Generated distributions when increasing pre-training epochs

Distribution after 50 pretraining epochs
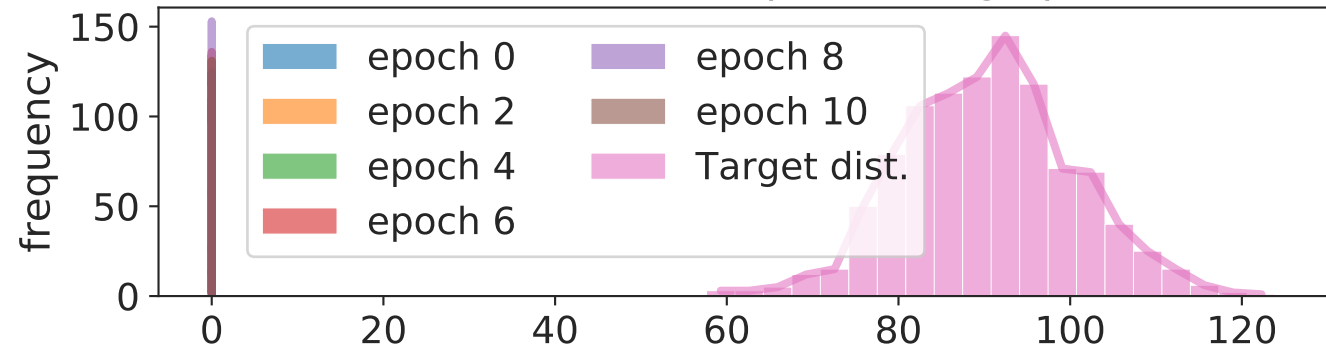
- epoch 0
- epoch 2
- epoch 4
- epoch 6
- epoch 8
- epoch 10
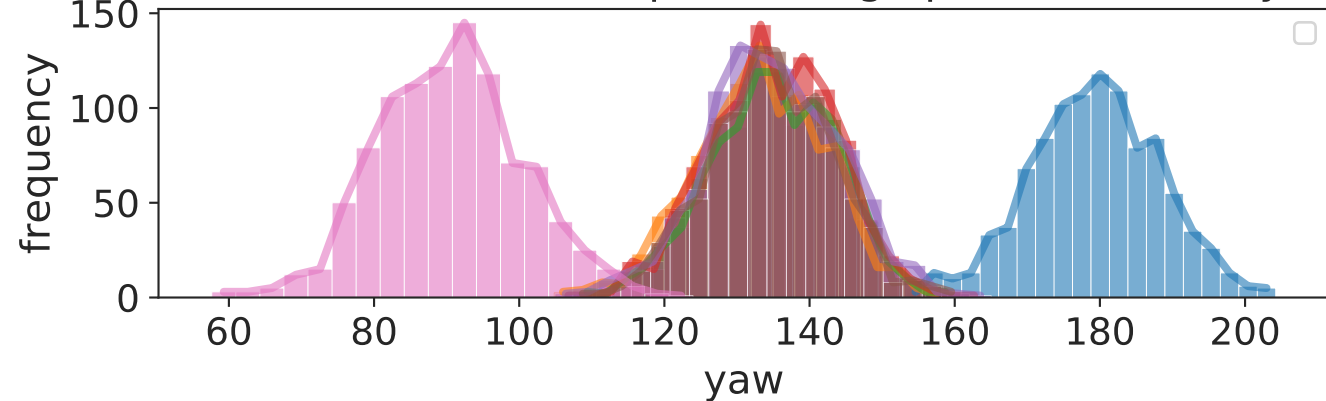- Target dist.

Distribution after 75 pretraining epochs
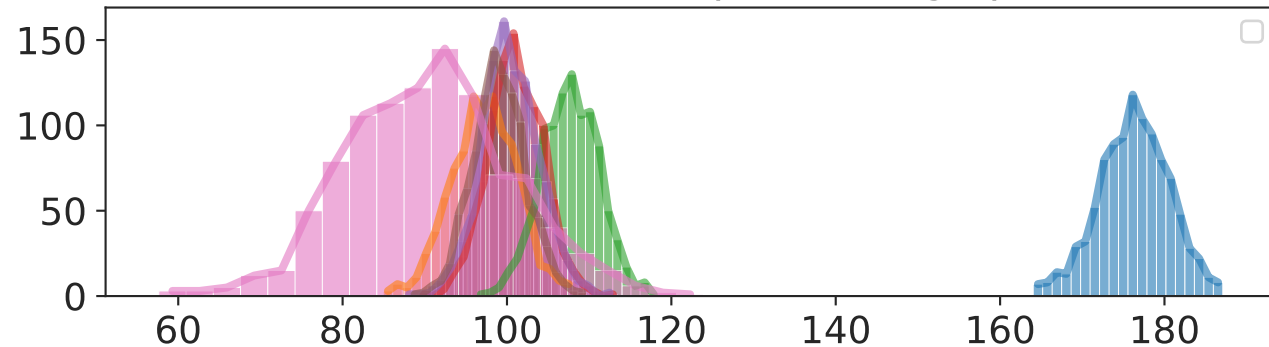
Distribution after 75 pretraining epochs + lr decay

Distribution after 75 pretraining epochs + lr decay (clipped)

frequency

yaw