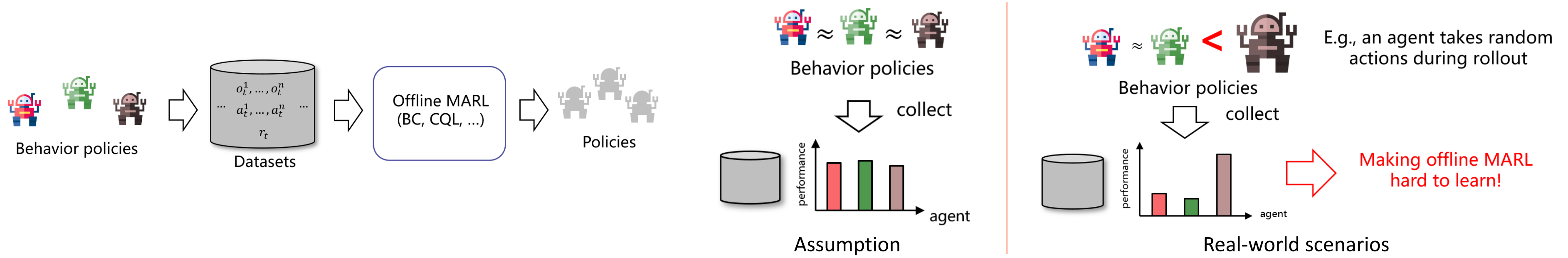


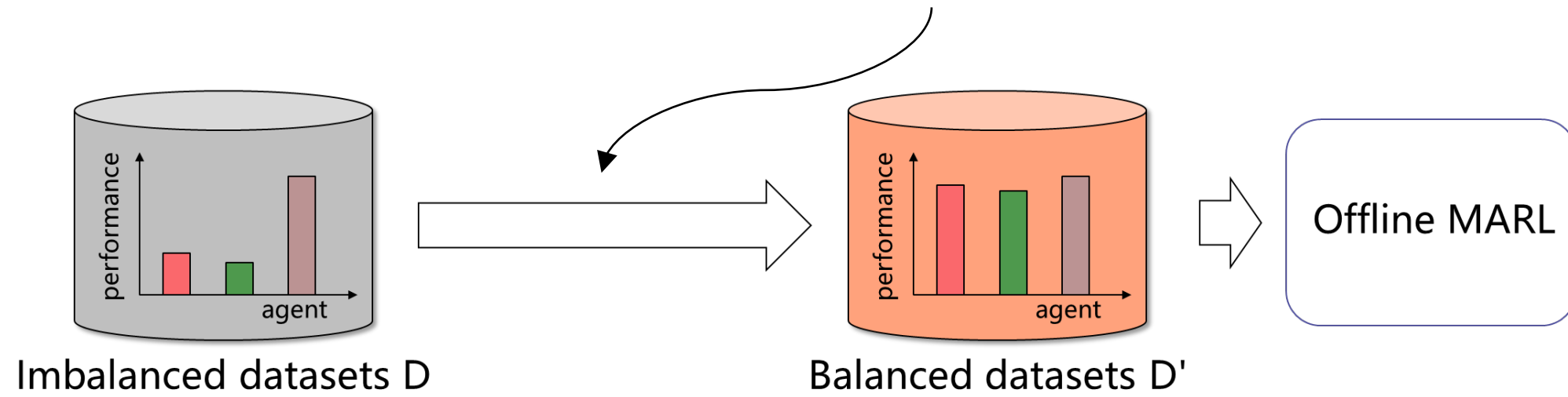
Introduction

1. Offline MARL is promising for learning coordination policies under costly tasks.

Most offline MARL work assume agent quality in behavior policies is **balanced**, which is not always the case!

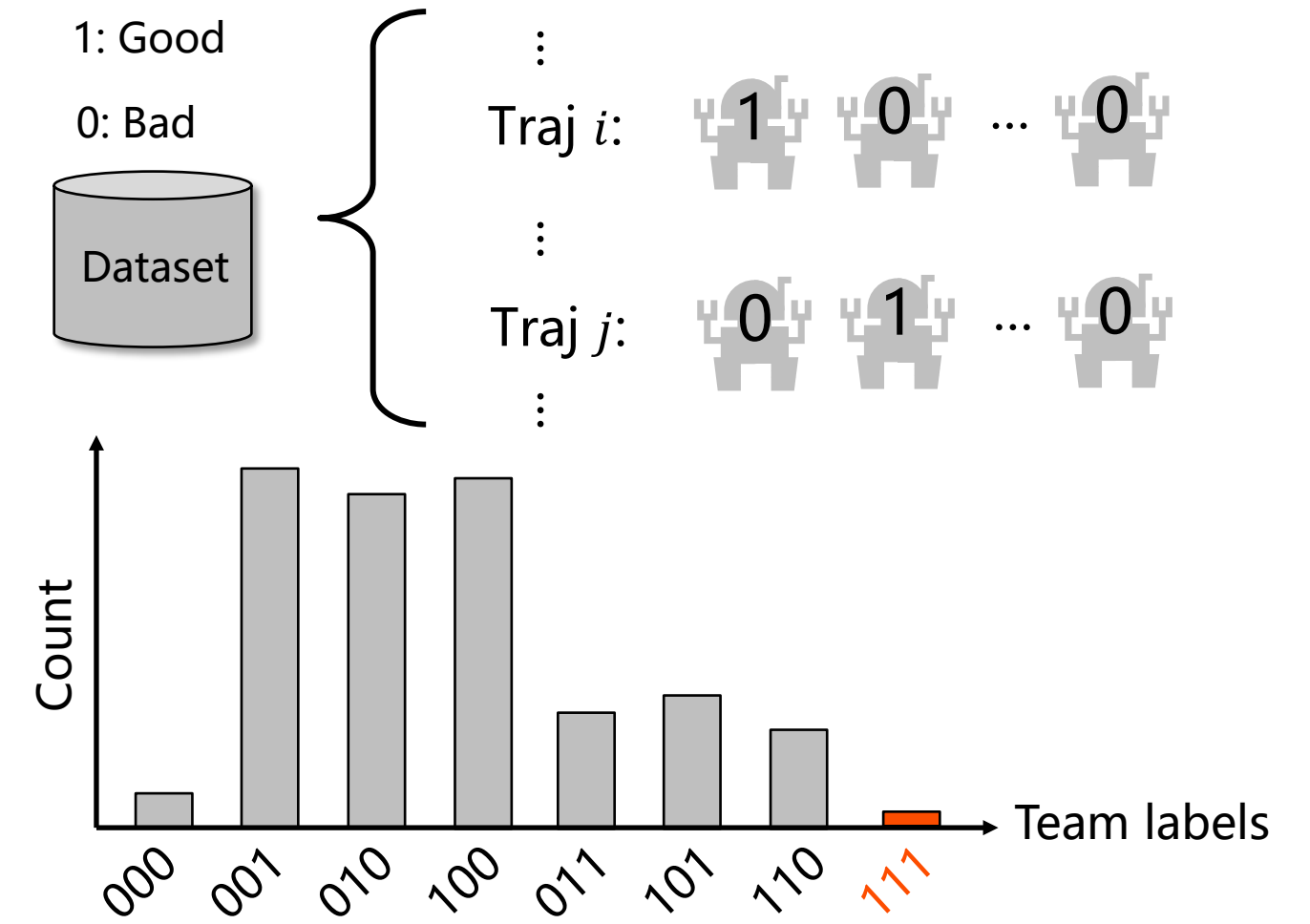


2. To bridge this gap, one solution is data augmentation using generative models.



3. New challenge: generalization to high-quality & balanced coordination data

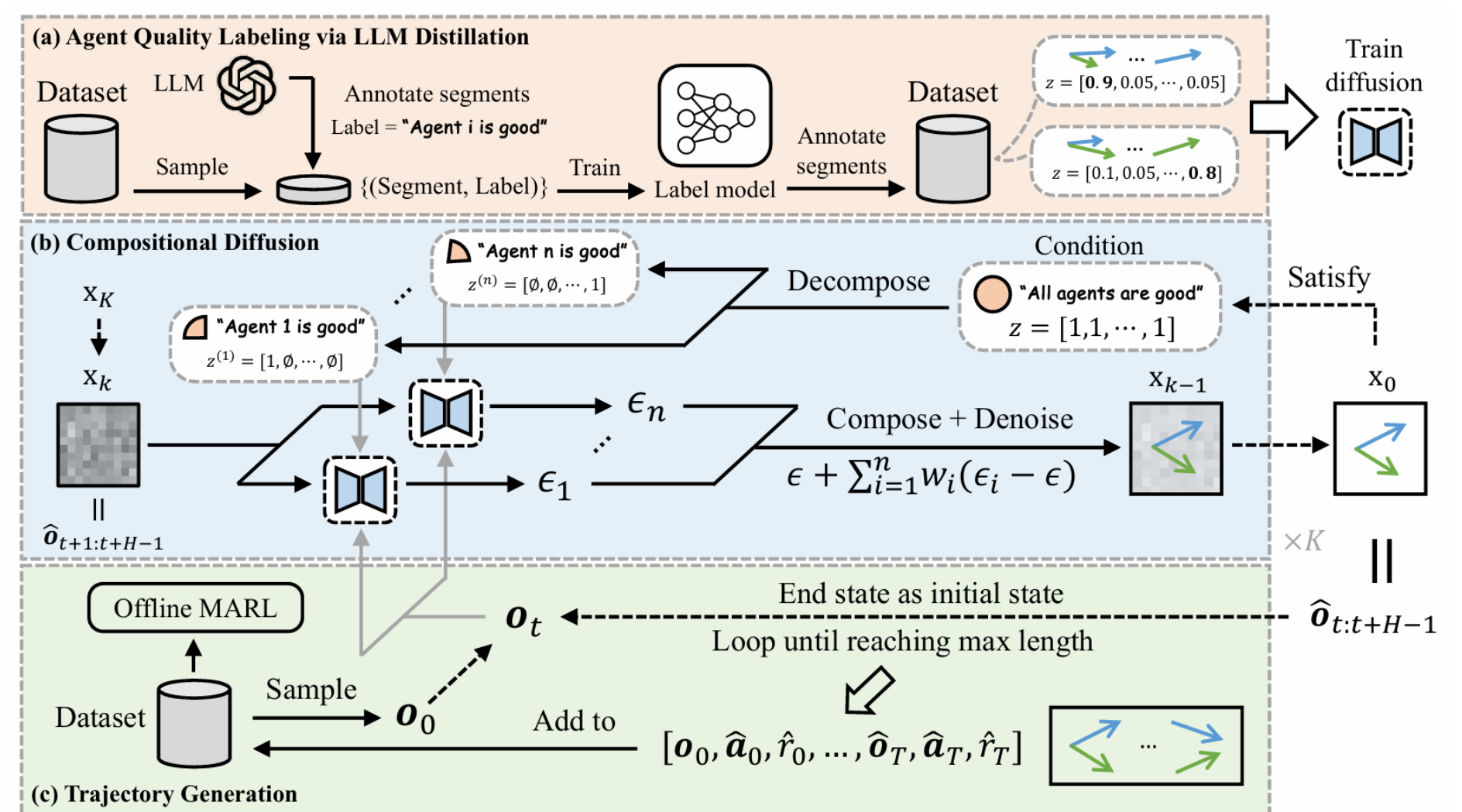
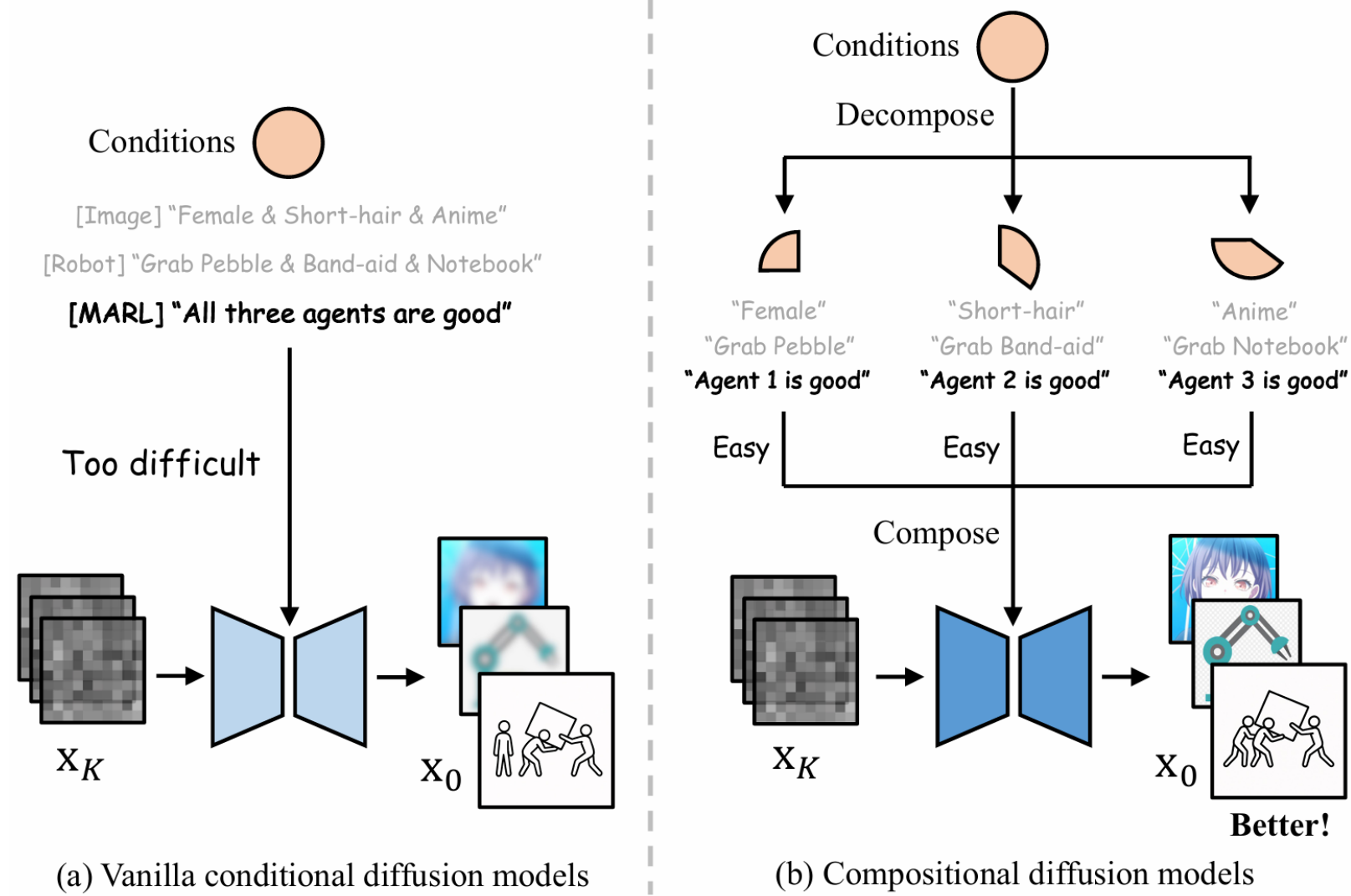
Our desired generation is 111 ("All 3 agents are good"), but it is out of distribution (OOD):



How to compose scattered individual high-quality behaviors for generalization?

Method

We propose **Compositional Diffusion for Imbalanced Datasets (CODI)**



Experiments

Performance of offline MARL policies learned from data augmented by different methods

Envs	Algs	Original	MBTS	MADiff	MADiTS	CODI w/o Com.	CODI w Pri.	CODI
CN	BC	0.21±0.06	-1.34±0.05	0.33±0.02	0.33±0.02	0.36±0.04	0.59±0.09	0.58±0.09
	50%BC	-0.02±0.03	-0.02±0.03	-0.25±0.06	-0.36±0.13	0.11±0.14	0.56±0.07	0.43±0.05
	OMAR	-4.06±1.20	-0.27±0.48	0.26±0.36	-0.10±0.95	0.35±0.43	0.76±0.36	0.50±0.83
	OMIGA	0.66±0.19	0.84±0.07	0.93±0.07	0.92±0.01	0.94±0.04	1.01±0.06	0.96±0.04
World	BC	0.43±0.15	-0.49±0.09	-0.18±0.24	-0.25±0.05	0.31±0.18	0.51±0.11	0.61±0.16
	50%BC	0.06±0.26	-0.49±0.19	-0.50±0.10	-0.52±0.03	0.00±0.11	0.27±0.10	0.38±0.21
	OMAR	-0.88±0.71	-1.55±0.11	-1.02±0.47	-0.92±0.49	-0.93±0.37	-0.43±0.38	-0.40±0.64
	OMIGA	0.66±0.10	-3.36±0.81	-3.60±0.17	-3.46±0.27	-2.42±1.10	-2.07±1.73	0.93±0.14
3m	BC	0.52±0.25	0.22±0.00	0.45±0.11	0.57±0.22	0.61±0.28	0.78±0.08	0.75±0.03
	50%BC	0.28±0.02	0.28±0.02	0.48±0.02	0.36±0.14	0.37±0.07	0.49±0.12	0.58±0.09
	OMAR	0.24±0.05	0.08±0.10	0.45±0.15	0.66±0.23	0.63±0.29	0.95±0.04	0.72±0.09
	OMIGA	0.36±0.04	0.34±0.03	0.28±0.09	0.37±0.03	0.26±0.05	0.44±0.07	0.40±0.10
Zerg_3v4	BC	0.74±0.00	0.69±0.68	0.96±0.31	0.87±0.19	1.01±0.35	0.95±0.19	1.07±0.12
	50%BC	1.11±0.20	0.93±0.06	0.87±0.35	0.96±0.10	1.10±0.10	1.23±0.16	1.25±0.07
	OMAR	0.62±0.08	0.32±0.07	0.63±0.06	0.59±0.06	0.59±0.10	0.68±0.28	0.65±0.10
	OMIGA	0.39±0.12	0.42±0.17	0.63±0.05	0.59±0.24	0.56±0.14	0.46±0.15	0.62±0.40
Average		0.08	-0.21	0.05	0.04	0.24	0.45	0.63

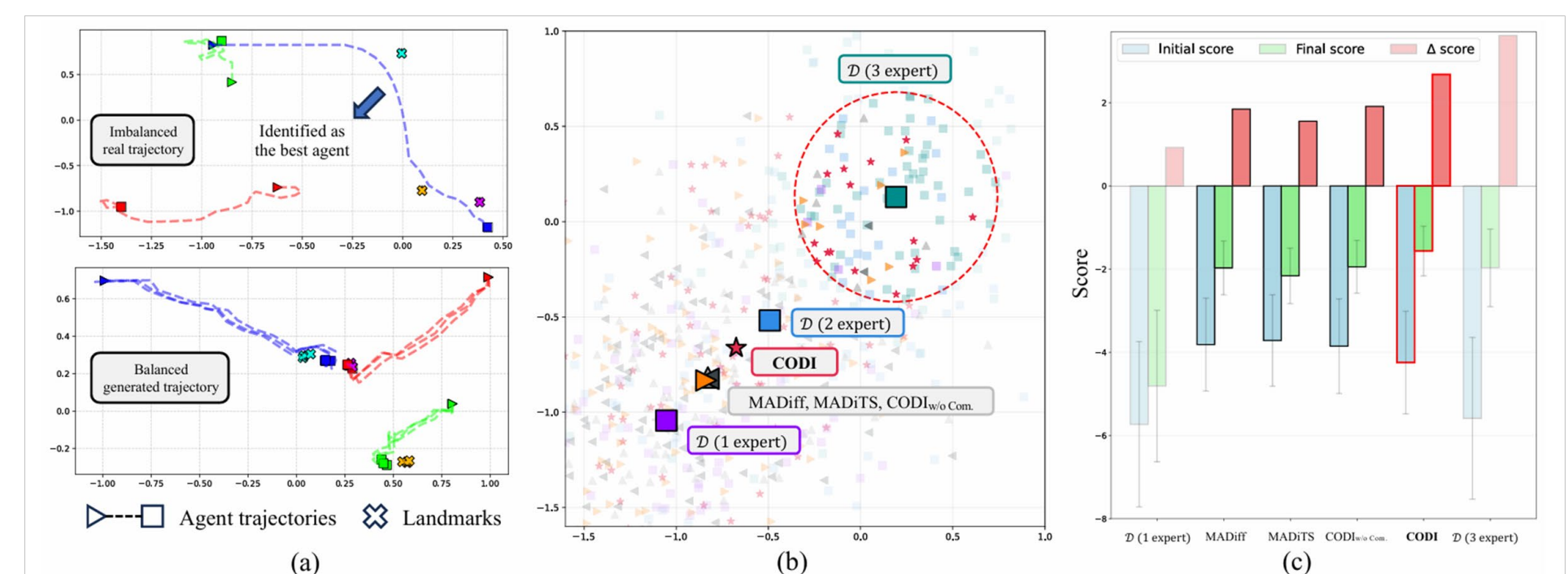


Figure 4: A case study on the CN task. (a) Trajectory visualization: a real imbalanced trajectory (top) with only one agent approaching the landmarks versus a balanced CODI-generated trajectory (bottom) where all agents navigate to the landmarks. (b) Trajectory embeddings: CODI brings the mean embedding (red star) closer to the balanced real dataset (green square), and yields more better trajectories (red circle). (c) Score improvement: CODI yields the highest improvement across methods.

CODI generates coordination trajectories with better quality!