Discussion of small batch training

ray/tune/custom_metrics/collide_gate_rate_mean
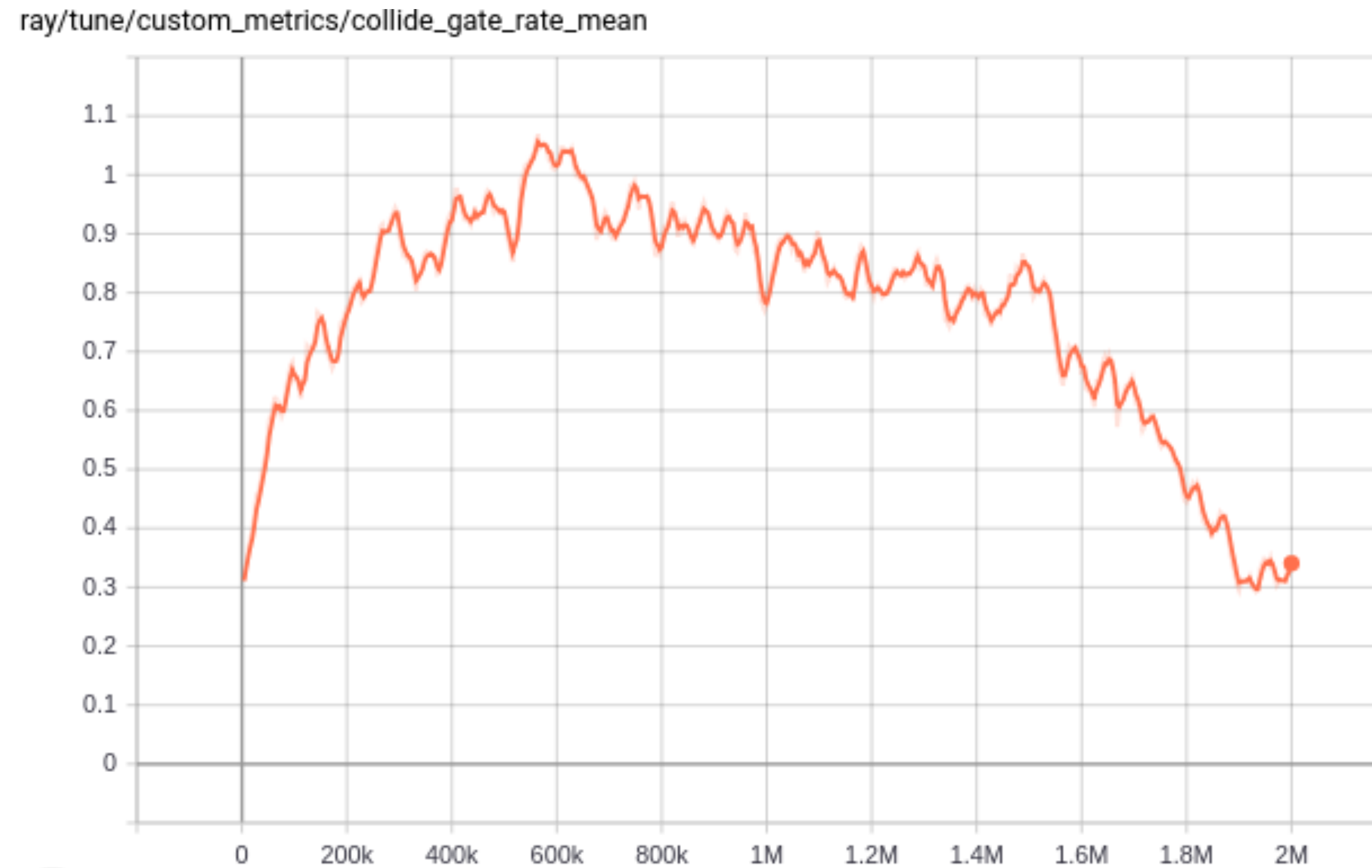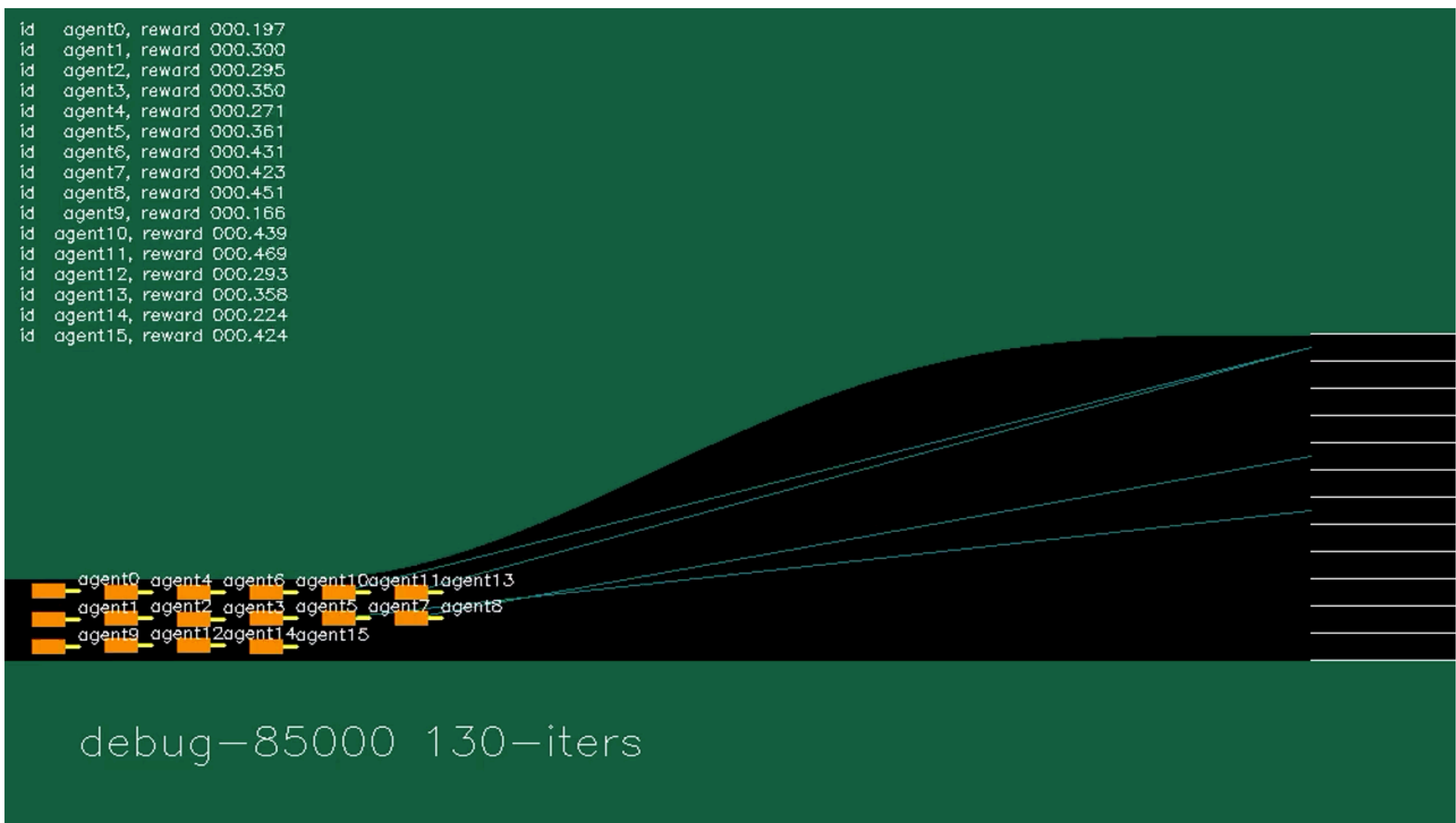
- Gate Collision Rate increase followed by decreasing.

- Agents first learn how to achieve tollgate area, then learn how to avoid gate collision.

- *Note that the values is wrong in this figure. It should times (9/16), due to a tiny bug which has been fixed.*

Performance of large batch training

Waiting and observing behaviors emerge! But still easy to collide.

Metric	Value
iteration	132
— **performance** —	
agent collision	0.0121
boundary col. rate	0
gate collision rate	0
pass rate	0.98796
success rate	0.16492
policy loss	-0.00010
value loss	0.68666
episode reward	1653.808
single reward	103.3630
episode length	194.8668
— **configuration** —	
num workers	2
env per worker	16
cpu per worker	26
sample batch size	200
train batch size	85000