

# Sweep 3

Report written after completing the third sweep.

Jo Falck-Ytter

Created on May 9 | Last edited on May 9

The data is SN-BAS extracted with VideoMAE-V2, early stopping was added to the model but it did not affect the results. Some outliers were explored late in the with one of the last runs being the first to achieve an average mAP of above 0.49. **70 runs**, the same as in earlier sweeps were used. I will prefer looking at the visual data in the graph from this, rather than the correlation and importance metrics. The distance from the best to the worst runs here are bigger than before, which is not great. But one negative outlier makes this appear worse. However, compared to sweep 2 there are more values below 0.45 now than then. Some added values can explain this.

I want to create config based on the results, not use the best run. I think creating a config based on averages will create a more robust model. It will be more fair to compare with. The risk of overfitting also reduces. The expected result will be between 0.46-0.48.

The bottom graph shows the mAP based on the categorical **crop\_ratio**. [0.95,1] and [0.8,0.9] performed worst. [0.8,1] performed better on average, [0.7,1] about the same but with higher tops. The choice is between them, but because the best run used **[0.7,1]** it is fair to pick that.

The **learning\_rate** seems to center somewhere around 0.00003-0.00004. The best run has a value of 0.000038 and the second best run has 0.000028. A fair value to pick would be **0.000035**.

**weight\_decay** is difficult to understand. While higher values seem to perform better, the best run had  $\sim 0.00001$  and the worst had  $\sim 0.000005$ . In sweep 2 weight decay {0.04,  $1e-6$ } got a negative correlation. (Geometric mean = 0.0002). In this run [geometric mean {0.004,  $1e-8$ }= $6.32e-6$ ] got a positive correlation. The plot of weight

decay shows that by itself it matters little. But the mean of the means, sends me to the area at about  $3e-5$  which performs relatively solid. That is interesting because it is almost the same as the learning rate at **0.00003**.

**droppath** appears to be centered around **0.12**, so that will be picked.

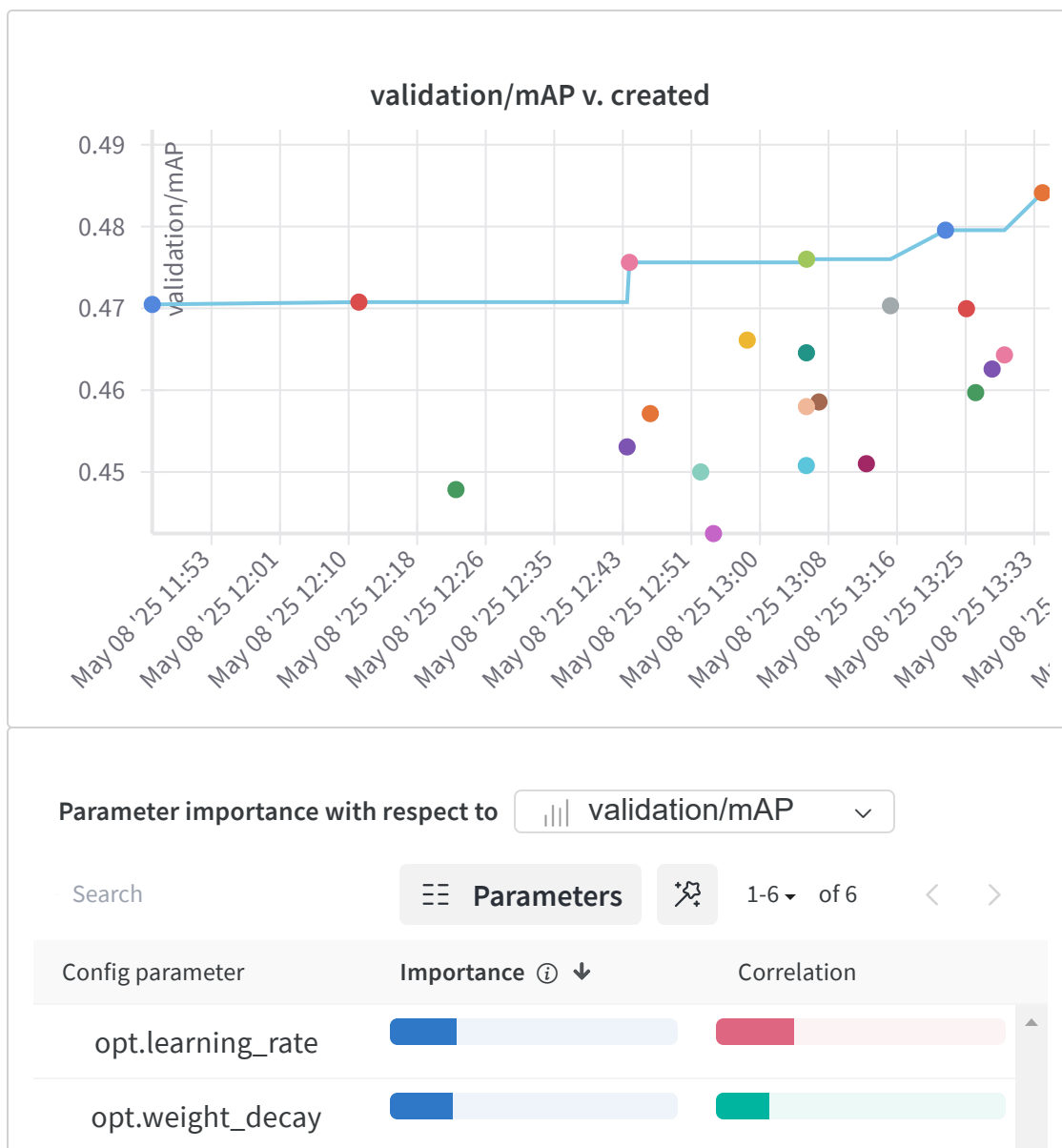
**trunc\_thresh** is centered at 0.5-0.7. With the two best runs being in each end of the section. A value of **0.6** is selected.

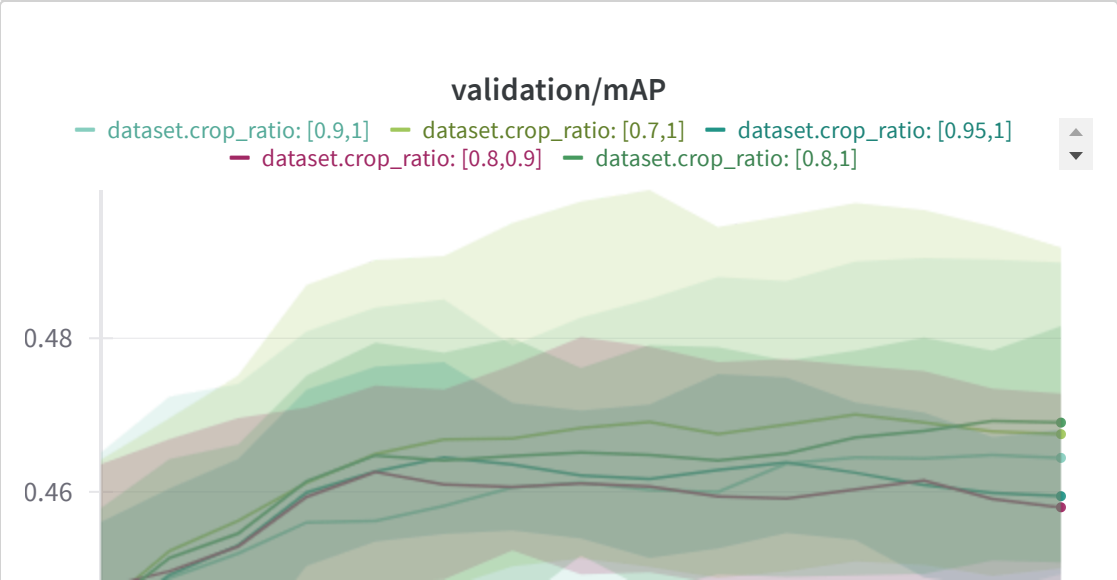
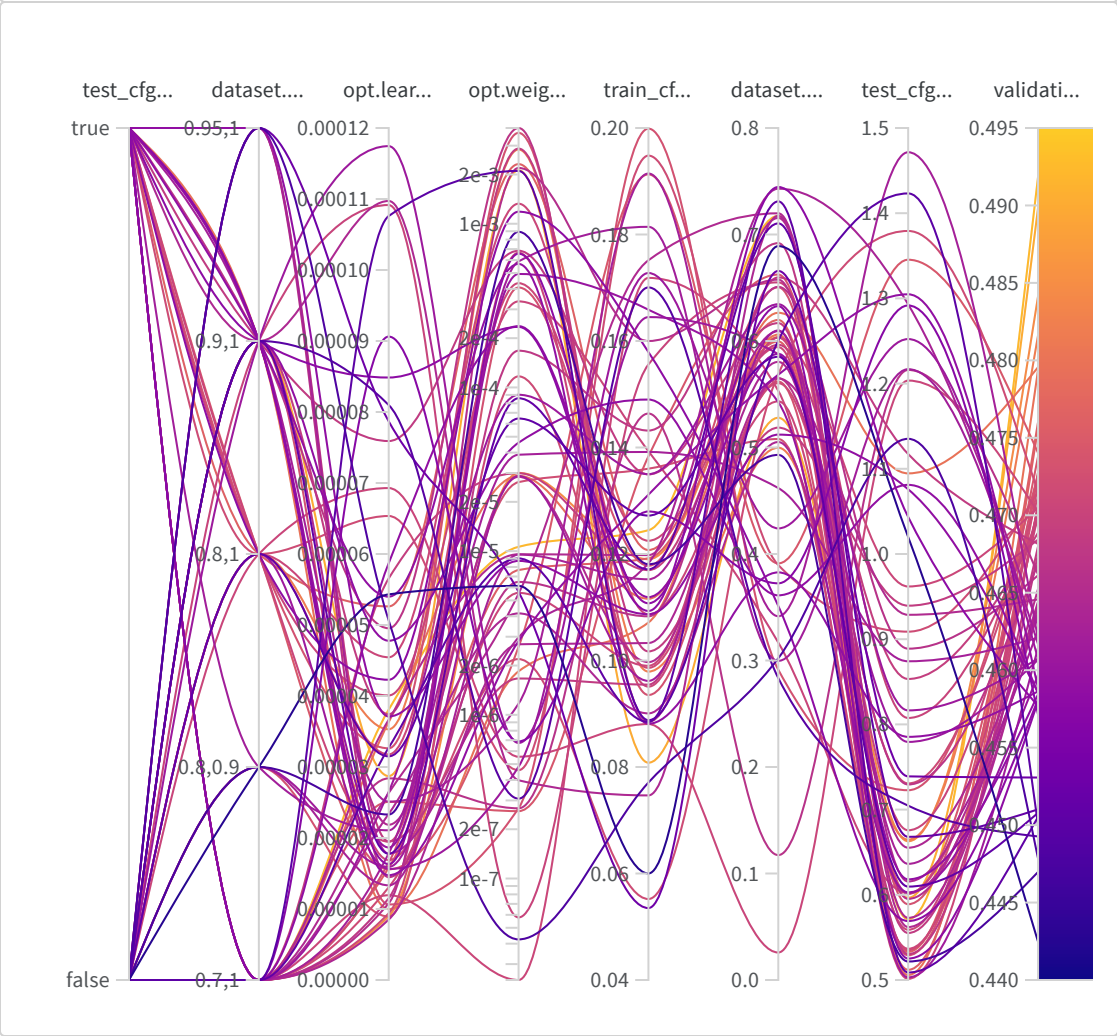
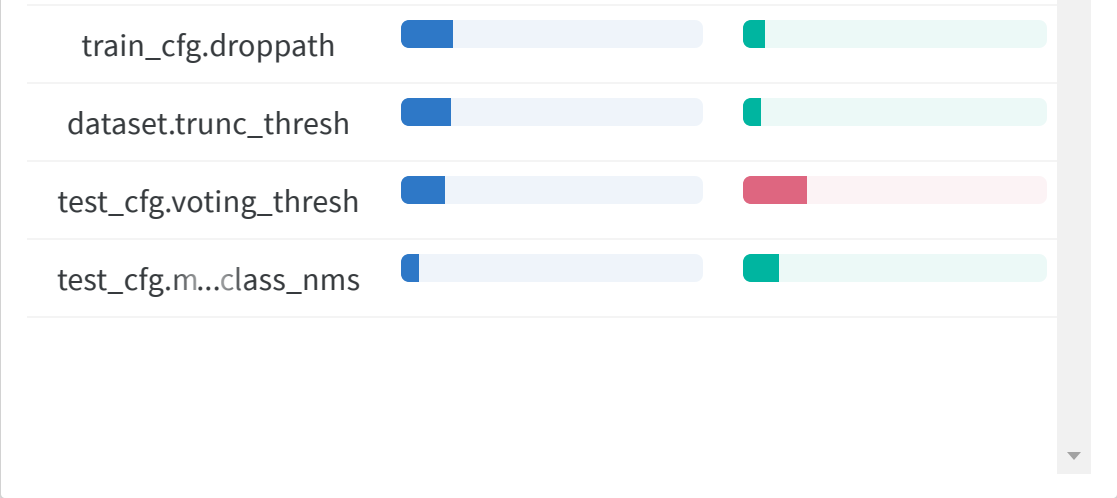
**voting\_thresh** performs best at **0.58**.

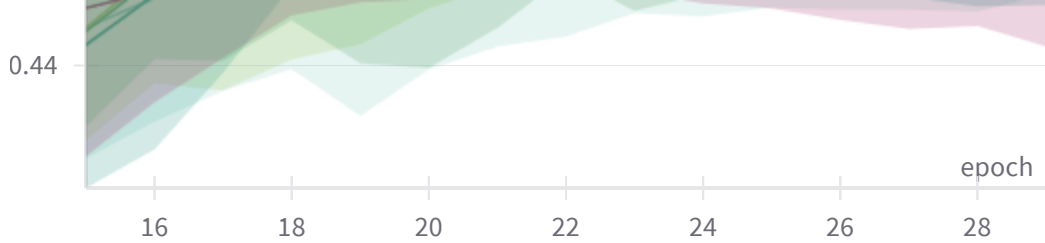
**muticlass\_nms** is set to **true** as that is the default value and it has little impact on the result.

Below are some plots and the configuration of the sweep.

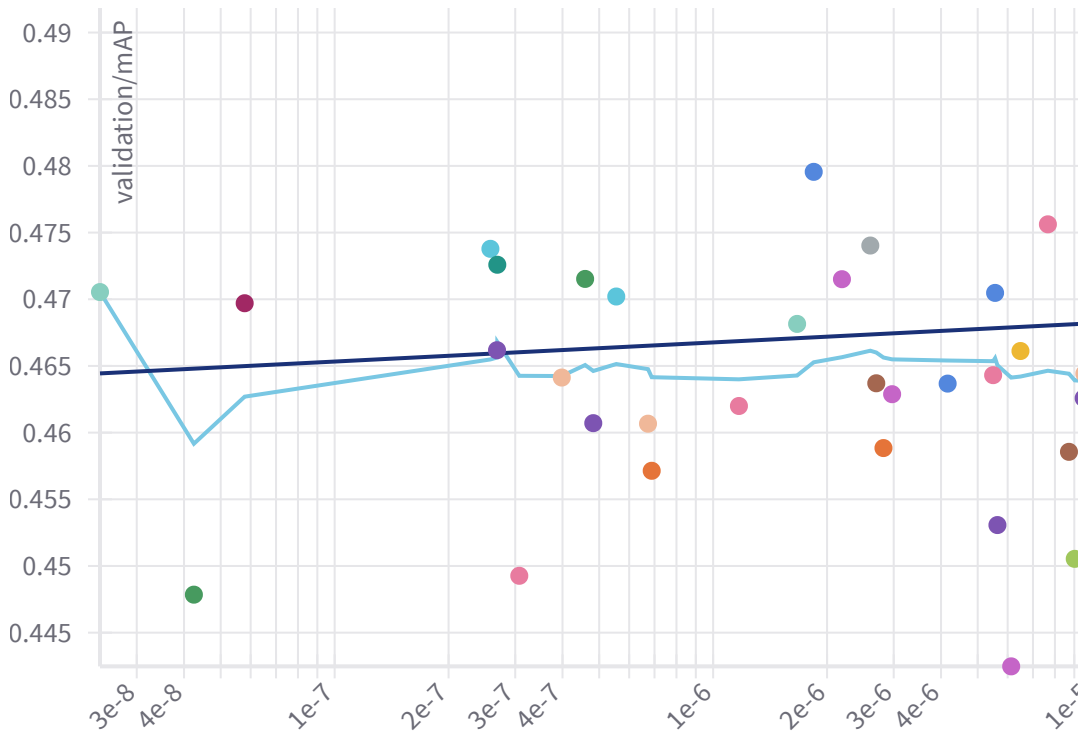
## ▼ Section 1







validation/mAP v. opt.weight\_decay



command:

- python
- train\_eval.py
- configs/mamba\_custom.yaml
- --wandb
- --wandb\_project
- video-mamba-suite
- --wandb\_entity
- jofalck-ntnu

entity: jofalck-ntnu

method: bayes

metric:

goal: maximize

name: validation/mAP

parameters:

dataset.crop\_ratio:

```
distribution: categorical
values:
  - 0.7
  - 1
  - 0.8
  - 0.9
  - 0.8
  - 1
  - 0.9
  - 1
  - 0.95
  - 1

dataset.trunc_thresh:
  distribution: uniform
  max: 0.75
  min: 0.025

opt.learning_rate:
  distribution: uniform
  max: 0.000125
  min: 8e-06

opt.weight_decay:
  distribution: log_uniform_values
  max: 0.004
  min: 1e-08

test_cfg.multiclass_nms:
  distribution: categorical
  values:
    - true
    - false

test_cfg.voting_thresh:
  distribution: uniform
  max: 1.5
  min: 0.5

train_cfg.droppath:
  distribution: uniform
  max: 0.2
  min: 0.05

program: train_eval.py
project: video-mamba-suite
```