**训练FixThreePolicy**

**第一阶段，finetune PEL**

**以switch = 50 total =200 all\_window target =70 收集数据**

**得到新的encoder 模型**

model/PA-unseen-3oppo-ours-a1-l1-W5-K20-20240630013352/pel\_encoder\_iter\_499

**第二阶段，训练ICD**

**以switch = 40 total = 400 all\_window target =70 PEL fine tune ICD 来收集数据，并以其中超过35的数据作为collected \_data 来训练原始 PEL ICD**

**微调后的PEL ,ICD 模型为**

model/PA-6oppo-ours-a1-l1-W5-K20-20240630032440/res\_encoder\_iter\_1999

model/PA-6oppo-ours-a1-l1-W5-K20-20240630032440/res\_decoder\_iter\_1999

**（评估，原始基线为34.75，微调后为36.589）**

**训练 ChaseBouncePolicy**

model/PA-6oppo-ours-a1-l1-W5-K20-20240703071700/res\_encoder\_iter\_1999

model/PA-6oppo-ours-a1-l1-W5-K20-20240703071700/res\_decoder\_iter\_1999

**479到494**

**新的**

model/PA-6oppo-ours-a1-l1-W5-K20-20240704062739/res\_encoder\_iter\_1999

model/PA-6oppo-ours-a1-l1-W5-K20-20240704062739/res\_decoder\_iter\_1999