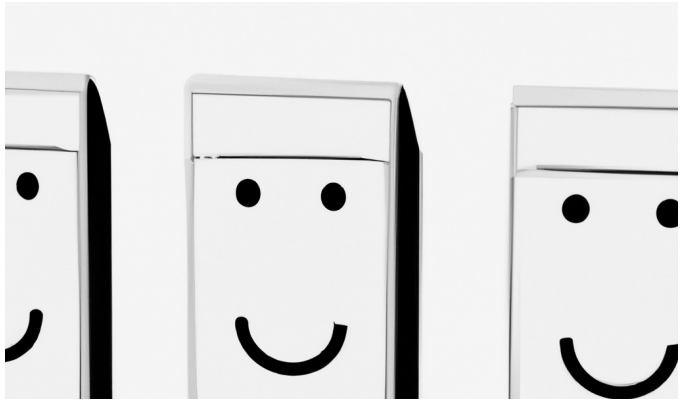


IBPT
Institute for Beam Physics and Technology

KIT
Karlsruhe Institute of Technology

“Clever Switches” Team 4

AI HERO 2 Challenge: Use Case Energy

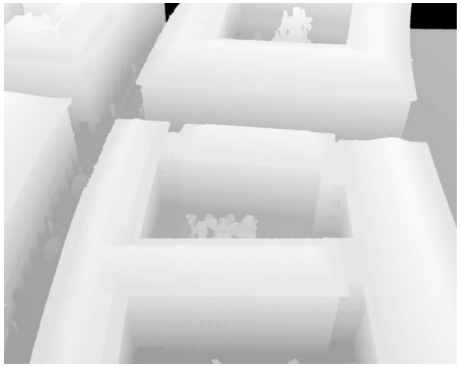
Andrea Santamaria Garcia, Stephan Kötter, Chenran Xu

AI4Accelerators Team @Institute for Beam Physics and Technology (KIT-IBPT)

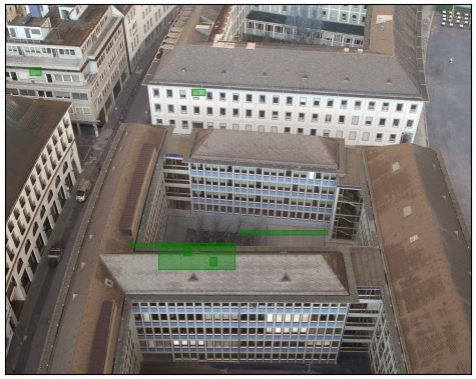
Understanding the data & goal

RGB

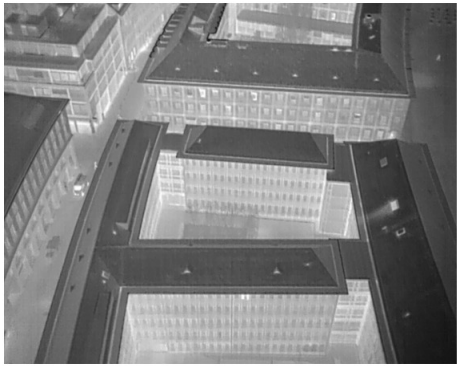
Depth



Boxes

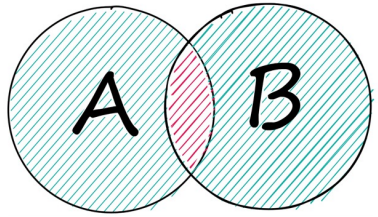


Thermal



Masks

$$\text{Jaccard} = \frac{\text{intersection}(A, B)}{\text{union}(A, B)}$$



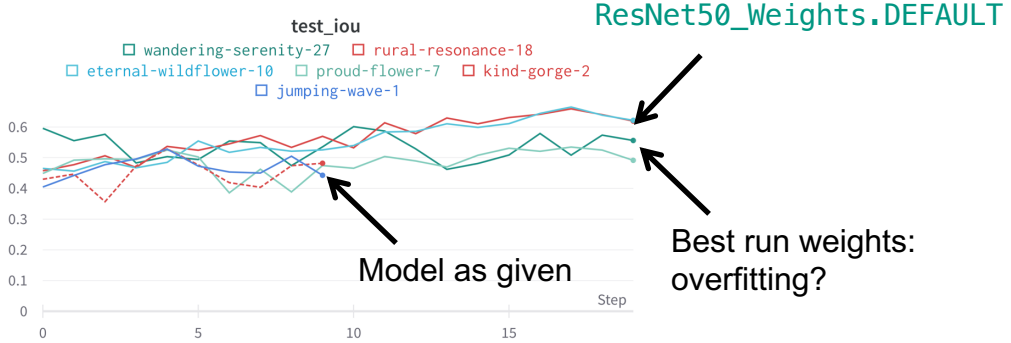
source

Intersection-over-Union or IoU

Andrea Santamaria Garcia, Stephan Kötter, Chenran Xu (KIT IBPT)

Day 1 strategies

- Used preloaded weights for backbone `ResNet50_Weights.DEFAULT`
- We added mixed precision (`autocast`)
  - no real impact (backbone loaded, not training from scratch)
- Normalized the input data to what the model expects
  - it was not being passed
- Using only 1 GPU
- Loading weights of best run
  - probably overfitting
  - didn't load optimizer checkpoint



Tracking with weights & biases ☺

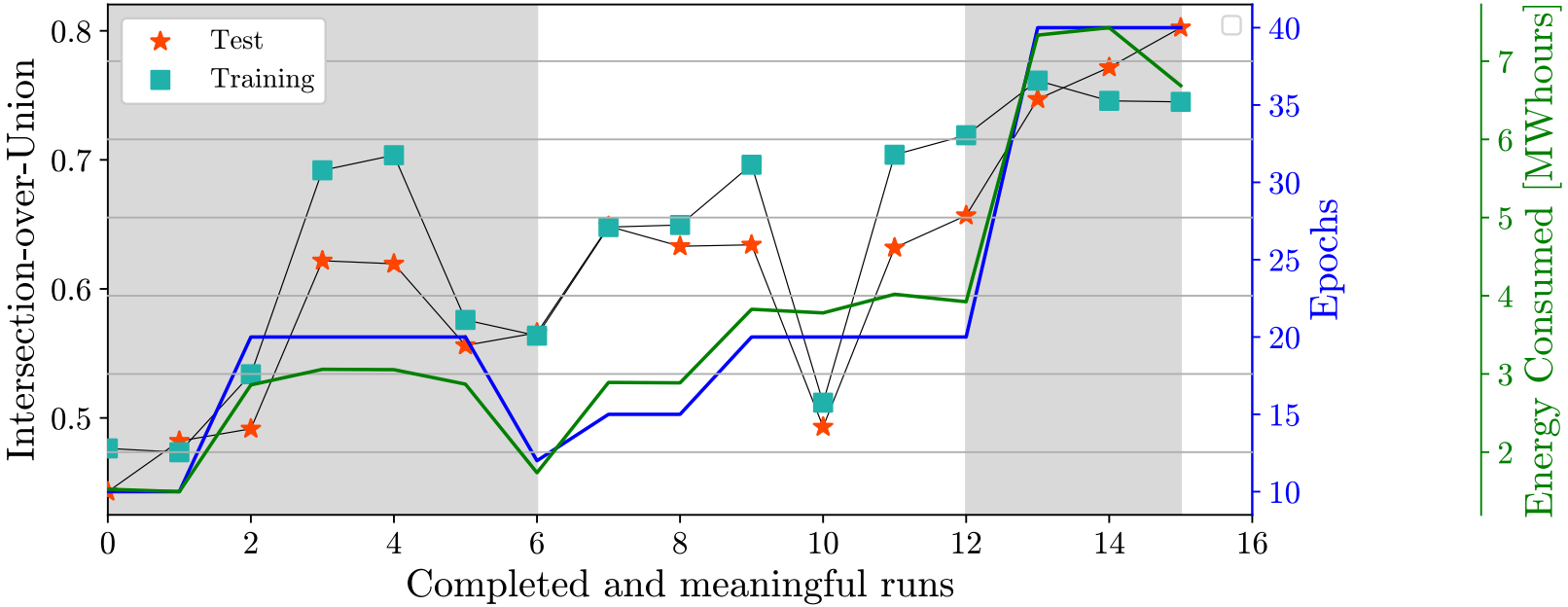
Day 2 strategies

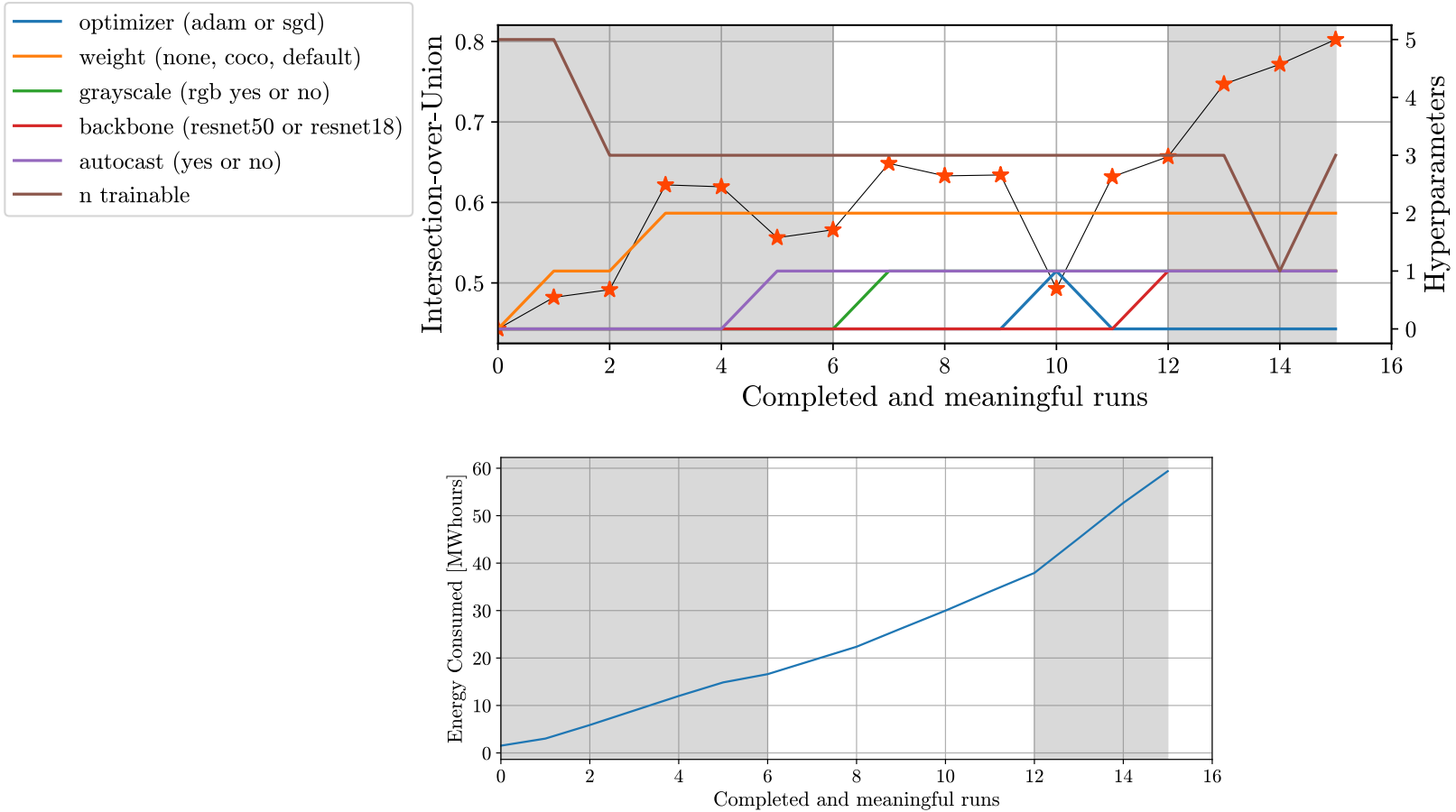
- Parallelized code to run in 4 GPUs with `DistributedDataParallel,DistributedSampler`
  - increased GPU utilization but more time per epoch
  - quite difficult to implement
- Changed backbone to `resnet18`
  - lighter model → faster inference + faster training
- Changed 3 RGB channels for 1 channel in grayscale
  - backbone weights pre-trained expects 3 channels not 5
  - no need to re-train first conv layer
- Realizing we can only use models that return masks
  - we wanted to use semantic segmentation method, no time to adapt
  - stick with original model, tune hyperparameters

Day 3 strategies

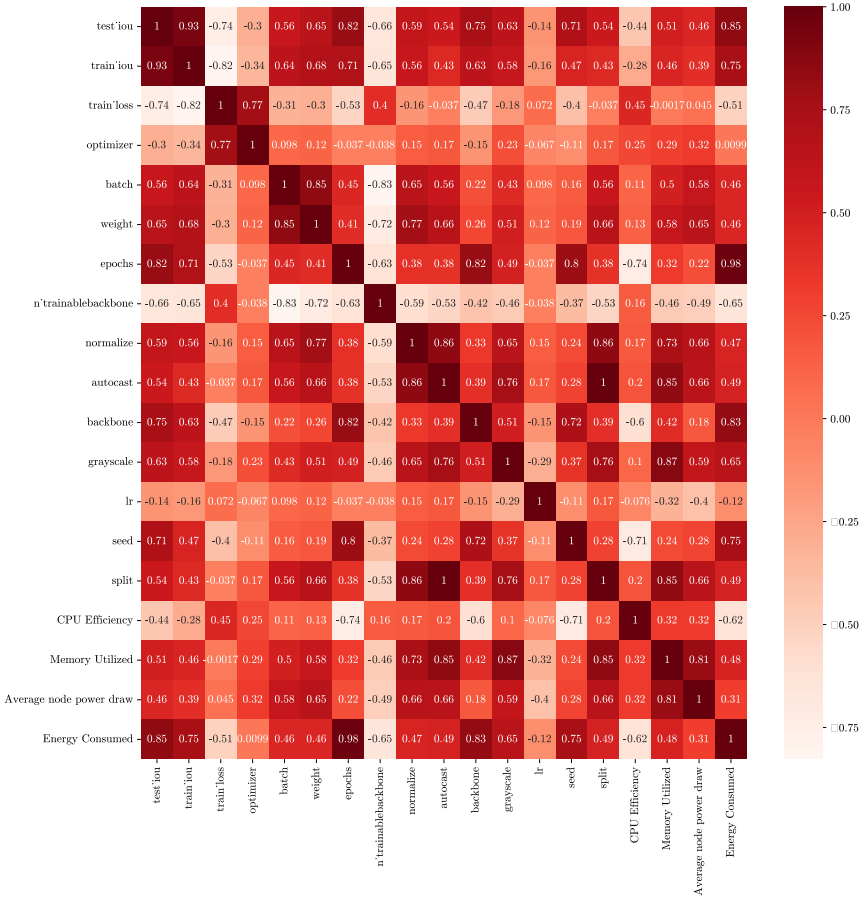
- Reduce number of trainable backbone layers
  - faster training (?)
- Increase number of epochs for training
  - better accuracy
- Increase number of batches for prediction
  - faster inference

Andrea Santamaria Garcia, Stephan Kötter, Chenran Xu (KIT IBPT)


Test
Training
Intersection-over-Union
0.8
0.7
0.6
0.5
0
2
4
6
8
10
12
14
16
Completed and meaningful runs
Epochs
40
35
30
25
20
15
10
Energy Consumed [MWhours]
7
6
5
4
3
2


optimizer (adam or sgd)
weight (none, coco, default)
grayscale (rgb yes or no)
backbone (resnet50 or resnet18)
autocast (yes or no)
n trainable
Intersection-over-Union
Hyperparameters
Completed and meaningful runs
Energy Consumed [MWhours]
Completed and meaningful runs

Trying some feature correlation…



We had a lot of fun ☺