Rebuttal

1. Response to reviewer 1

1.1. Training stability

We believe the proposed method does not incur significant instability during training. As shown in Figure 2 in the paper, the standard deviation of the proposed method are similar with the SOTA works CaDM and DOMINO in experiments. Thus, it demonstrates that the domain adaptation methods of CDMetaRL does not cause instability but effectively improves performance.

1.2. Lack of novelty

The mentioned GAIL and a similar work SimGAN cited in the paper are essentially the generative approaches that generate identical decisions corresponding to experts. In contrast, CDA-RL is an adversarial domain adaptation algorithm that is designed to extract domain-invariant features to mitigate the impact of continuous dynamics variations on robot control.

1.3. Weak experimental results

CDAMetaRL achieves significant improvements over the SOTA works as listed in Table 1 in the paper. Specifically, CDAMetaRL brings performance improvements compared with CaDM by 10.17%, 10.50%, 17.34%, and 27.5% in HalfCheetah, CrippleHalfCheetah, Ant, and SlimHumanoid environments, respectively. CDAMetaRL brings performance improvements compared with DOMINO by 9.55%, 5.5%, 19.12%, and 18.77% in the preivous environments, respectively. The experiment results are not well presented and explained in tables, which is to be revised. However, we believe CDAMetaRL demonstrates significant improvements in dynamics domain adaptation.

A demonstration video is presented in the following link that includes details and real experiments conducted in an elevator cabin. https://anonymous.4open.science/r/CDAMetaRL-7357

1.4. Writing clarity

Thanks for your suggestion. We will provide the algorithm with pseudo-code in the revision.

2. Response to reviewer 2

2.1. Domain parameters setting

We sincerely appreciate your suggestion. There are typos in target domain settings mentioned in lines 268-270 in the section 4.1.1 in the paper. We explicitly lists the source and target domain settings in Table 4 in the supplement materials in lines 644-658. Specifically, the source domain is a span of 0.7-1.3 for all five environments, and the target domain is a span of 0.4-0.6 and 1.4-1.6 for four experiments and 0.5-0.7, 1.3-1.5 for the Pendulum environment, which are similar with the settings adopted in the SOTA works CaDM and DOMINO. This settings involves significant domain gaps for all baselines and our methods.

2.2. Dynamics-extrapolate experiments

Thanks for your suggestion. We will supplement the dynamics-extrapolate experiments.

2.3. Lack of analysis on the ”few-shot” statement

As described in lines 304-305 in the paper, the support set of CDAMetaRL is set to one trajectory, which is essentially the one-shot setting. As listed in Table 1 and 2 in the paper, CDAMetaRL outperforms baselines in five environments in the one-shot scenario. We believe the proposed method will consistently perform well in the few-shot (3, 5, 7 trajectories) scenarios. We will provide detailed experiment results in the revision version.

Morover, as listed in Table 2 in the paper, the ablation study shows the performance of CDAMetaRL without meta optimization indicated as ‘CDA-RL’, and the completed CDAMetaRL with meta optimization. It shows the performance of CDA-RL drops 30.2%, 13.2%, and 19.7% in the HalfCheetah, Ant, and SlimHumanoid environments, respectively. As such, the meta-optimization method significantly enhance few-shot domain transfering capability.

Table 1. Average and standard deviation of return for each method.

Env	CaDM	DOMINO	Ours
SlimHumanoid	5813.4±1200.2	6244.1±1093.2	**7416.3**±**783.63**
Ant	234.6±19.98	231.1±42.16	**275.3**±**51.58**
HalfCheetah	1376.9±115.03	1384.2±127.58	**1516.4**±**132.89**
CrippleHalfCheetah	1768.5±109.14	1852.3±134.58	**1954.2**±**117.62**
Pendulum	-261.7±38.6	-376.4±84.2	**-254.4**±**42.1**

2.4. Experimental validity

CDAMetaRL achieves significant improvements over the SOTA works as listed in Table 1 in the paper. Specifically, CDAMetaRL brings performance improvements compared with CaDM by 10.17%, 10.50%, 17.34%, and 27.5% in HalfCheetah, CrippleHalfCheetah, Ant, and SlimHumanoid environments, respectively. CDAMetaRL brings performance improvements compared with DOMINO by 9.55%, 5.5%, 19.12%, and 18.77% in the previous environments, respectively. The experiment results are not well presented and explained in tables, which is to be revised. However, we believe CDAMetaRL demonstrates significant improvements in dynamics domain adaptation.

A demonstration video is presented in the following link that includes details and real experiments conducted in an elevator cabin. https://anonymous.4open.science/r/CDAMetaRL-7357

3. Response to reviewer 3

3.1. Experiment details

Thanks for your suggestion. The Mujoco version is 2.1 in experiments. Additionally, we employ the source code in the referenced papers for environments such as Ant, SlimHumanoid, HalfCheetah, etc. For more details, please refer to the supplementary materials.

3.2. Enhancing reliability of results

We omit the standard deviation due to paper length limits and only list the mean in Table 1 in the paper. Herein, the standard deviations of two SOTA works and ours in three typical environments are presented in the following. In the SlimHumanoid environment, the standard deviations of CaDM, DOMINO, and CDAMetaRL are 1200.22, 1093.2, and 783.63, respectively. In the Ant environment, their standard deviations are 19.98, 42.16, and 51.58, respectively. In the Halfcheetah environment, their standard deviations are 115.0, 127.58, and 120.89, respectively. As listed in the following table, our proposed CDAMetaRL consistently outperforms the two SOTA works with less deviations. The detailed experimental results would be added in the supplementary materials in the revision version.

A demonstration video is presented in the following link that includes details and real experiments conducted in an elevator cabin. https://anonymous.4open.science/r/CDAMetaRL-7357

3.3. Shortcomings and future work

CDAMetaRL demonstrates excellent performance in addressing the dynamic mismatches. Such problems also exist in the images as inputs, such as variations in shooting angles and lighting conditions. We plan to extend our proposed method on images as inputs for robotic control applications.