COMBO: Compositional World Models for Embodied Multi-Agent Cooperation

Hongxin Zhang[1]*, Zeyuan Wang[2]*, Qiushi Lyu[3]*, Zheyuan Zhang[4]
Sunli Chen[2], Tianmin Shu[5], Yilun Du[6], Chuang Gan[1,7]

[1] University of Massachusetts Amherst [2] IIIS, Tsinghua University
[3] Peking University [4] University of Michigan [5] Johns Hopkins University
[6] Massachusetts Institute of Technology [7] MIT-IBM Watson AI Lab

Abstract. In this paper, we investigate the problem of embodied multi-agent cooperation, where decentralized agents must cooperate given only partial egocentric views of the world. To effectively plan in this setting, in contrast to learning world dynamics in a single-agent scenario, we must simulate world dynamics conditioned on an arbitrary number of agents' actions given only partial egocentric visual observations of the world. To address this issue of partial observability, we first train generative models to estimate the overall world state given partial egocentric observations. To enable accurate simulation of multiple sets of actions on this world state, we then propose to learn a compositional world model for multi-agent cooperation by factorizing the naturally composable joint actions of multiple agents and compositionally generating the video. By
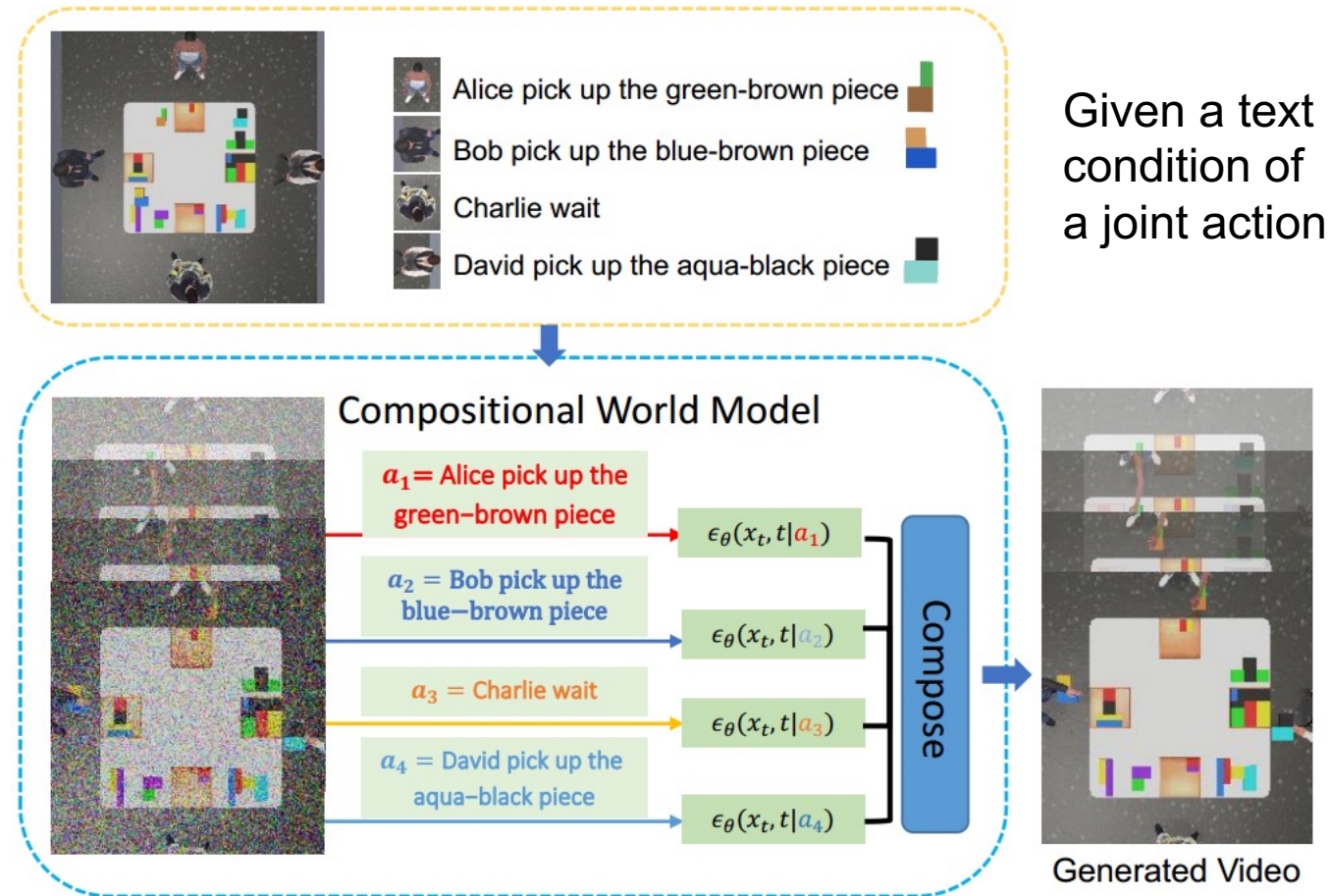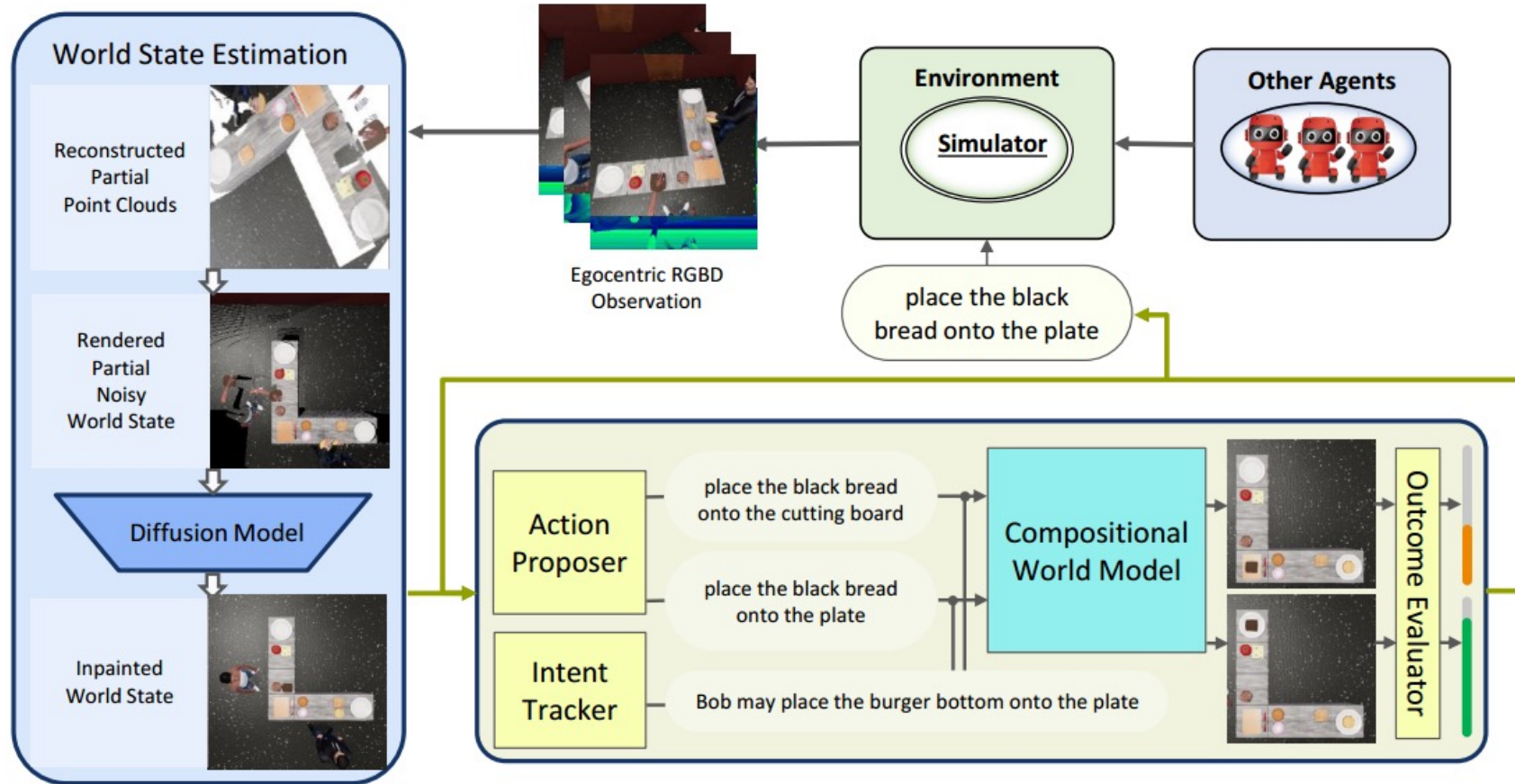
Introduction
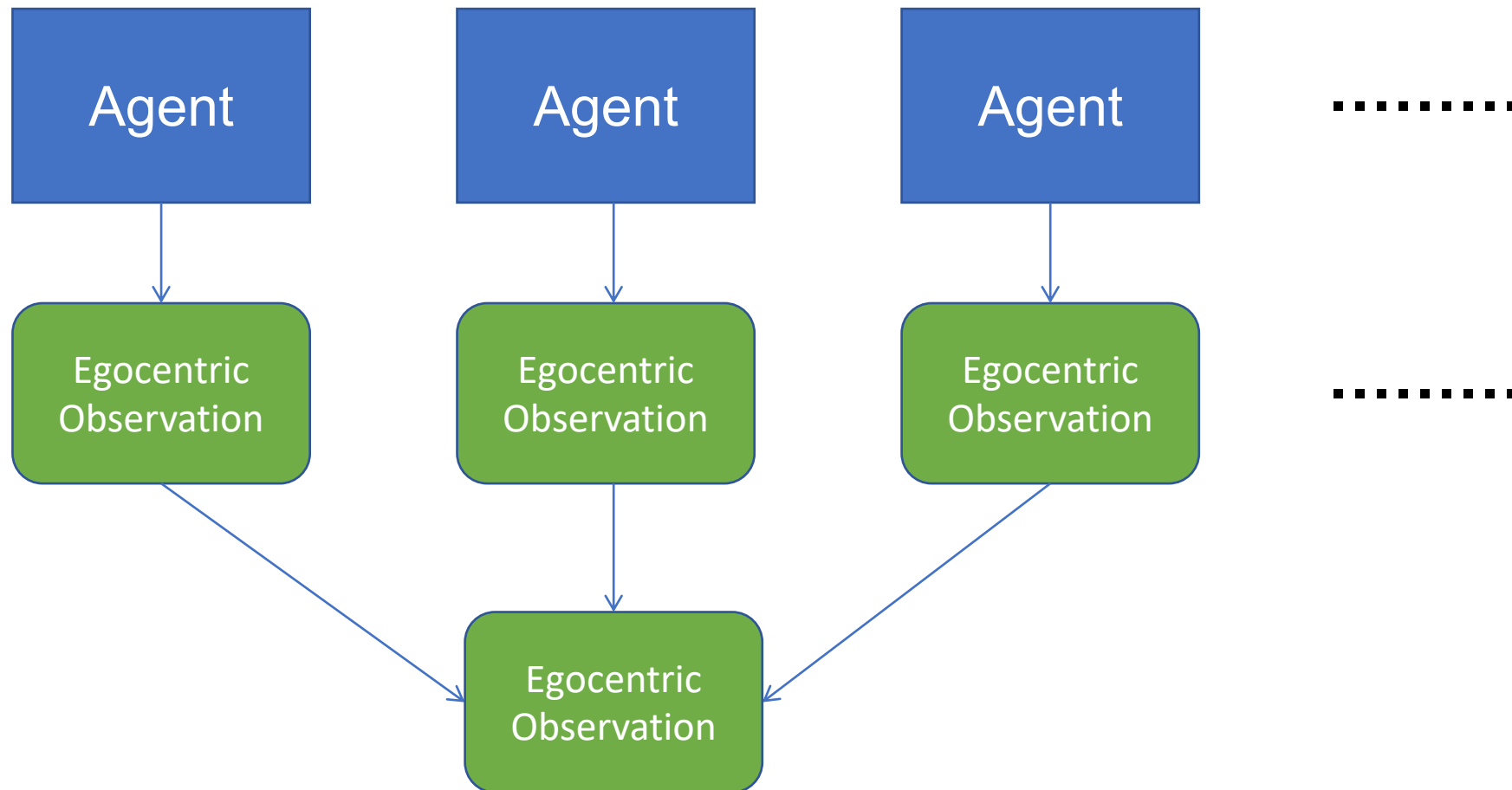
Compositional World Model + VLM

Compositional World Model

Given a text condition of a joint action

separate it into several components corresponding to each agent

COMBO

Experiment

Cooperating with		TDW-Game		TDW-Cook	
		Agent 1	Agent 2	Agent 1	Agent 2
Recurrent World Models [23]	Success Rate	0	0	0	0
	Average Steps	/	/	/	/
MAPPO [78]	Success Rate	0	0	0	0
	Average Steps	/	/	/	/
CoELA (Oracle Vision) [80]	Success Rate	1.0	0.8	0.5	0.2
	Average Steps	17.8	26.1	38.2	29.5
LLaVA [42]	Success Rate	0.9	1.0	1.0	0.9
	Average Steps	26.0	31.1	31.5	36.7
COMBO (w/o *IT*)	Success Rate	0.8	0.7	0.9	0.9
	Average Steps	15.6	17.4	24.8	21.0
COMBO (Ours)	Success Rate	**1.0**	**1.0**	**1.0**	**1.0**
	Average Steps	16.8	18.1	**21.3**	**21.7**
Oracle Cooperator (Oracle Vision)	Success Rate	1.0	1.0	1.0	1.0
	Average Steps	14.9	16.8	23.4	19.3

Thank You!

Avenida da Universidade, Taipa, Macau, China
Tel： (853) 8822 8833　　Fax： (853) 8822 8822
Email： xiongyilee@outlook.com　　Website： www.um.edu.mo