchmarking a single component cannot reflect the overall system's effects

Wanling Gao, Fei Tang, Jianfeng Zhan, Xu Wen, Lei Wang, Zheng Cao, Chuanxin Lan, Chunjie Luo and Zihan Jiang. AIBench: Scenario-distilling AI Benchmarking. arXiv preprint arXiv:2005.03459.

Model Accuracy vs. QoS

- For *E-commerce Search Intelligence*
  - Model accuracy improvement **1.5%** => overall system 99th percentile latency deteriorates by **9.7X**
    - Replace ResNet50 with ResNet152 for image classification
      - Overall system 99th percentile latency
        - 1136.79 millisecond => 10985.49 millisecond

Benchmarking a single component cannot reflect the tradeoff between model accuracy and QoS

Wanling Gao, Fei Tang, Jianfeng Zhan, Xu Wen, Lei Wang, Zheng Cao, Chuanxin Lan, Chunjie Luo and Zihan Jiang. AIBench: Scenario-distilling AI Benchmarking. arXiv preprint arXiv:2005.03459.

Statistical Model + Component Benchmarks ?

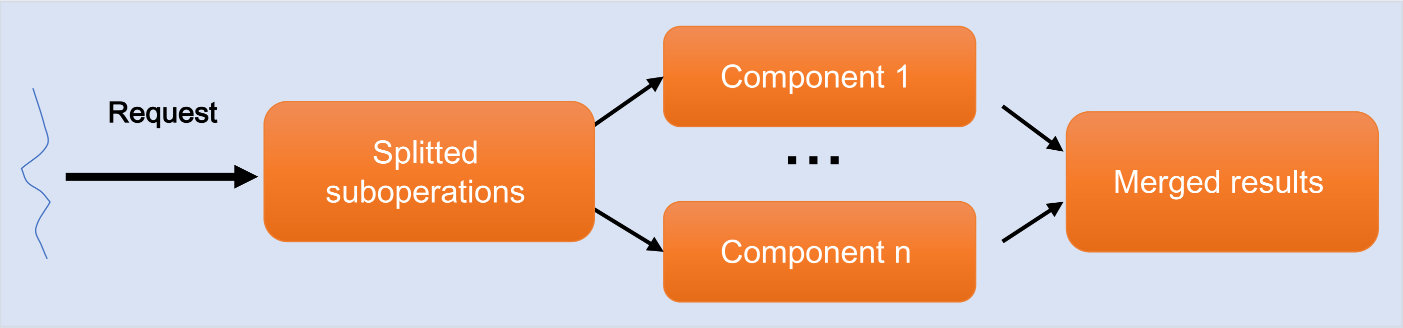
- Whether a statistical model can predict the overall system tail latency, through profiling many components' tail latency performance ?
  - **NO !**
- Simple queueing model
  - E-commerce Search Intelligence Scenario
    - **8.6X** between the actual average latency and the theoretical one
    - **3.3X** between the actual 99th percentile latency and the theoretical one
- Sophisticated queueing network model
  - E-commerce Search Intelligence Scenario
    - **4.9X** between the actual average latency and the theoretical one
    - **Difficult** for tail latency predicting: non-superposition property

Wanling Gao, Fei Tang, Jianfeng Zhan, Xu Wen, Lei Wang, Zheng Cao, Chuanxin Lan, Chunjie Luo and Zihan Jiang. AIBench: Scenario-distilling AI Benchmarking. arXiv preprint arXiv:2005.03459.

Scenario Benchmark is needed !

- A proxy of a realistic application scenario
  - The real one is treated as first-class confidential issues

- Capturing the critical path and primary modules
  - The **permutations** of a series of AI and non-AI components



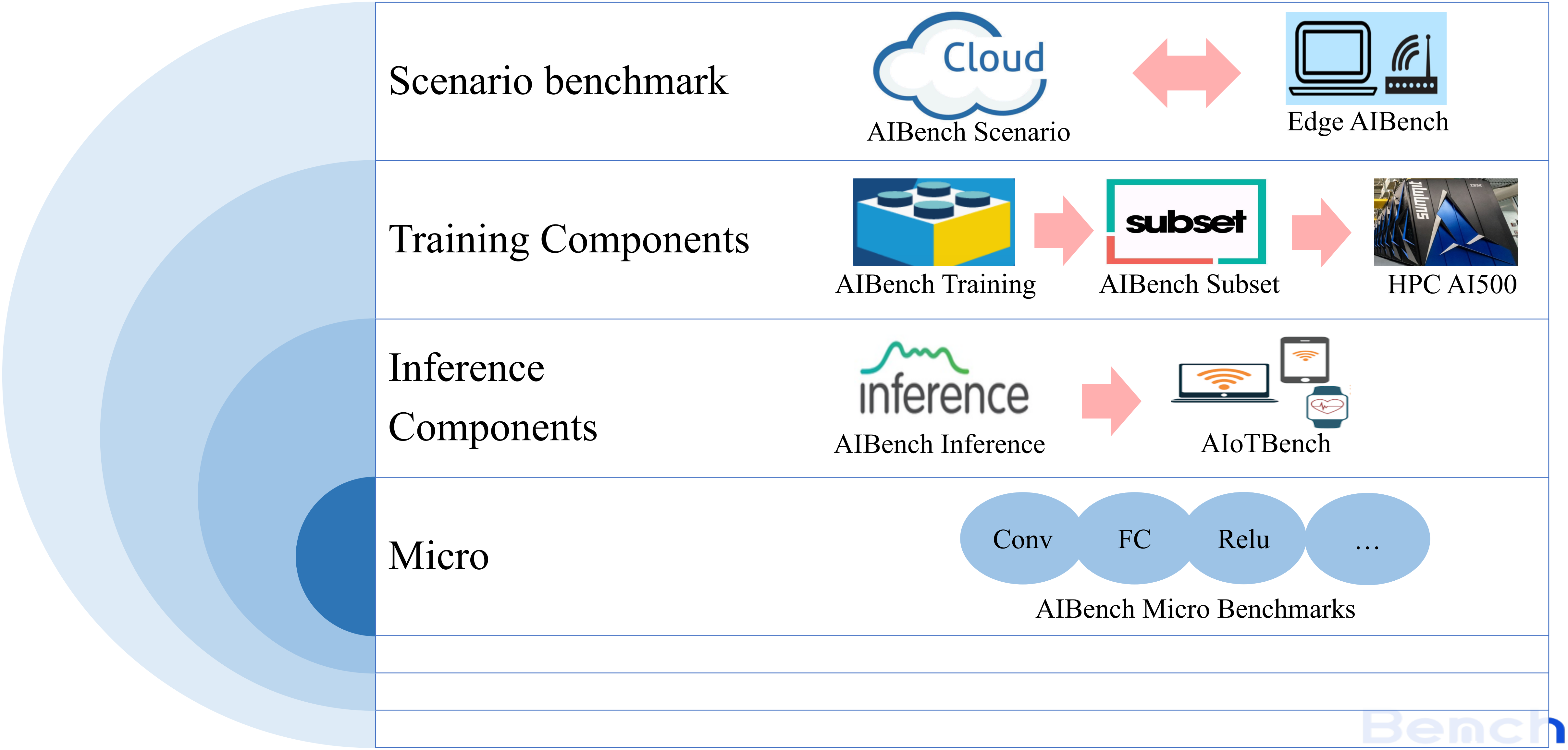
Our Methodology

- Scenario benchmarks
  - Overall system performance

Consider Conflicting Benchmarking Requirements

- Benchmarking at different stages
  - Earlier-stage evaluations of a new architecture or system：
    - Portability（Micro benchmarks）
    - Simplicity
  - Later-stage evaluations or purchasing off-the-shelf systems：
    - Comprehensiveness/Representativeness
    - Reality and system performance （Component or scenario benchmarks）

AIBench Summary



Overview

- Challenges
- **Related Work**
- AIBench
  - AIBench Scenario
    - Edge AIBench
  - AIBench Training
    - HPC AI500
  - AIBench Inference
    - AIoTBench
- Conclusion

AI Benchmarks for Datacenters

Fathom
arXiv 2016
IISWC 2016
Eight workloads
Training and inference
No quality metric

DNNMark
GitHub 2016
GPGPU 2017
Eight micro benchmarks

DeepBench, Baidu
Github 2017
AI basic operators, containing gemm, convolution, recurrent layer and all reduce
Only has micro benchmarks

BenchIP
arXiv 2017
JCST 2018
10 microbenchmarks
11 neural network models

DAWNBench, NIPS 2017
Image classification and question answer
Use time-to-accuracy as metric

TBD Suite
GitHub 2018
IISWC 2018
Eight workloads, six domains

AIBench, Bench 2018
arXiv 2019, 2020
First proposing scenario benchmarks
17 tasks, 17 workloads, 3 subsets

MLPerf, 2018
GitHub 2019
SysML 2020
Five domains
seven workloads

AIIA-DNN
GitHub 2019
Designed to support training and inference, but only provides inference implements now

Bench Council

AI Benchmarks for HPC Systems

DAWNBench, NIPS 2017

Image classification and question answer; the first AI benchmark that uses time-to-accuracy as the metric.

HPC AI500, Bench 2018, arXiv 2020

Bench'18:
Cover 3 representative application of scientific deep learning.
arXiv 2020:
Hierarchical benchmarking methodology;
3 benchmarking levels and rules;
Use Valid FLOPS as the metric;
Two representative and repeatable AI workloads (Business + Scientific).

Deep500, IPDPS 2019

A framework covering 4 level benchmarking;
No concrete reference implementation.

HPL-AI, 2019

Micro benchmark based on LU decomposition;
Scalable but can not reflect model quality.

MLPerf, arXiv 2019, SysML 2020

7 workloads covering 5 domains;
2 benchmarking levels and rules;
Use time-to-train as the metric.

AIPerf, 2020

Based on AutoML;
Scalable but hard to ensure repeatability.

Bench Council

AI Benchmarks for Edge Computing

Edge AIBench，Bench 2018, arXiv 2019

Scenario benchmarking
ICU patient monitoring, camera monitoring, smart home, and automatic driving
Integrated federal learning

EdgeBench，UCC Companion 2018

Speech recognition, and image classification

EEMBC MLMark，2019

Image classification, object detection, translation, and speech recognition
Closed source

AI Benchmarks for IoT

AIOT，Bench 2018

Vision, audio, and NLP domain

Supports Android and Raspberry Pie

TensorFlow Lite、Caffe 2

End-to-end、microbenchmarks

ETH Zurich AI Benchmark，ECCV 2018

Only supports vision domain

Only supports Android and TensorFlow Lite

End-to-end

Bench Council

Overview

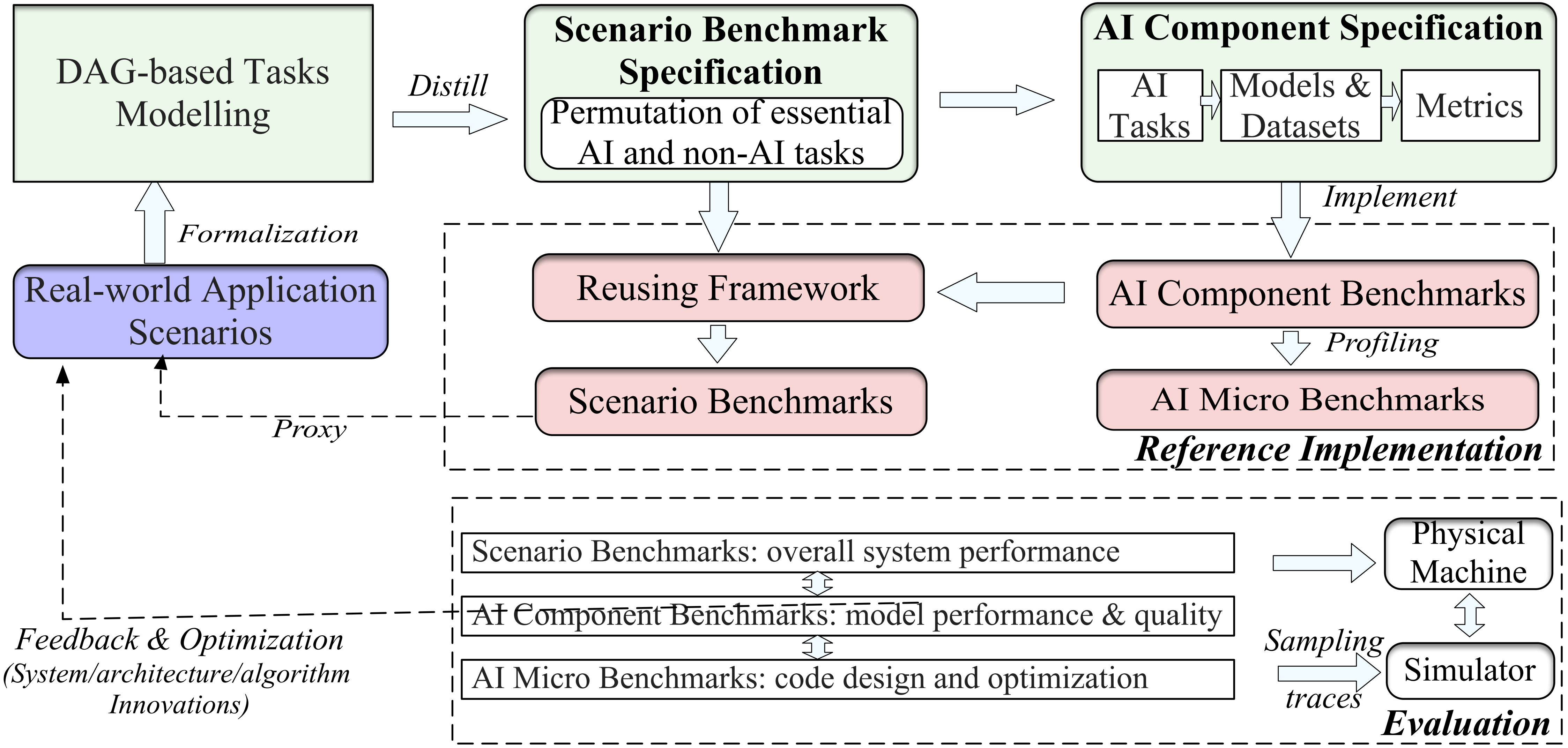
- Challenges
- Related Work
- **AIBench**
  - **AIBench Scenario**
    - Edge AIBench
  - AIBench Training
    - HPC AI500
  - AIBench Inference
    - AIoTBench
- Conclusion

Bench Council

BenchCouncil

- **International Open Benchmark Council (BenchCouncil)**
  - http://www.benchcouncil.org
  - a non-profit international organization
    - Aiming to promote the standardization, benchmarking, evaluation, incubation, and promotion of HPC, Chip, AI, Big Data, Block Chain, and other emerging techniques.

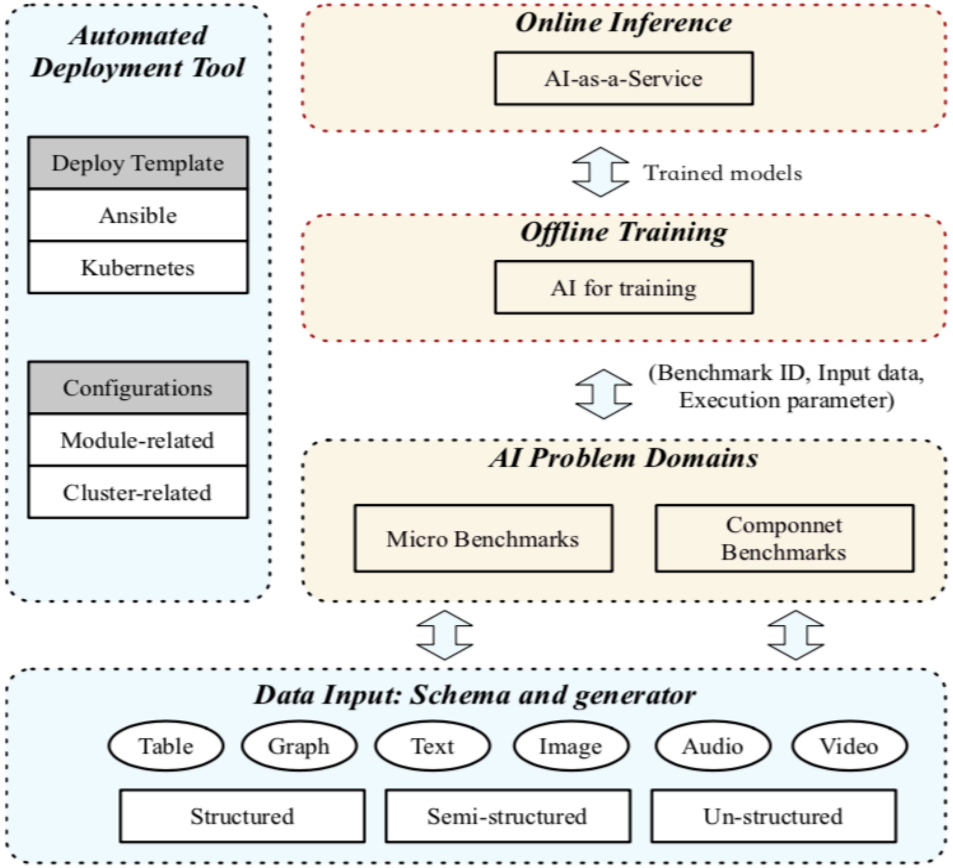
Scenario-distilling Benchmarking Methodology



BenchCouncil

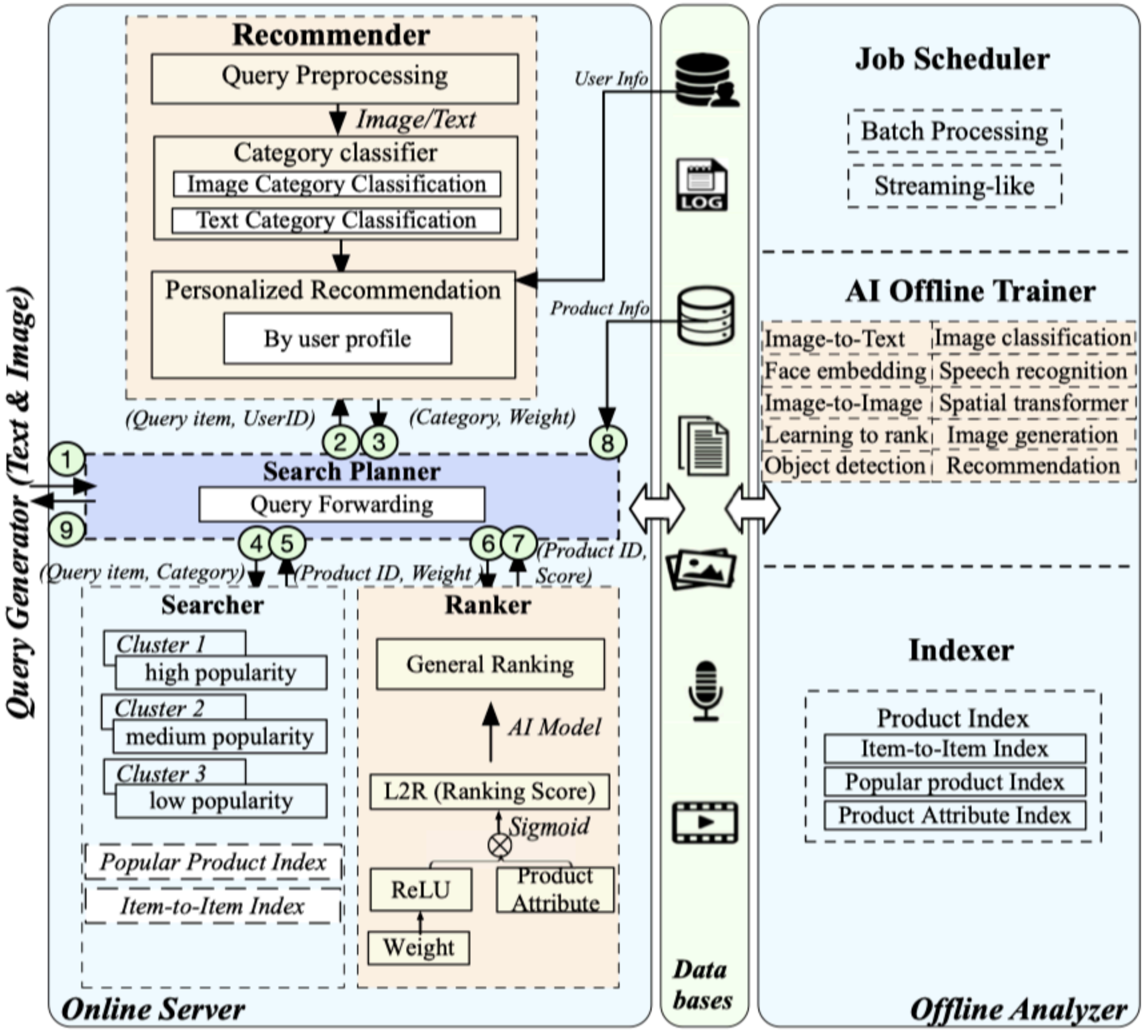
Reusing Framework

- **The First** Reusing Framework for easily constructing scenario benchmarks
  - A highly extensible, configurable, and flexible benchmark framework
  - AI-related and non AI-related Library
  - Support critical paths and primary modules modelling
  - Multiple loosely coupled modules
    - Individually
      - Micro/Component benchmarks
    - Collectively
      - Scenario benchmarks

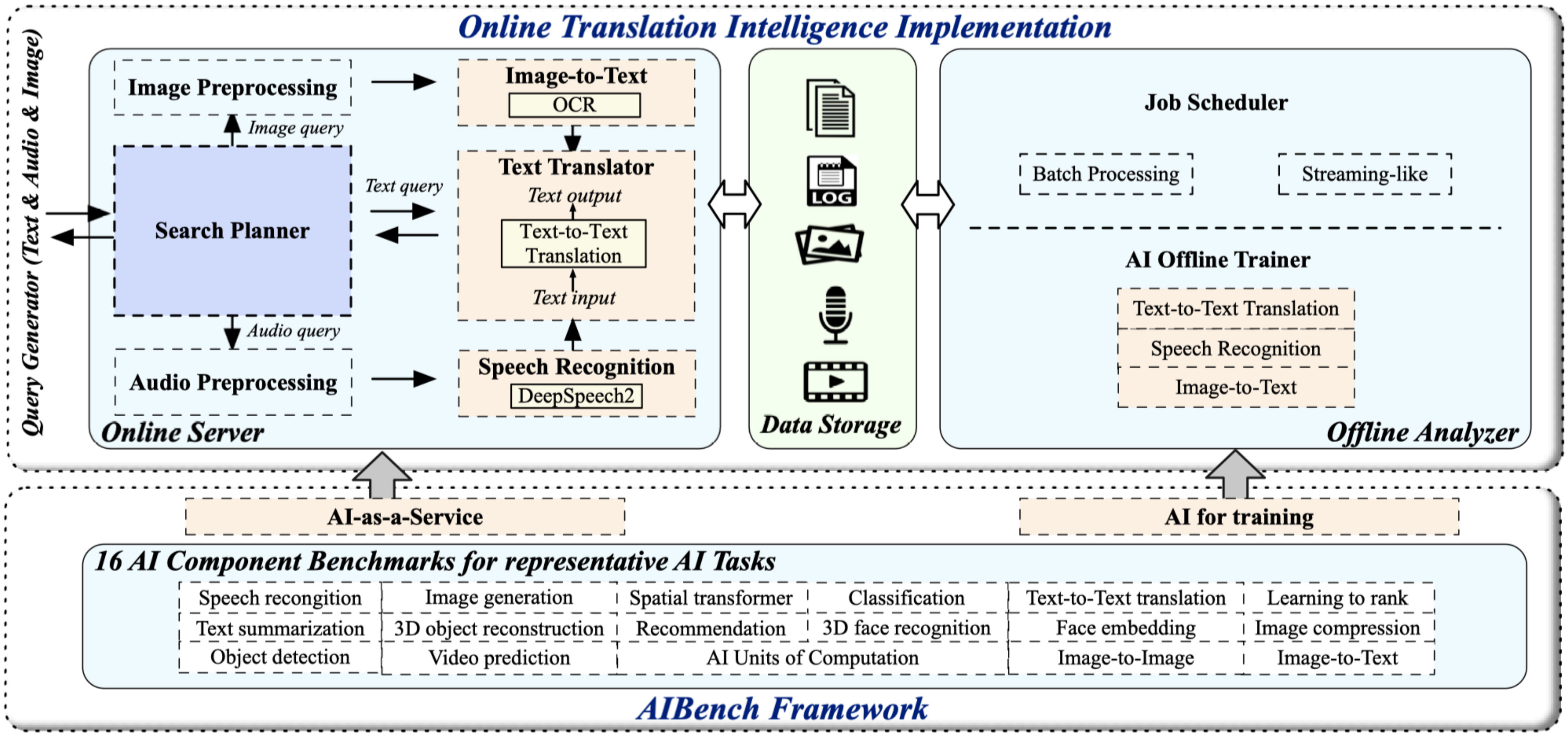


Scenario Benchmark: E-commerce Search Intelligence

- **Query generator**
  - simulate concurrent users and send query requests
- **Online module**
  - personalized searching and recommendations
- **Offline module**
  - a training stage to generate a learning model
- **Data storage module**
  - data storage, e.g., user database, product database



Scenario Benchmark: Online Translation Intelligence

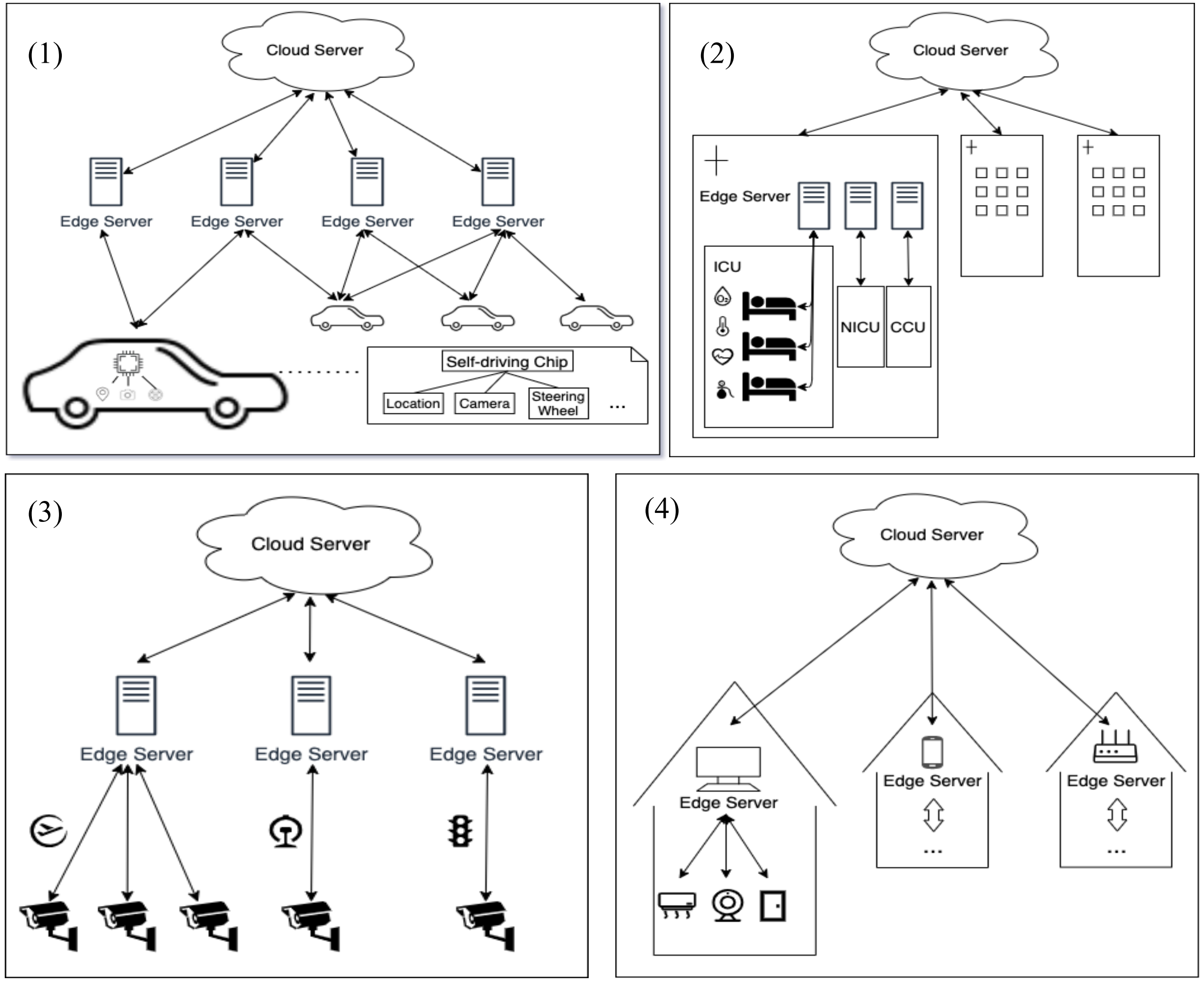


Overview

- Challenges
- Related Work
- **AIBench**
  - AIBench Scenario
    - **Edge AIBench**
  - AIBench Training
    - HPC AI500
  - AIBench Inference
    - AIoTBench
- Conclusion

Four Typical Edge AI Scenarios

- (1) Autonomous Vehicle
  - Latency-sensitive
  - High-accuracy
  - Mobile
- (2) ICU Patient Monitor
  - Latency-sensitive
  - Parallel
- (3) Surveillance Camera
  - Enormous Data
- (4) Smart Home
  - Heterogenous devices and data



Bench Council

Nine Typical Edge AI Tasks

Task Name	Edge AI Scenarios	Models	Datasets	Implementations
Lane Detection	Autonomous Vehicle	LaneNet	Tusimple/ CULane	Pytorch/Caffe
Traffic Sign Detection	Autonomous Vehicle	Capsule Network	German Traffic Sign Recognition Benchmark	Keras
Heart Failure Prediction	ICU Patient Monitor	LSTM	MIMIC-III	Tensorflow/Keras
Decompensation Prediction	ICU Patient Monitor	LSTM	MIMIC-III	Tensorflow/Keras
Death Prediction	ICU Patient Monitor	LSTM	MIMIC-III	Tensorflow/Keras
Person Re-identification	Surveillance Camera	DG-Net	Market-1501	Pytorch
Action Detection	Surveillance Camera	ResNet18	UCF101	Pytorch/Caffe
Face Recognition	Smart Home	Facenet/Sphere network	LFW/CASIA-Webface	Tensorflow/Caffe
Speech Recognition	Smart Home	DeepSpeech2	LibriSpeech	Tensorflow

Overview

- Challenges
- Related Work
- **AIBench**
  - AIBench Scenario
    - Edge AIBench
  - **AIBench Training**
    - HPC AI500
  - AIBench Inference
    - AIoTBench
- Conclusion

Bench Council

Typical Internet service applications (with 17 industry partners)

Internet Service	Core Scenario	Involved AI Tasks
Search Engine	Content-based image retrieval (e.g., face, scene)	Object detection; Classification; Spatial transformer; Face embedding; 3D face recognition
	Advertising and recommendation	Recommendation
	Maps search and translation	3D object reconstruction; Text-to-Text translation; Speech recognition; Neural architecture search
	Data annotation and caption (e.g., text, image)	Text summarization; Image-to-Text
	Search result ranking	Learning-to-rank
	Image resolution enhancement	Image generation; Image-to-Image
	Data storage space and transfer optimization	Image compression; Video prediction
Social Network	Friend or community recommendation	Recommendation; Face embedding; 3D face recognition;
	Vertical search (e.g., image, people)	Classification; Spatial transformer; Object detection;
	Language translation	Text-to-Text translation; Neural architecture search
	Automated data annotation and caption	Text summarization; Image-to-Text; Speech recognition
	Anomaly detection (e.g., spam image detection)	Classification
	Image resolution enhancement	Image generation; Image-to-Image
	Photogrammetry (3D scanning)	3D object reconstruction
	Data storage space and transfer optimization	Image compression; Video prediction
	News feed ranking	Learning-to-rank
E-commerce	Product searching	Classification; Spatial transformer; Object detection
	Product recommendation and advertising	Recommendation
	Language and dialogue translation	Text-to-Text translation; Speech recognition; Neural architecture search
	Automated data annotation and caption	Text summarization; Image-to-Text
	Virtual reality (e.g., virtual fitting)	3D object reconstruction; Image generation; Image-to-Image
	Data storage space and transfer optimization	Image compression; Video prediction
	Product ranking	Learning to rank
	Facial authentication and payment	Face embedding; 3D face recognition;

Bench Council

AIBench Training Workloads

- Coverage of diverse network architectures（CNN、ResNet、LSTM、GRU、Attention, etc.）
  - **Text processing (5)**
    - Text-to-Text, Text summarization, Learning to Rank, Recommendation, Neural Architecture Search
  - **Image processing (8)**
    - Image Classification, Image Generation, Image-to-Text, Image-to-Image, Face Embedding, Object Detection, Image Compression, Spatial Transformer
  - **Audio processing (1)**
    - Speech Recognition
  - **Video processing (1)**
    - Video Prediction
  - **3D data processing (2)**
    - 3D Face Recognition, 3D Object Reconstruction

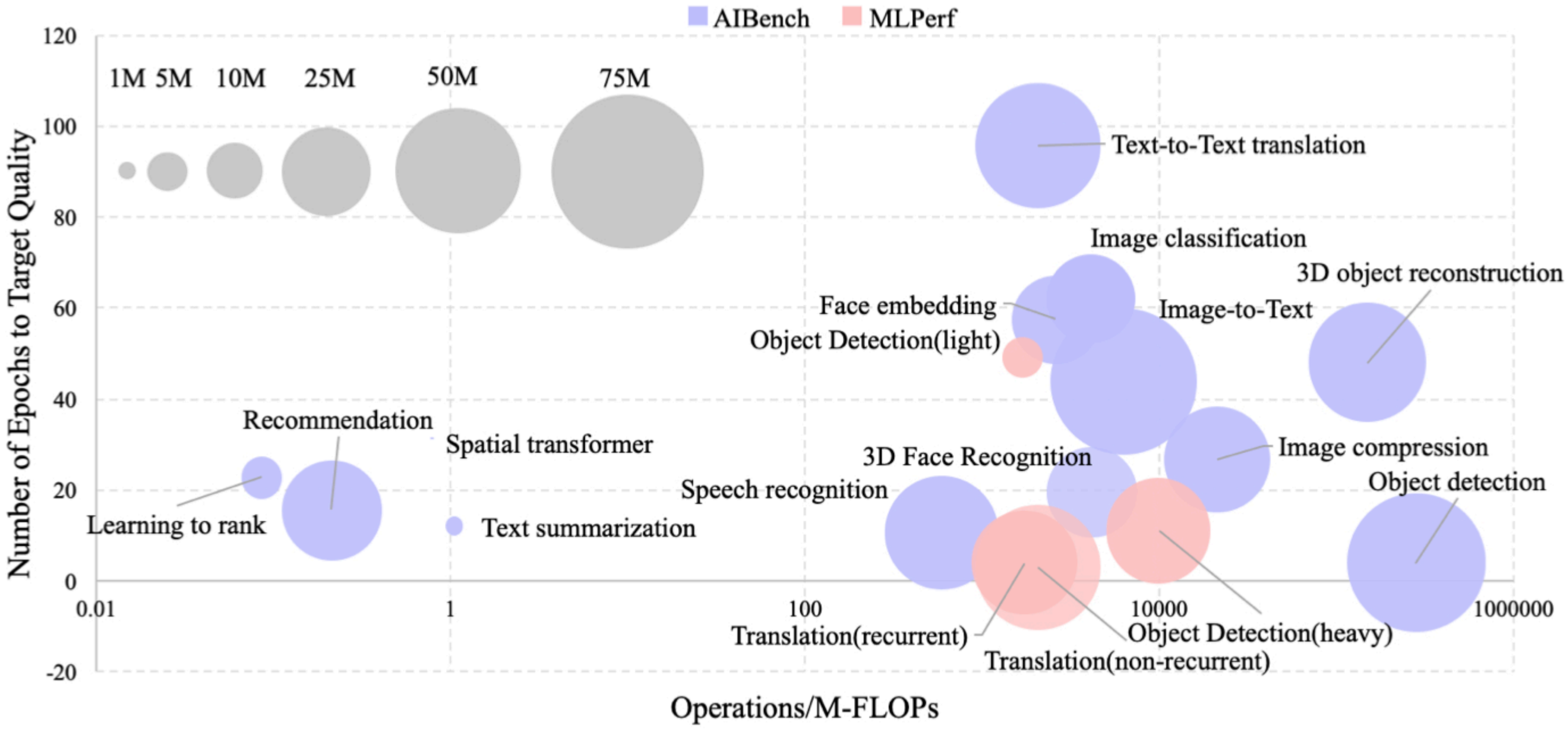
Bench Council

AIBench Training vs. MLPerf Training

- Concurrent work
- AIBench Training has wider coverage
  - Tasks
  - Model complexity
  - Diverse Characteristics
    - Microarchitecture
      - FLOPs computation、memory access pattern、computation pattern、I/O pattern
    - System
      - Evaluation time cost、variation、and convergence of hot functions
    - Algorithms
      - Model architectures and parameters

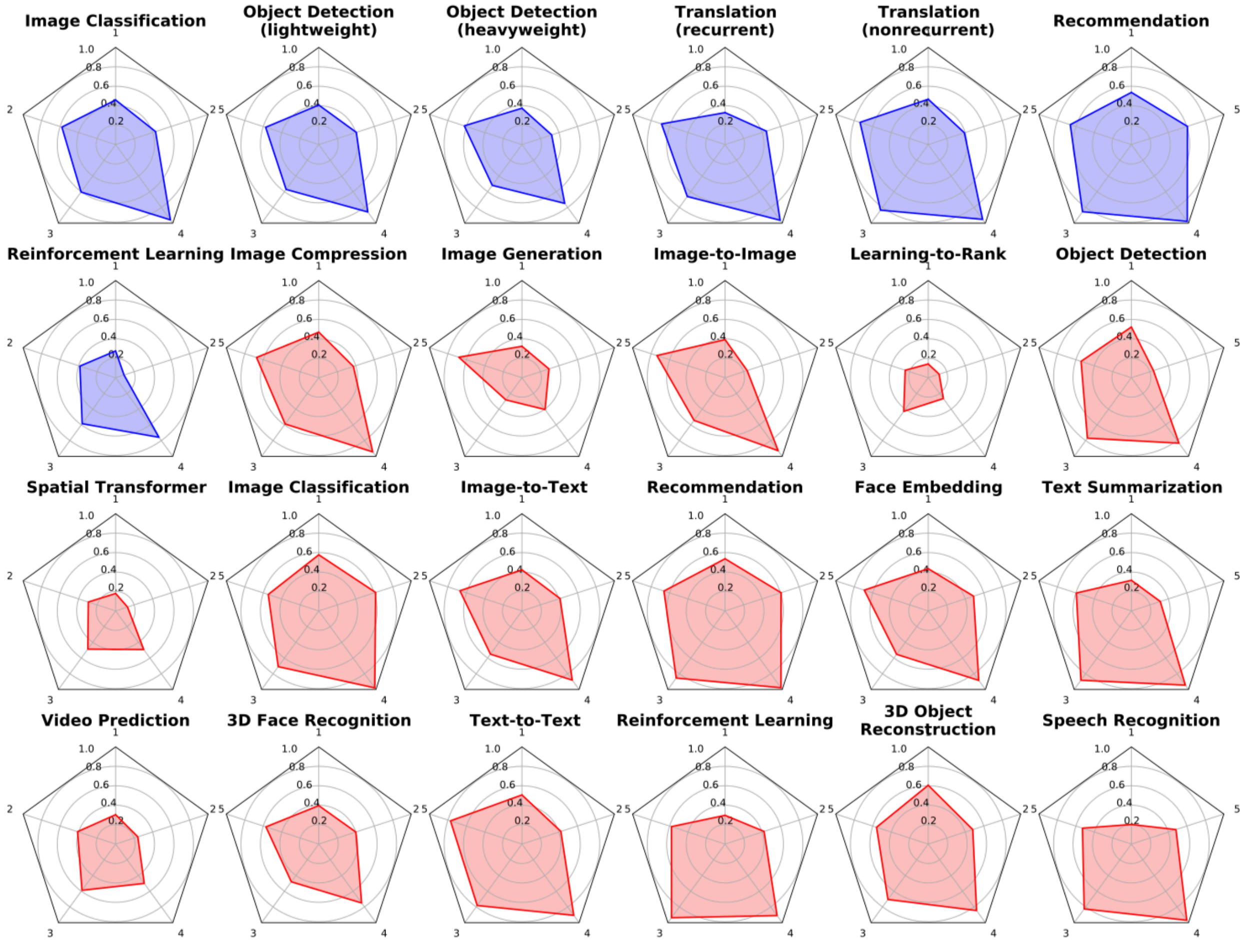
		AIBench Training v1.0	MLPerf Training V0.5
Methodology		Balanced methodology considering conflicting requirements	According to commercial and research relevance
Algorithm		Seventeen tasks and models	Five tasks and seven models
Dataset		Text, image, 3D, audio, and video data	Text and image data
Model behavior	Computation	0.09 to 282830 MFLOPs	0.21 to 24500 MFLOPs
	Complexity	0.03 to 68.4 million parameters	5.2 to 49.53 million parameters
	Convergence	6 to 96 epochs	3 to 49 epochs
System behavior		30 hot functions	9 hot functions
Micro-architecture behavior	Achieved occupancy	0.14 to 0.61	0.28 to 0.54
	IPC efficiency	0.25 to 0.77	0.39 to 0.74
	Gld efficiency	0.28 to 0.94	0.52 to 0.85
	Gst efficiency	0.27 to 0.98	0.75 to 0.98
	DRAM utilization	0.12 to 0.61	0.52 to 0.61

Bench Council


AIBench
MLPerf
120
100
80
60
40
20
0
-20
1M 5M 10M 25M 50M 75M
Number of Epochs to Target Quality
Text-to-Text translation
Image classification
3D object reconstruction
Face embedding
Image-to-Text
Object Detection(light)
Recommendation
Spatial transformer
3D Face Recognition
Image compression
Object detection
Speech recognition
Learning to rank
Text summarization
0.01
1
100
10000
1000000
Translation(recurrent)
Object Detection(heavy)
Translation(non-recurrent)
Operations/M-FLOPs

Micro-architectural Characteristics

- Distinct computation and memory access behaviors
  - AIBench has a wider coverage than MLPerf



1: achieved occupancy
Warps utilization rate

2: ipc efficiency
IPC efficiency

3: gld efficiency
Global memory load efficiency

4: gst efficiency
Global memory store efficiency

5: dram utilization
DRAM utilization

Overview

- Challenges
- Related Work
- **AIBench**
  - AIBench Scenario
    - Edge AIBench
  - AIBench Training
    - **HPC AI500**
  - AIBench Inference
    - AIoTBench
- Conclusion

HPC AI500 Benchmarking Methodology

- The criteria for choosing the workloads
  - Repeatability
  - Representativeness and Affordability
  - Scalability

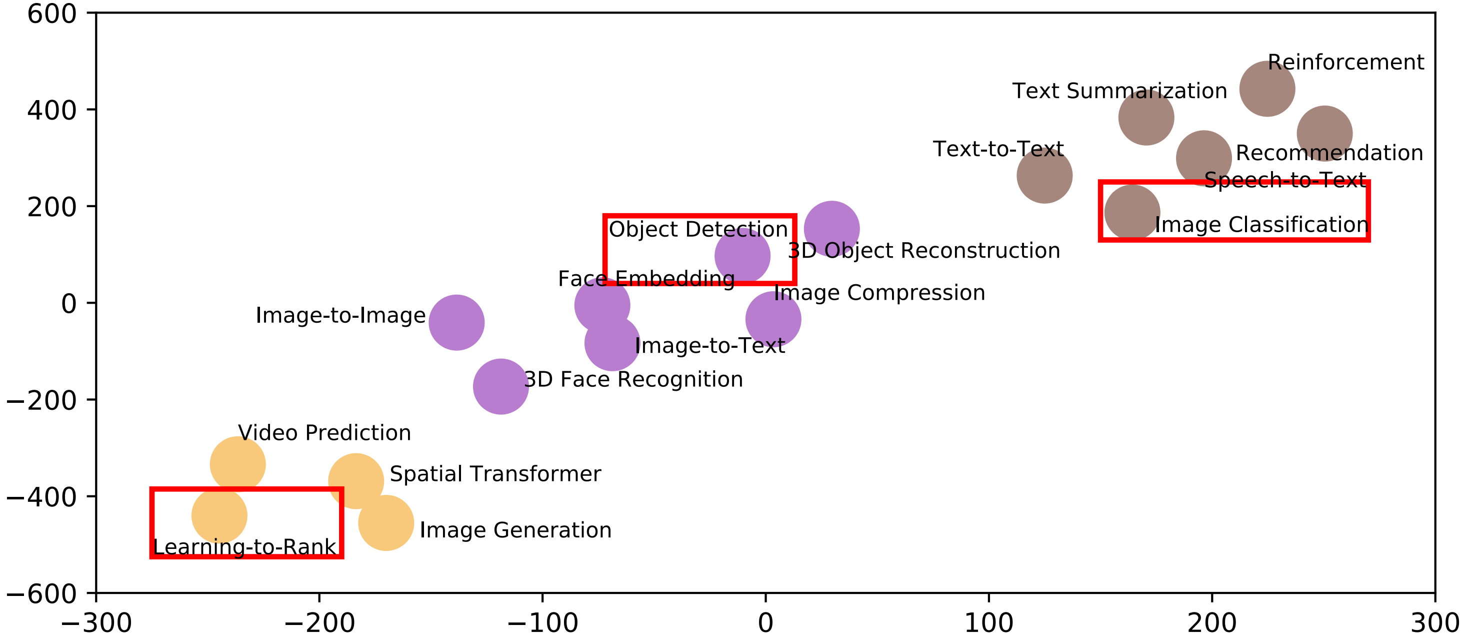
BenchCouncil

Repeatability: Randomness of Workloads

No.	Component Benchmark	Time Per Epoch (second)	Total Time (hour)	Variation	Repeat Times
TrC1	Image Classification	4440	76.25	1.12%	5
TrC2	Image Generation	3935.75	N/A	N/A	N/A
TrC3	Text-to-Text translation	64.83	1.72	9.38%	6
TrC4	Image-to-Text	845.02	10.21	23.53%	5
TrC5	Image-to-Image	251.67	N/A	N/A	N/A
TrC6	Speech Recognition	14326.86	42.78	12.08%	4
TrC7	Face Embedding	214.73	3.43	5.73%	8
TrC8	3D Face Recognition	36.99	12.02	38.46%	4
TrC9	Object Detection	1859.96	2.06	0	10
TrC10	Recommendation	36.72	0.16	9.95%	5
TrC11	Video Prediction	24.99	2.11	11.83%	4
TrC12	Image Compression	763.44	5.67	22.49%	4
TrC13	3D Object Reconstruction	28.41	0.38	16.07%	4
TrC14	Text Summarization	1923.33	6.41	24.72%	5
TrC15	Spatial Transformer	6.38	0.06	7.29%	4
TrC16	Learning-to-Rank	60.1	0.14	1.90%	4
TrC17	Neural Architecture Search	932.79	7.47	6.15%	6

Representativeness and Affordability

- Using K-Means to cluster all seventeen benchmarks based on system behavior metrics (system occupancy、IPC、load、store、dram utilization)



- The selected workloads have low randomness and good repeatability
  - Image Classification, Object Detection, and Learning to Rank

Scalability

- AIBench subset computation comparison (Single training batch).

Workloads	Computation （FLOPs）
Image Classification	23 G
Object Detection	691 G
Learning to Rank	0.08 M

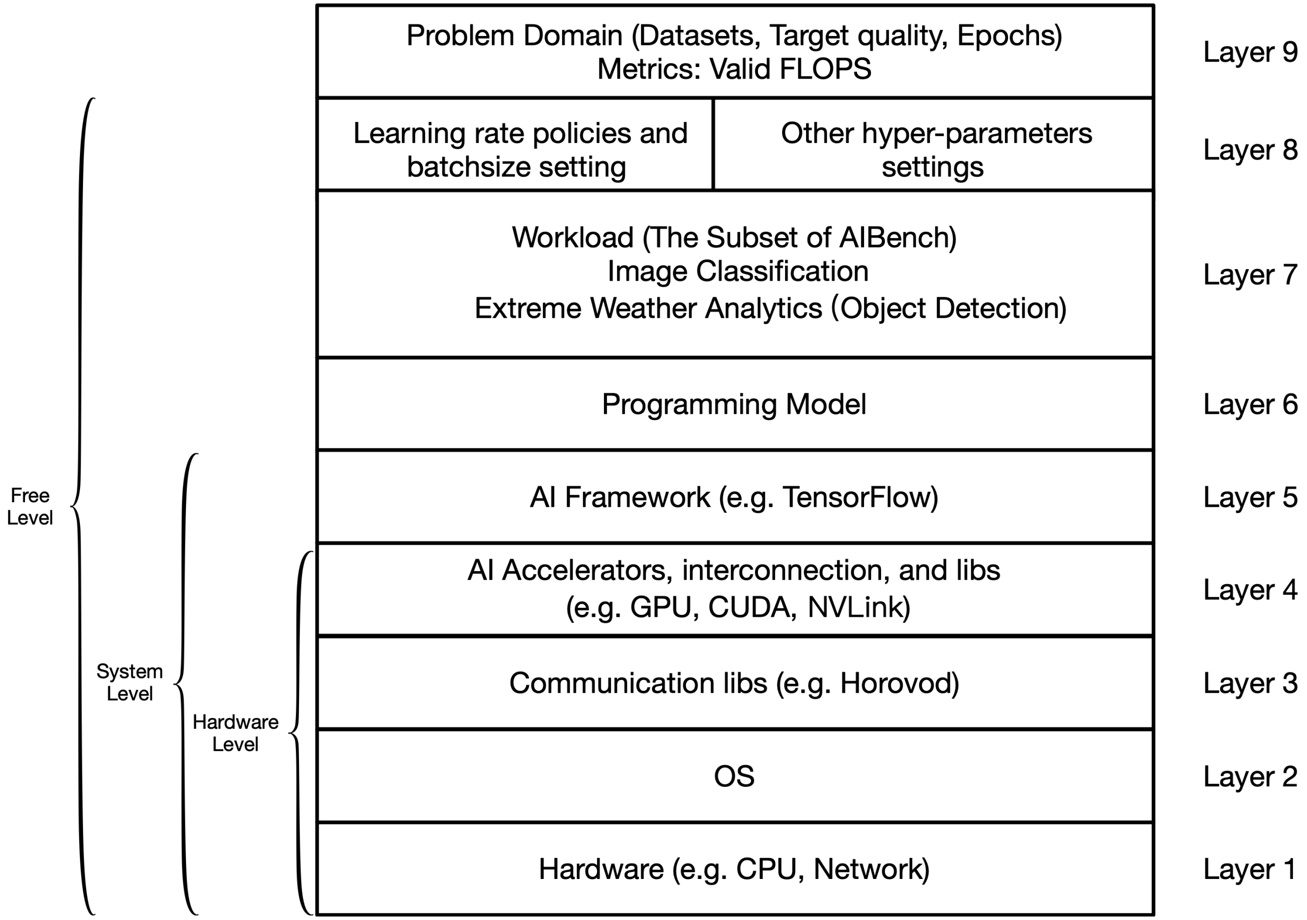
Image Classification and **Object Detection** meet the scalability requirement and are chosen as two typical workloads for HPC AI benchmarking.

Tasks, Dataset, and Model of HPC AI500

- Tasks
  - Extreme Weather Analysis: detect the patterns of extreme weather, essentially Object Detction. The application that wins Gordon Bell Prize.
  - Image Classification：ResNet50/ImageNet is a de facto benchmark for optimizing HPC AI systems.
- Dataset
  - The extreme weather dataset: 16 channels, 768*1052, 2 TB
  - ImageNet 2012: 3 channels, 256*256, 136 GB
- Model
  - Faster-RCNN
  - ResNet-50 V1.5

HPC AI500 Hierarchical Benchmarking Rules

- HPC AI500 defines a comprehensive benchmarking methodology based on nine-layers system abstraction, divided into the following three level: hardware level, system level, and free level.



Metric

- Using VFLOPS to unify the computation and model quality.

$$VFLOPS = FLOPS \times \left(\frac{achieved\_quality}{target\_quality}\right)^n$$

achieved_quality refers to the achieved quality in the evaluation.

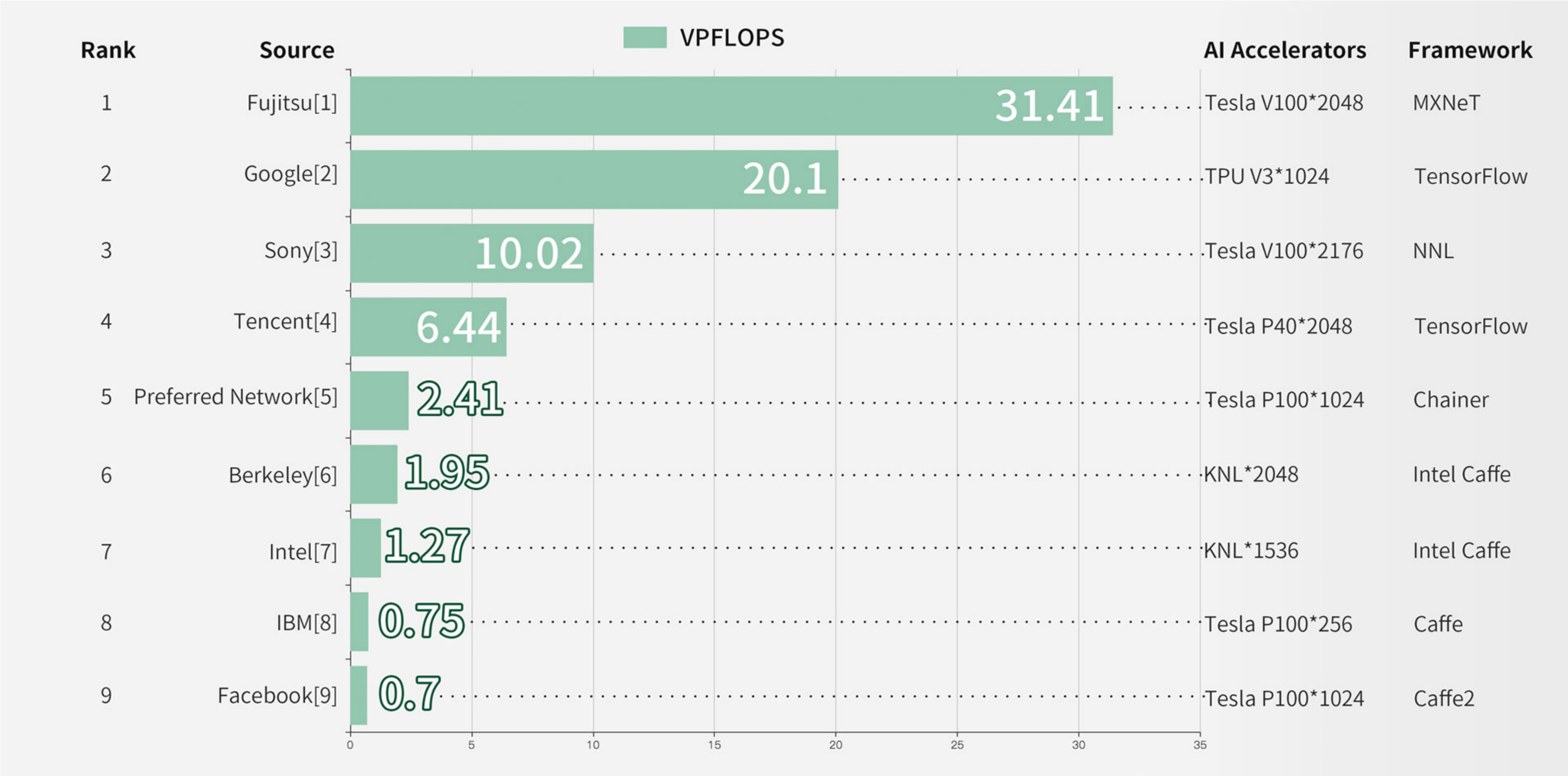
target_quality refers to the target quality defined in HPC AI500 problem domain.

The value of n is a positive integer, which is used to define the sensitivity to the model quality. The higher the number of n, the more loss of quality drop. As EWA (Object Detection) has much more stringent quality requirement than that of Image Classification. We set n as 10 for EWA and 5 for Image Classification by default.

Bench Council

HPC System Ranking

HPC AI500 Image Classification, Free Level



For more details about benchmarking rules, metrics, and performance data:
https://www.benchcouncil.org/ranking.html

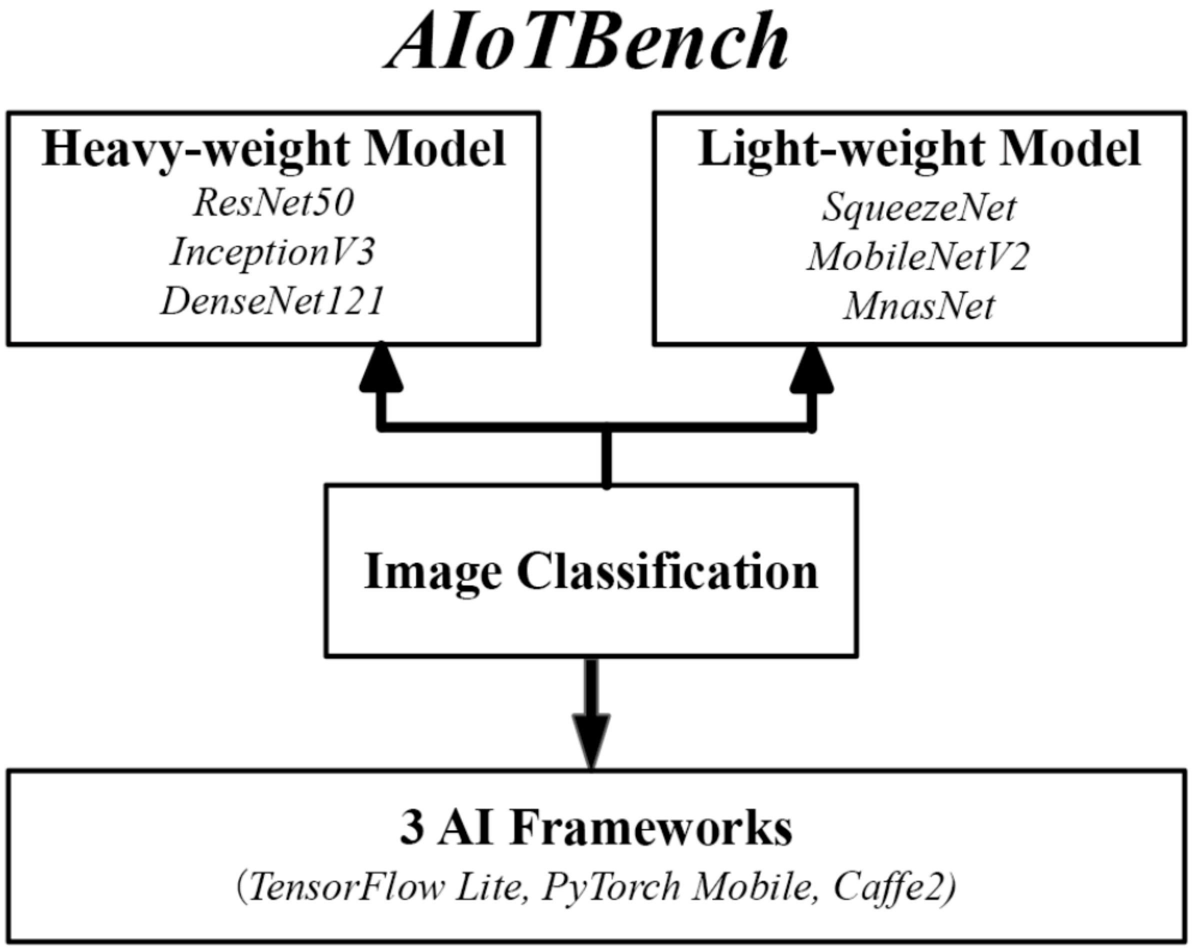
Overview

- Challenges
- Related Work
- **AIBench**
  - AIBench Scenario
    - Edge AIBench
  - AIBench Training
    - HPC AI500
  - **AIBench Inference**
    - AIoTBench
- Conclusion

AIBench Inference

- 17 Inference Workloads (CNN、ResNet、LSTM、GRU、Attention, etc.)
  - **Text processing (5)**
    - Text-to-Text, Text summarization, Learning to Rank, Recommendation, Neural Architecture Search
  - **Image processing (8)**
    - Image Classification, Image Generation, Image-to-Text, Image-to-Image, Face Embedding, Object Detection, Image Compression, Spatial Transformer
  - **Audio processing (1)**
    - Speech Recognition
  - **Video processing (1)**
    - Video Prediction
  - **3D data processing (2)**
    - 3D Face Recognition, 3D Object Reconstruction

AIoTBench Overview



Conclusion

- **Scenario AI benchmarking is needed !**
- **BenchCouncil AIBench (**https://www.benchcouncil.org/aibench.html**)**
  - *Scenario, Training, Inference, and Micro Benchmarks across Datacenter, HPC, IoT, Edge*
  - *Scenario-distilling benchmarking methodology*
    - considering different benchmarking requirements
      - *Scenario benchmarks*
        - **The first AI benchmark** that provides real-world scenario modelling
        - E.g., the complete use cases of autonomous driving scenario in edge computing
        - For overall system evaluation
      - *Component benchmarks*
        - Comprehensive workload behaviors
        - Algorithm/System/Micro-architectural Characteristics
        - Providing component subset for ranking
        - Fairness, affordability, representativeness
      - *Micro benchmarks*
        - Hotspot functions and code optimizations

Bench Council

Conclusion (Cont')

- *If you feel interested in BenchCouncil or AIBench, you are very welcome to join us !*

Bench Council

References

- **AIBench** (https://www.benchcouncil.org/aibench.html)
  - *Scenario-distilling AI Benchmarking*
    - *https://arxiv.org/abs/2005.03459*
  - *AIBench Training: Balanced Industry-Standard AI Training Benchmarking*
    - *Accepted by ISPASS 2021*
- **HPC AI500** (https://www.benchcouncil.org/HPCAI500/index.html)
  - *HPC AI500: The Methodology, Tools, Roofline Performance Models, and Metrics for Benchmarking HPC AI Systems*
    - *https://www.benchcouncil.org/file/HPC_AI500TR.pdf*
- **Edge AIBench** (https://www.benchcouncil.org/EdgeAIBench/index.html)
  - *Edge AIBench: towards comprehensive end-to-end edge computing benchmarking.*
    - *https://arxiv.org/pdf/1908.01924.pdf*
- **AIoTBench** (*http://www.benchcouncil.org/AIoTBench/index.html*)
  - *AIoTBench: Towards Comprehensive Benchmarking Mobile and Embedded device Intelligence*
    - *https://www.benchcouncil.org/AIoTBench/files/AIoTBench-Bench18.pdf*

Bench Council

Download

- AIBench for datacenter AI benchmarking
  - AIBench Micro Benchmark
    - http://www.benchcouncil.org/benchhub/AIBench/DC_AIBench_Micro/
  - AIBench Component Benchmark
    - http://www.benchcouncil.org/benchhub/AIBench/DC_AIBench_Component/
  - AIBench Scenario Benchmark
    - http://www.benchcouncil.org/benchhub/AIBench/AIBench_Application_Benchmark/
    - http://www.benchcouncil.org/benchhub/AIBench/AIBench_DCMIX
  - AIBench Framework
    - http://www.benchcouncil.org/benchhub/AIBench/AIBench_Framework/

- Sign in/up BenchHub to get access !

Download (Cont')

- HPC AI500 for benchmarking HPC AI systems
  - http://www.benchcouncil.org/benchhub/hpc-ai500-benchmark
- Edge AIBench for edge computing
  - http://www.benchcouncil.org/benchhub/edge-aibench/
- AIoTBench for IoT
  - http://www.benchcouncil.org/benchhub/aiotbench/
- Sign in/up BenchHub to get access !

Thank You !

Bench Council