ACT Tuning Tips

- Chunk size is the most important param to tune when applying ACT to a new environment. One chunk should correspond to ~1 secs wall-clock robot motion.
- High KL weight (10 or 100), or train **without** CVAE encoder.
- Consider removing temporal_agg and increase query frequency [here](#) to be the same as your chunk size. I.e. each chunk is executed fully.
- train for **very long** (well after things plateaus, see picture)
- Try to increase batch size as much as possible, and increase lr accordingly. E.g. batch size 64 with learning rate 5e-5 versus batch size 8 and learning rate 1e-5
- Have separate backbones for each camera (requires changing the code, see [this commit](#))
- L1 loss > L2 loss (not precise enough)
- Abs position control > delta/velocity control (harder to recover)
- Try multiple checkpoints

For real-world experiments:
- Train for even longer (5k - 8k steps, especially if multi-camera)
- If inference is too slow -> robot moving slowly: disable temporal_agg and increase query frequency [here](#). We tried as high as 20.

Example loss curve (L1)