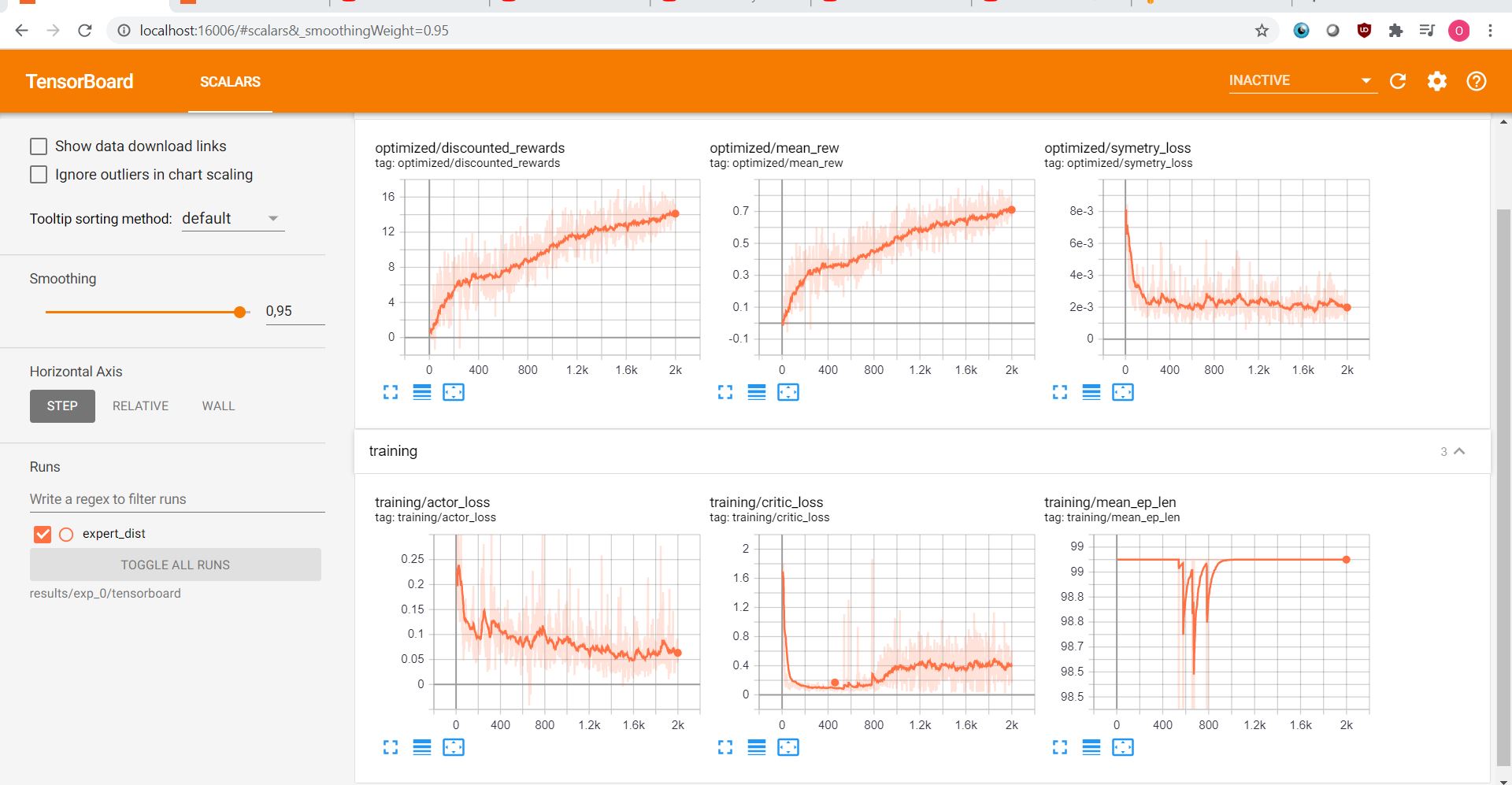
Baseline

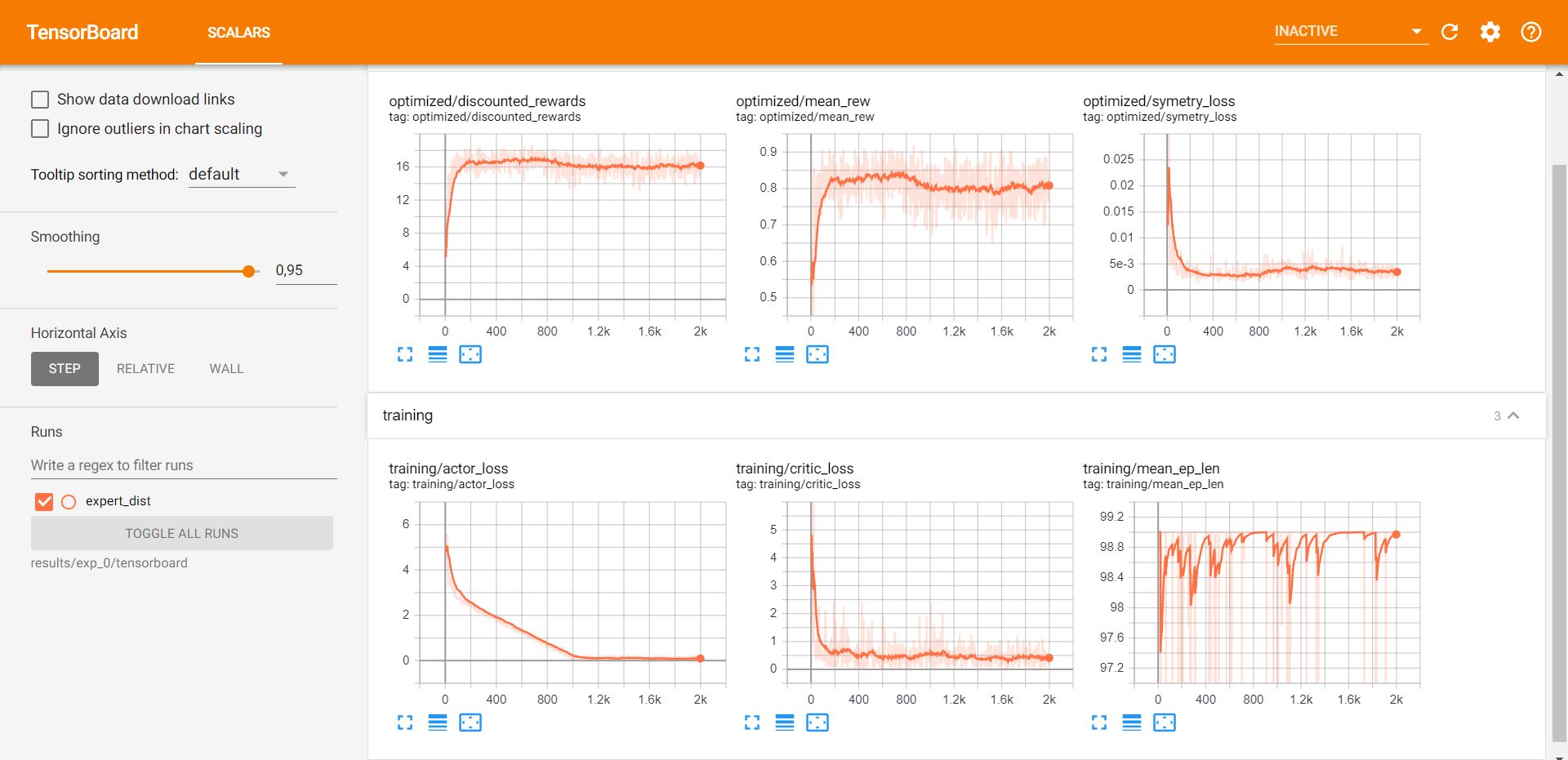
Env reward += square to the primitive action

* Worse than baseline



Actor gradient += neglogp to primitive action

* Actually kinda good



Actor gradient += neglogp to primitive action / 20

* Da best

