

2020 科技部AI跨域交流觀摩會

實現深度學習於產業服務之邊端智慧系統架構與其設計流程

國立交通大學資訊工程學系 陳添福教授



GoEdge.ai

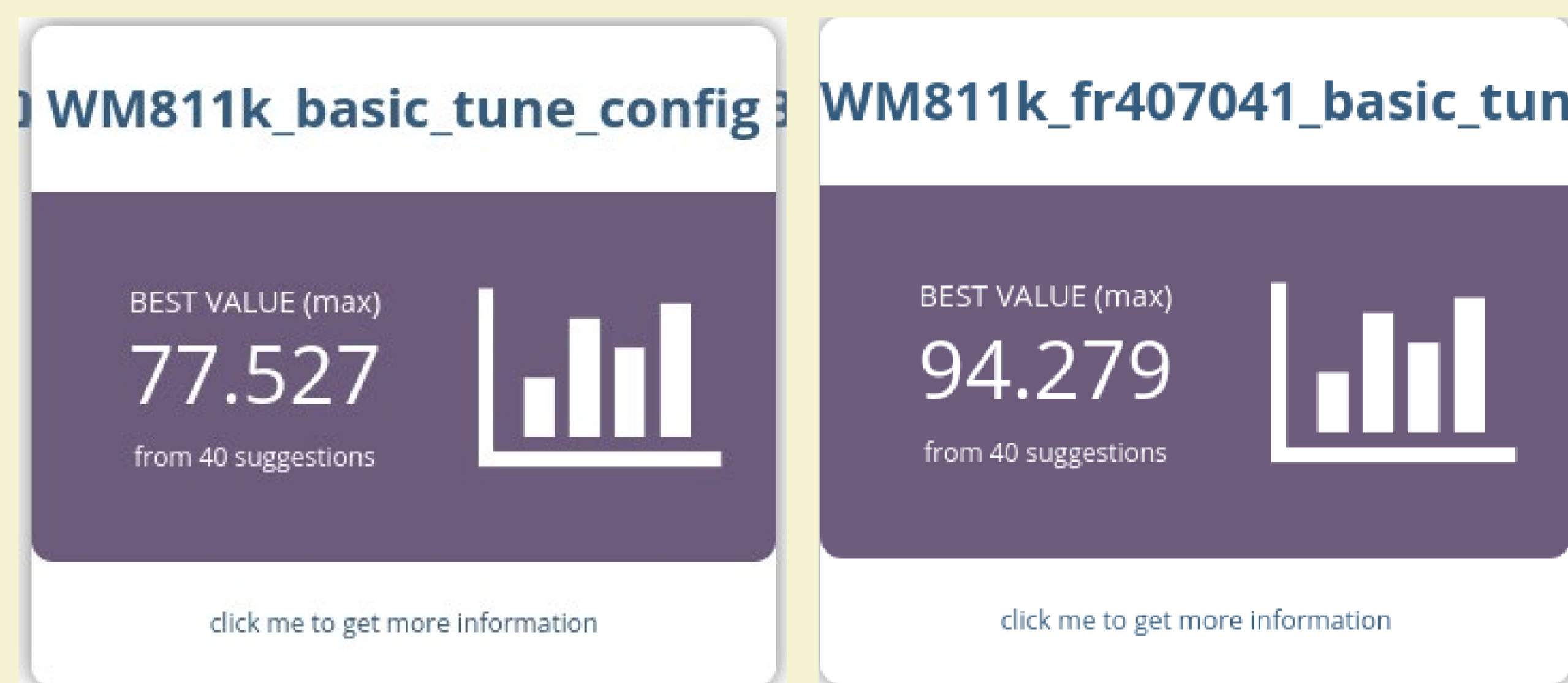
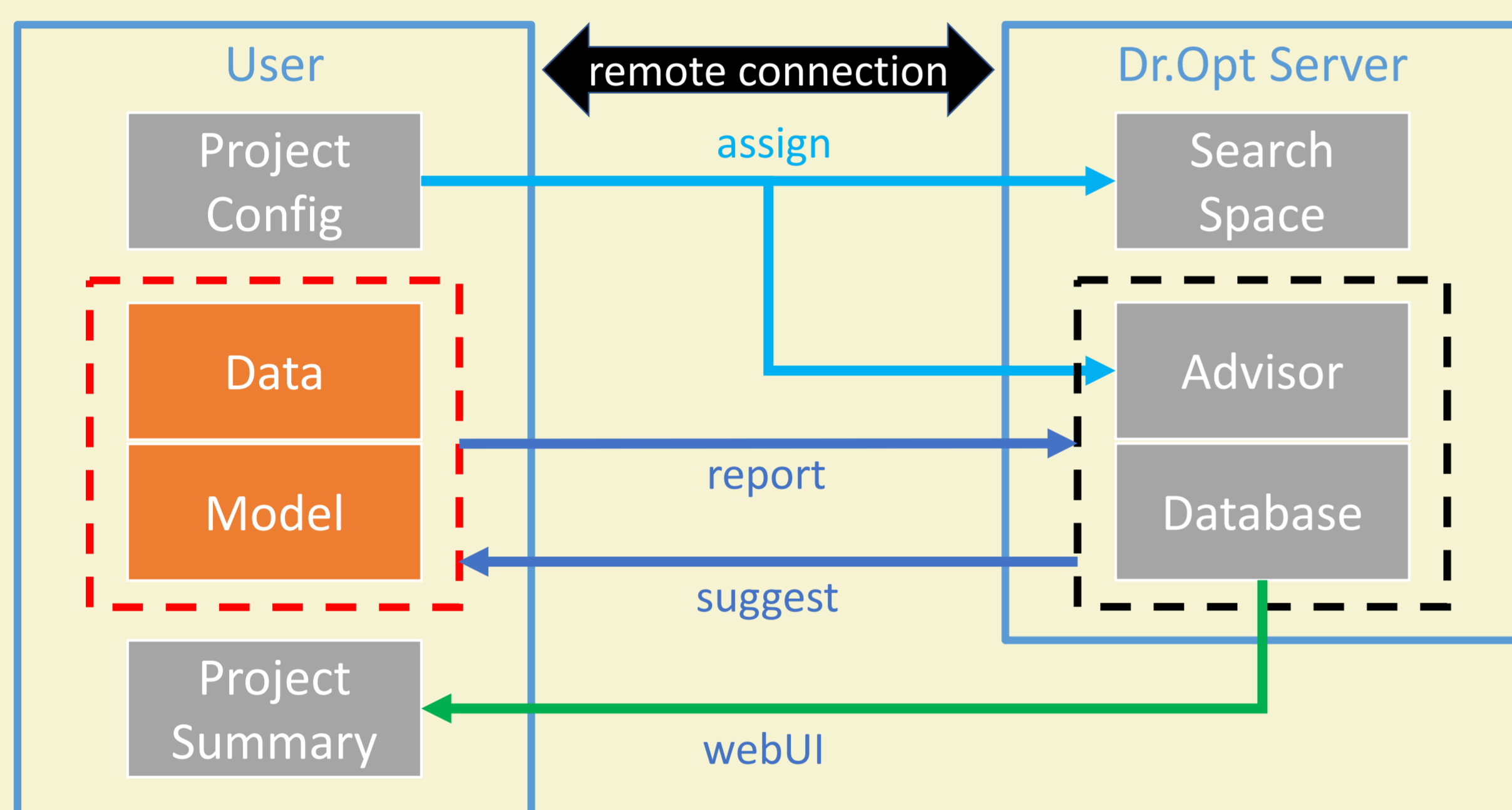
What is Dr.Opt?

- An auto-tuning service that finds
 - hyper-parameter configurations for improved performance
 - manufacture parameters for quality products
- In general, Dr.Opt deals with optimization of a black-box

Guided model accuracy tuning

- Kaggle wafer map dataset
- Basic accuracy: 77.5%
- Guided accuracy: over 94%
 - Search for 7 hyper-parameters: LR, batch size, ...
 - 40 trials

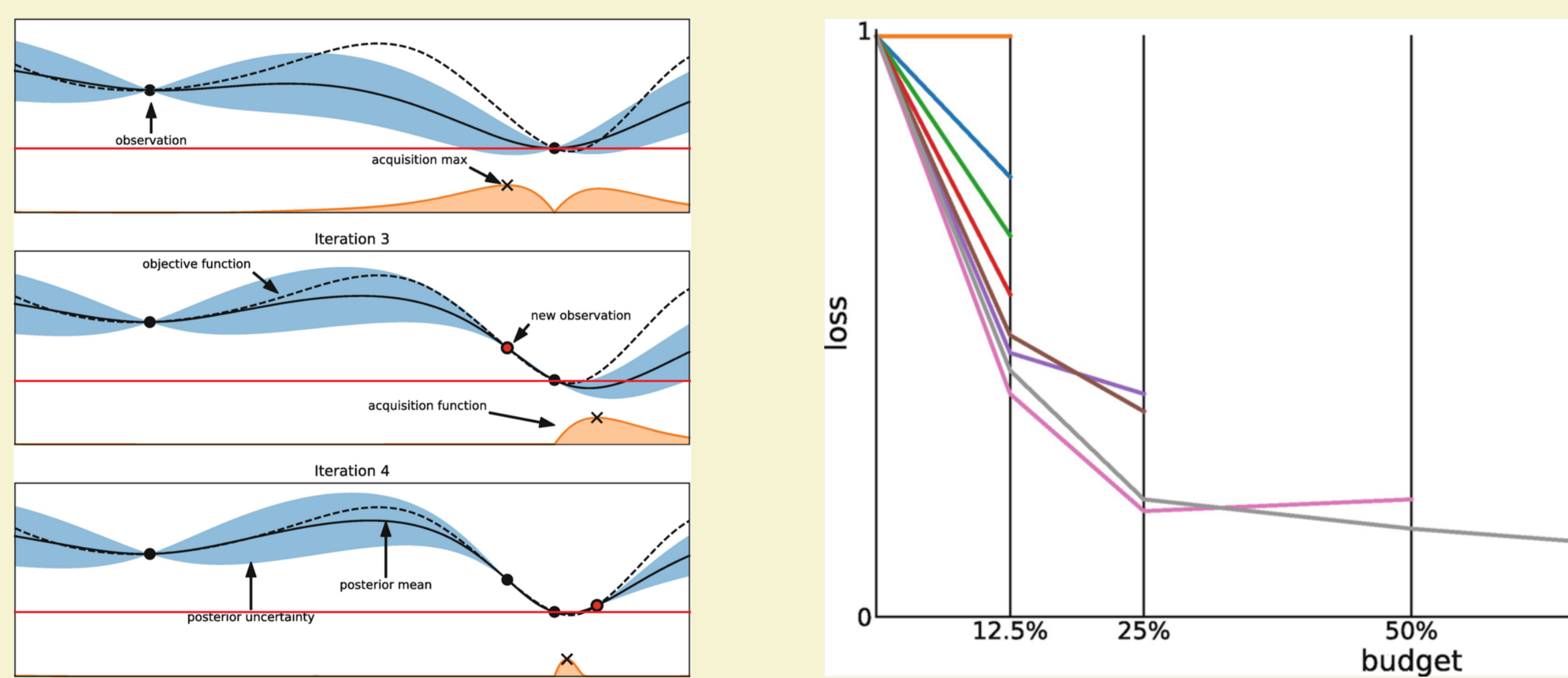
How does Dr.Opt work?



Basic

Guided

How does advisor work?



Math modeling

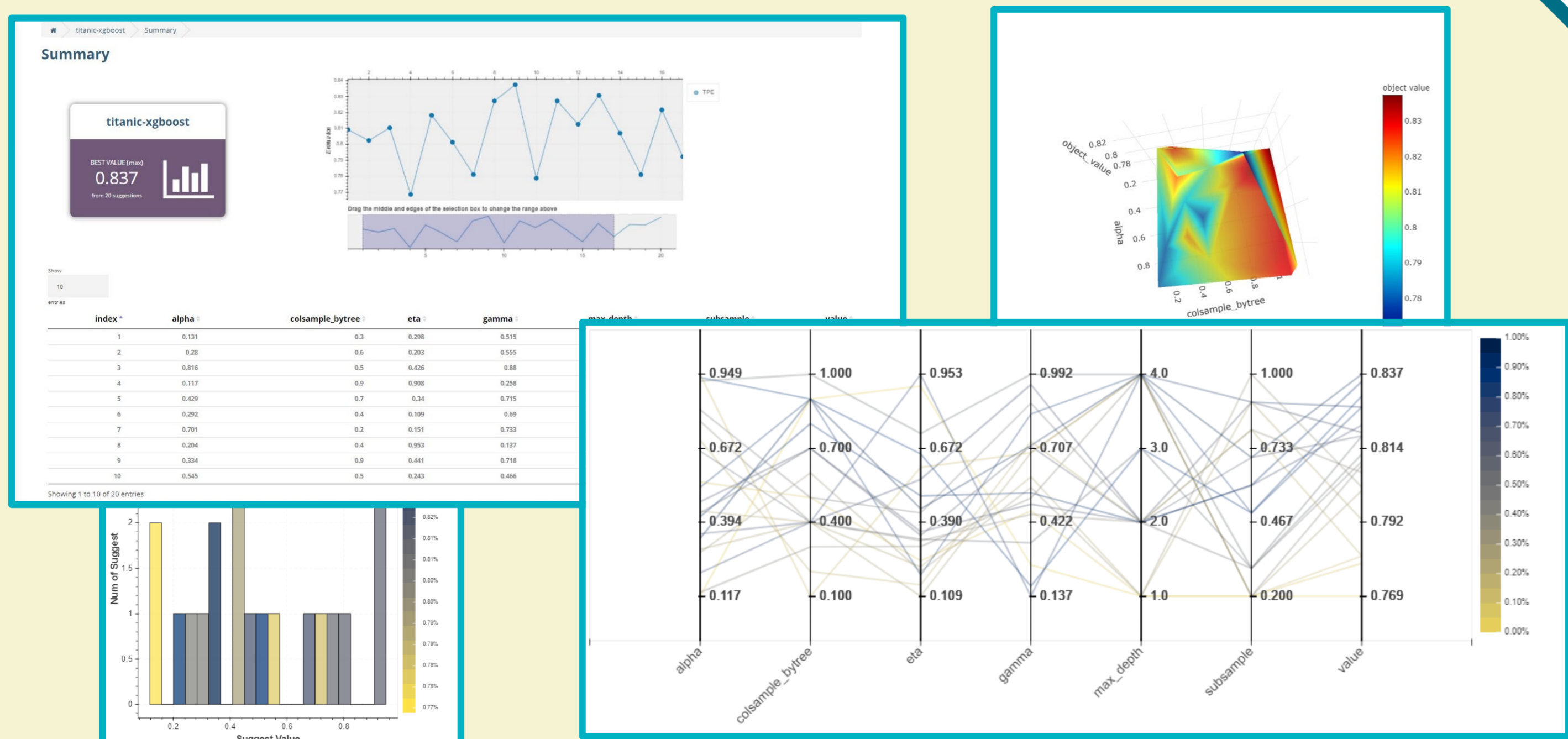
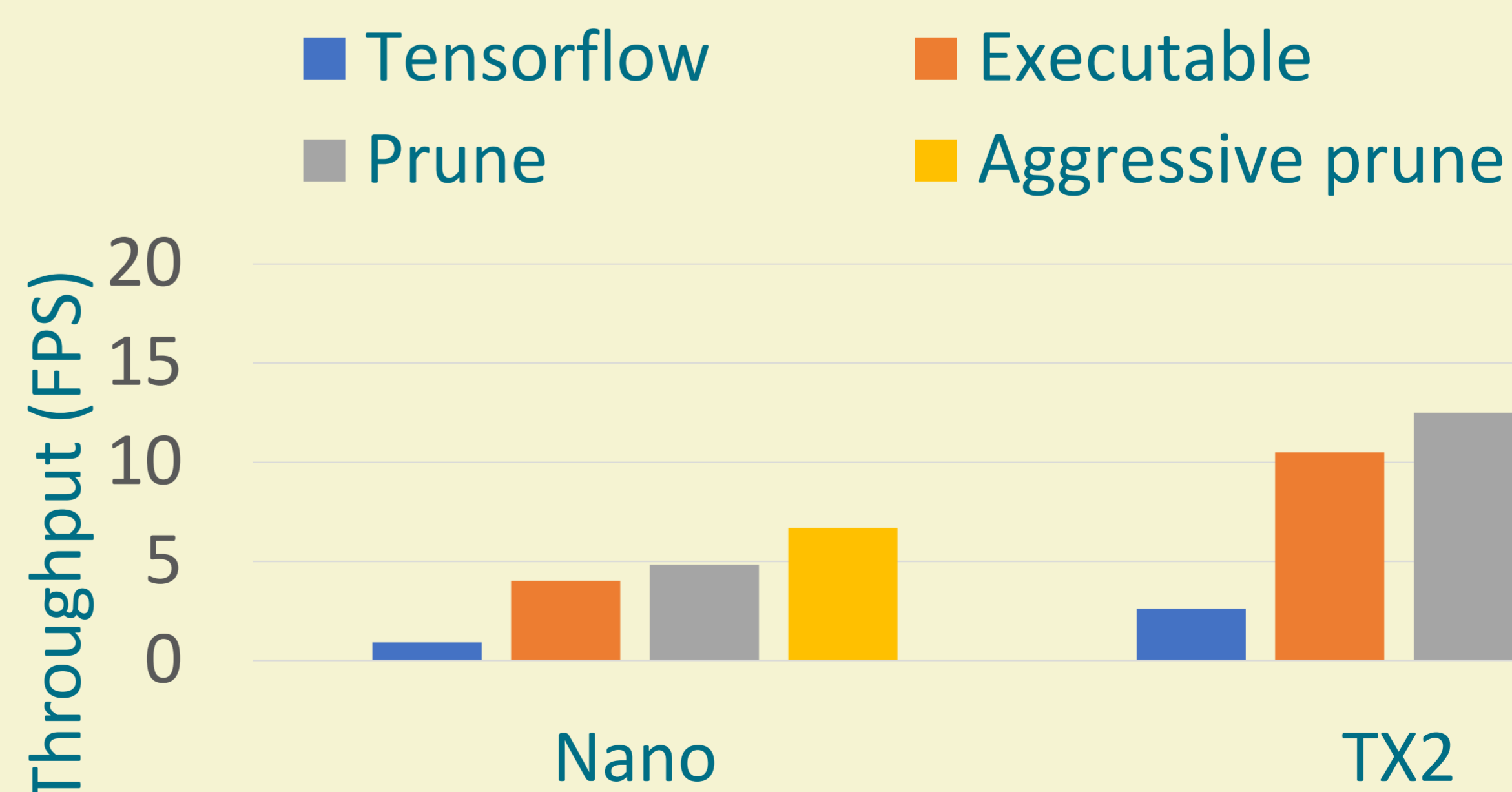
Resource allocation

Guided model inference time tuning

- Model repo. - qqwweee/keras-yolo3
- Our simple dataset
 - About 1K images and 1 category
- Four configurations
 - Tensorflow, Executable(Data independent), 2 level pruned model(Data dependent)

| | Tensorflow/ Executable | Prune | Aggressive prune |
|-------------|---------------------------|------------|------------------|
| Weight Size | 235MB | 132MB(56%) | 52MB(22%) |
| mAP | 68.6% | 66.3% | 65.6% |

| | Tensorflow | Optimized |
|------|------------|-----------|
| Nano | ≅ 4.4GB | 1.5~1.6GB |
| TX2 | ≅ 7.2GB | 2.2~2.3GB |



Dr.Opt Server Dr.Opt GitHub Dr.Opt GitPage Dr.Opt PyPI