We thank all reviewers for their constructive comments and are glad that our contributions are largely recognized. Below, we address the reviewer's concerns point by point.

To Reviewer #2:

Q1. Additional experiments on harder environments: We agree with the reviewer that experiments of robotic manipulation tasks other than locomotion benchmarks will further emphasize the benefits of our method. Hence in Tab. A, we provide results of three MuJoCo manipulation examples: *Pusher*, *Striker* and *Thrower*. It reads that our method achieves significant improvement over GAIfO and behaves comparably to GAIL. Applying our method for real human demonstrations is a direct extension of our paper, but it demands more practical considerations (*e.g.* domain adaption from human to robot, feature extraction of observations, etc) and will be left for future exploration.

Q2. Our improvement or insight over GAIL: Naive exclusion of actions from GAIL leads to GAIfO, which is demonstrated to perform much worse than GAIL from Table 1 in the paper. In contrast, by bridging the gap between GAIL and GAIfO, our method is able to outperform all other LfO baselines. We believe that our method is preferably a practical choice for imitation learning from observations when GAIL is no longer applicable.

Q3. Codes and reference citation: We thank the reviewer for the reminding. Our code, including detailed instructions on reproducing the results will be made public. Besides, we will cite the reference [Sun et al., 2019] raised by the reviewer. In spite of focusing on the same topic, [Sun et al., 2019] provides theoretical guarantee on the sampling efficiency of LfO over pure RL, while our core insight is improving LfO by investigating the gap between LfO and LfD.

To Reviewer #3:

Q1. Results of DeepMimic: Per the reviewer's suggestion, we have tuned the reward function of DeepMimic on HalfCheetah and Ant by adjusting the weights of different reward terms. Even after careful tuning, DeepMimic is still much worse than our method as observed from Tab. B. Moreover, applying DeepMimic requires to design the hand-craft reward case by case, which makes it impracticable or even inapplicable for diverse types of agent mechanisms.

Q2. On additional tasks: The task with discrete actions has already been included in the main paper (namely, *GridWorld* and *CartPole*), on which our method still performs promisingly. Due to the time limit and the major focus of this work, we would like to make image-based imitation like Atari a future research as the reviewer suggested.

Table A: Additional experiments on manipulation tasks.

	Expert	GAIL	GAIfO	Ours
Pusher	-21.0±2.1	-20.2±1.0	-31.1±6.9	-21.8±1.3
Striker	-101.5±35.9	-118.3±6.0	-178.4±13.9	-127.6±8.3
Thrower	-26.8±0.3	-28.6±1.1	-74.4±5.6	-29.9±1.1

Table B: Reward tuning for DeepMimic.

	HalfCheetah	Ant
DeepMimic (Tuned)	202.6±4.4	-985.3±13.6
Ours	5699.3±51.8	1970.3±110.1

To Reviewer #4:

Q1. The explanations of Eq. 5 : We are sorry for the confusion caused by the explanations of Eq. 5. We will provide more illustrations on the relationship between LfD and LfO including that the divergence of LfD is always greater than LfO and optimizing LfD implies optimizing LfO but not vice versa. For the statement on Eq. 5 $\neq 0$, we apologize for the improper presentation and will add necessary restrictions to make it consistent with Corollary 1.

Q2. The clarification to deterministic systems: We thank the reviewer for reminding this and will make the applicable scope of our method (deterministic systems) clearer in the final revision. We also agree that employment to stochastic dynamics is important for some real-world tasks and will be an exciting direction for future research.

Q3. Learning curves for GAIL over number of interactions and over number of demos: We provide the learning curves under varying numbers of interactions for GAIL along with other methods in the left sub-figure of Fig. A. Besides, the learning curves of GAIL, GAIfO, and our method with different numbers of demos on HalfCheetah are reported in the right sub-figure of Fig. A (those on all other tasks will be included in the final revision due to the space limit here). As indicated by the results, our method is able to outperform GAIL if the number of demos we use is sufficiently larger than that of GAIL (*e.g.* our method with 50 demos vs. GAIL with 10 demos).

Q4. The codes and replication of results: Please refer to our response to Q3, Reviewer #2.

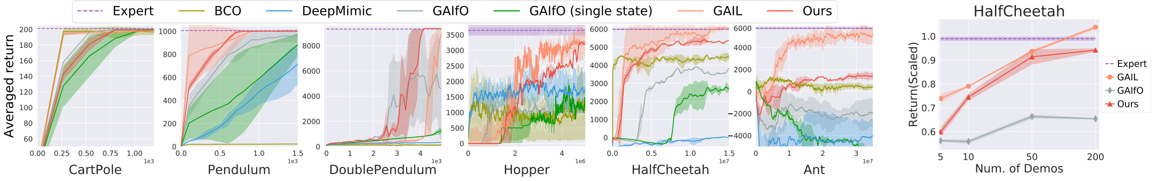

Figure A: **Left**: Learning curves w/ GAIL. **Right**: Results w/ different num. of demos on HalfCheetah task.