CausalVerse: Benchmarking Causal Representation Learning with Configurable High-Fidelity Simulations (Spotlight)

Guangyi Chen*[12], Yunlong Deng*[2], Peiyuan Zhu*[2], Yan Li*[2], Yifan Shen[2], Zijian Li[12], Kun Zhang[12]

[1] Carnegie Mellon University [2] Mohamed bin Zayed University of Artificial Intelligence   *equal contribution
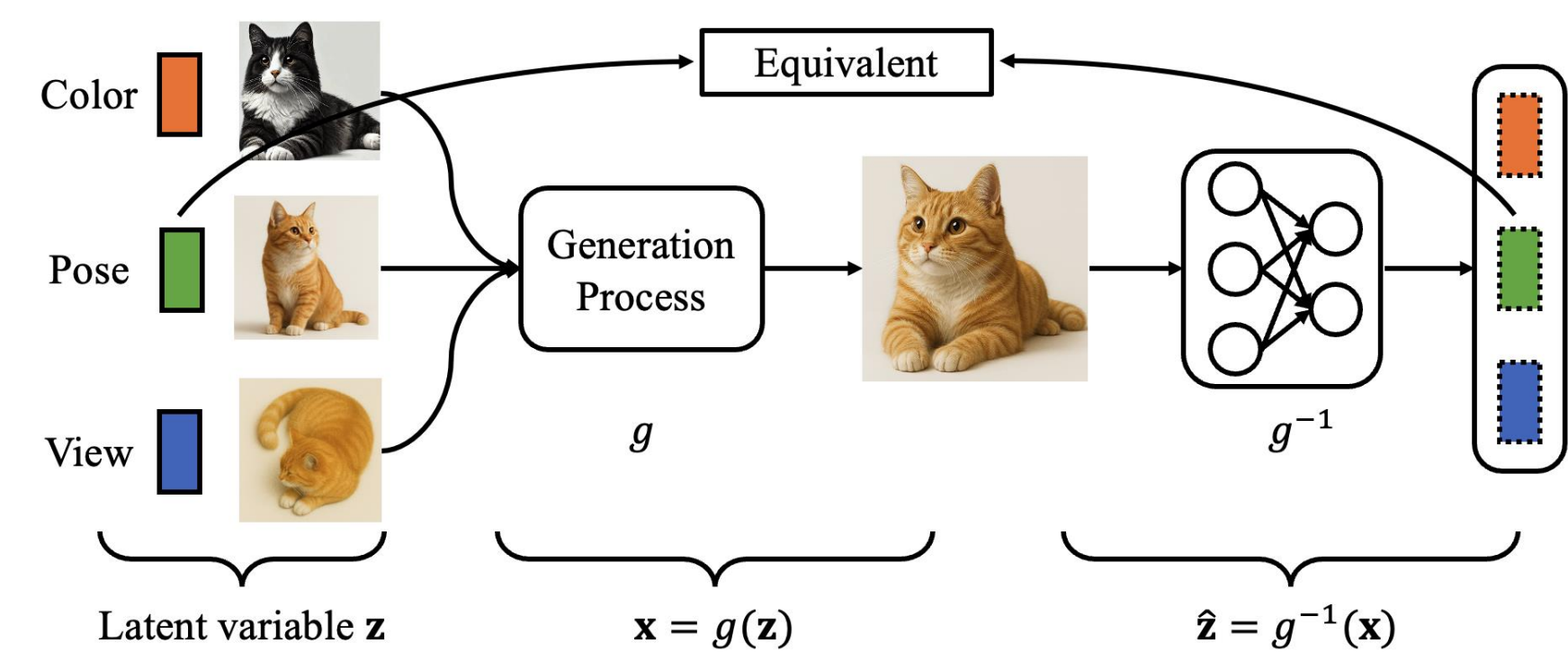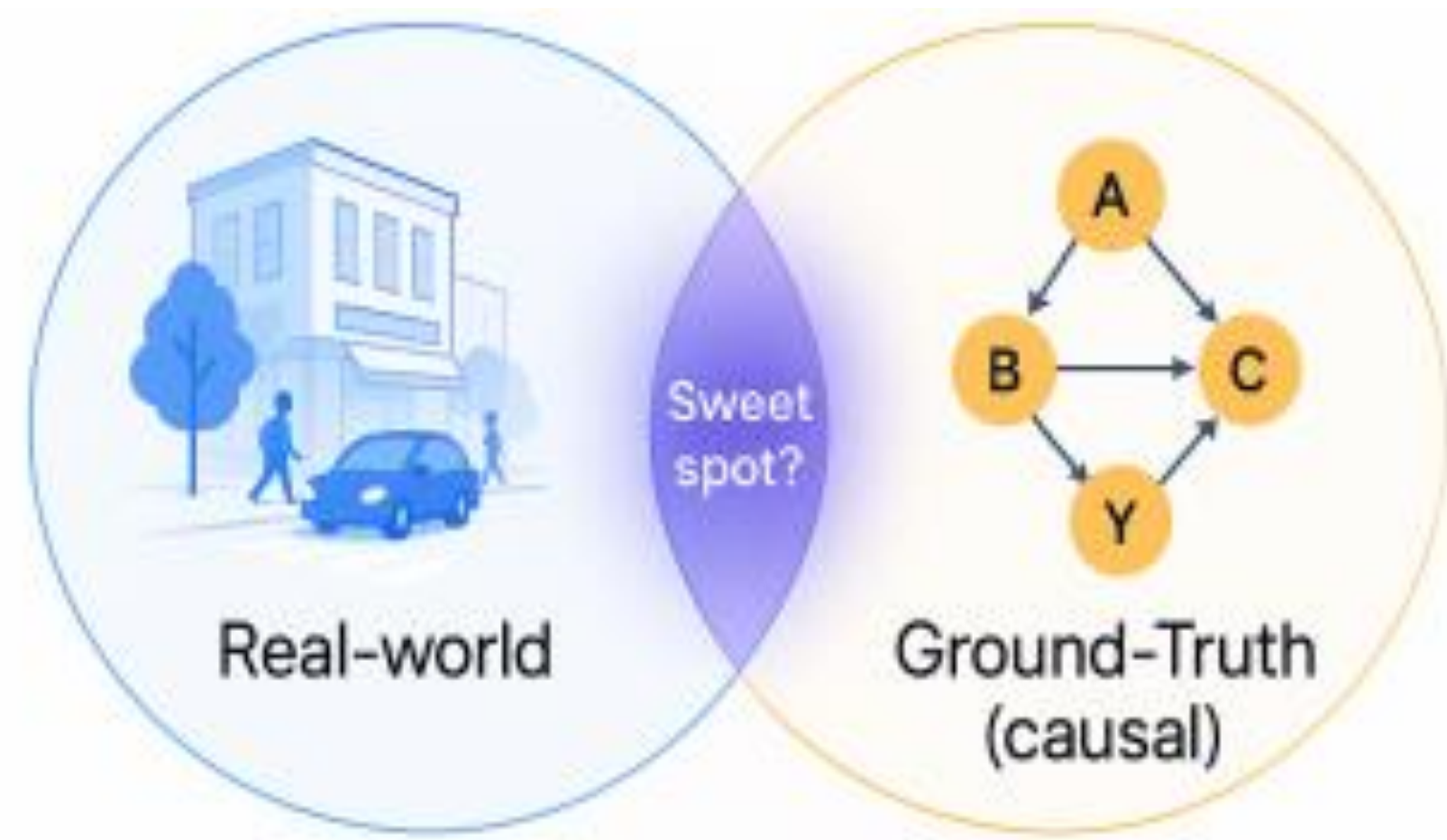
Motivation

Causal representation learning: invert the data generation process to identify the ground-truth causal factors.
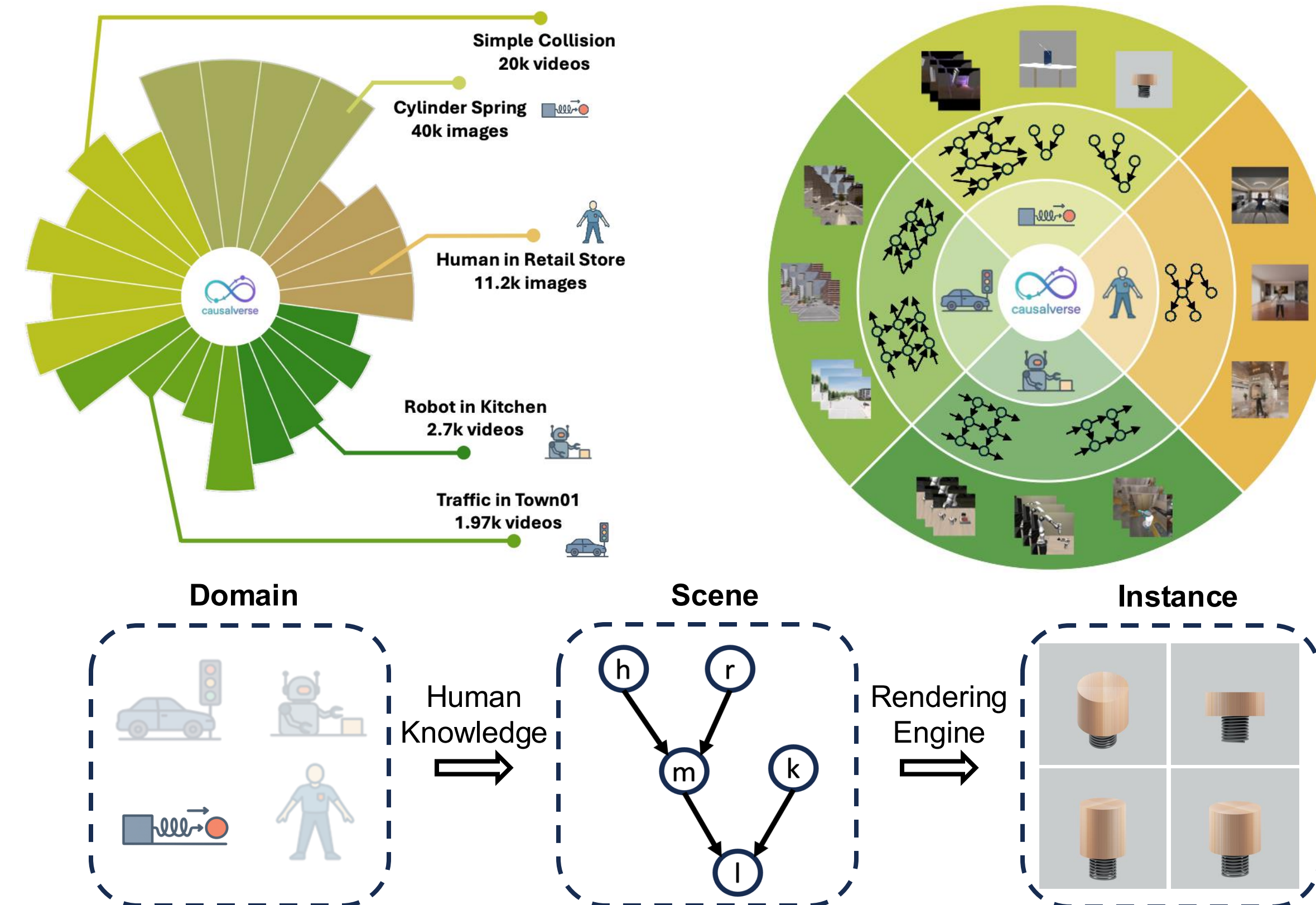

Trade-off: Realism ↔ Evaluative Precision

CausalVerse is the first comprehensive benchmark for causal representation learning with controllable high-fidelity simulations.
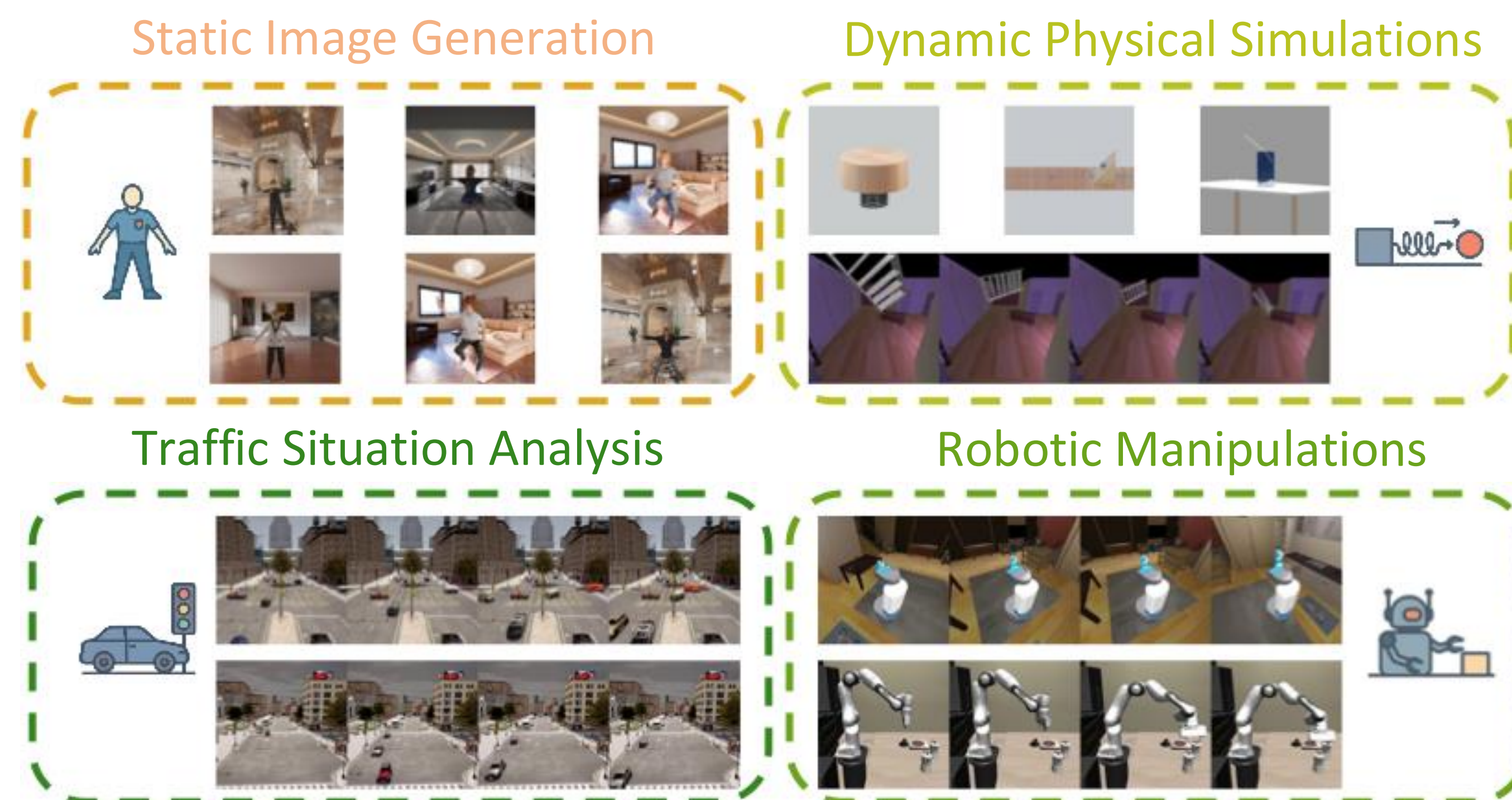
Benchmark	Design for CRL	Data Type	Scale	Domain	Latent Variables	Dynamic Data	Ground Truth
Pendulum [29]	✗	Image	7k	Physics	<10	✗	✓
Flow [29]	✗	Image	8k	Physics	<10	✗	✓
CAUSAL3D [4]	✗	Image	190k	Physics	<10	✗	✓
Causal3DIdent [16]	✓	Image	277k	3D scenes	10	✗	✓
3DIdent [5]	✓	Image	275k	3D scenes	10	✗	✓
CausalCircuit [18]	✓	Image	120k	Electronics	4	✗	✓
Ball [3]	✓	Video	2500k	Ball	<10	✓	✓
Cloth [3]	✓	Video	600k	Cloth	<10	✓	✓
Light Tunnel [30]	✓	Image	60k	Physical	3-5	✓	✓
ISTAnt [31]	✗	Video	792k	Biology	-	✓	✗
CausalVerse	✓	Image / Video	198.66k / 300m	Multi Domains	3-129	✓	✓

Construction

Choose domain -> Decide Scene->Causal graph->Render Image/Video

Dataset Structure

Static Image Generation

Dynamic Physical Simulations

Traffic Situation Analysis

Robotic Manipulations

Evaluation

Table 3: **MCC and R^2 on image-based scenes under unsatisfied assumptions.**

Algorithm	Scene	MCC	R^2
Supervised	Ball on the Slope	0.9878	0.9962
	Cylinder Spring	0.9970	0.9910
	Light Refraction	0.9900	0.9800
	Avg	0.9916	0.9891
Sufficient Change [6]	Ball on the Slope	0.4434	0.9630
	Cylinder Spring	0.6092	0.9344
	Light Refraction	0.6778	0.8420
	Avg	0.5768	0.9131
Mechanism Sparsity [41]	Ball on the Slope	0.2491	0.3242
	Cylinder Spring	0.3353	0.2340
	Light Refraction	0.1836	0.4067
	Avg	0.2560	0.3216
Multiview [16]	Ball on the Slope	0.4109	0.9658
	Cylinder Spring	0.4523	0.7841
	Light Refraction	0.3363	0.7841
	Avg	0.3998	0.8447
Contrastive Learning [42]	Ball on the Slope	0.2853	0.9604
	Cylinder Spring	0.6342	0.9920
	Light Refraction	0.3773	0.9677
	Avg	0.4323	0.9734

Table 3: **MCC and R^2 on video-based scenes.**

Algorithm	Fall Simple		Robotics Study	
	MCC	R^2	MCC	R^2
IDOL [24]	0.2527	0.5901	0.2500	0.6503
CaRiNG [8]	0.2280	0.5457	0.2225	0.6476
TDRL [1]	0.2003	0.5525	0.2440	0.6394
TCL [43]	0.1717	0.4892	0.2163	0.6150
iVAE [44]	0.1881	0.5233	0.1948	0.6165

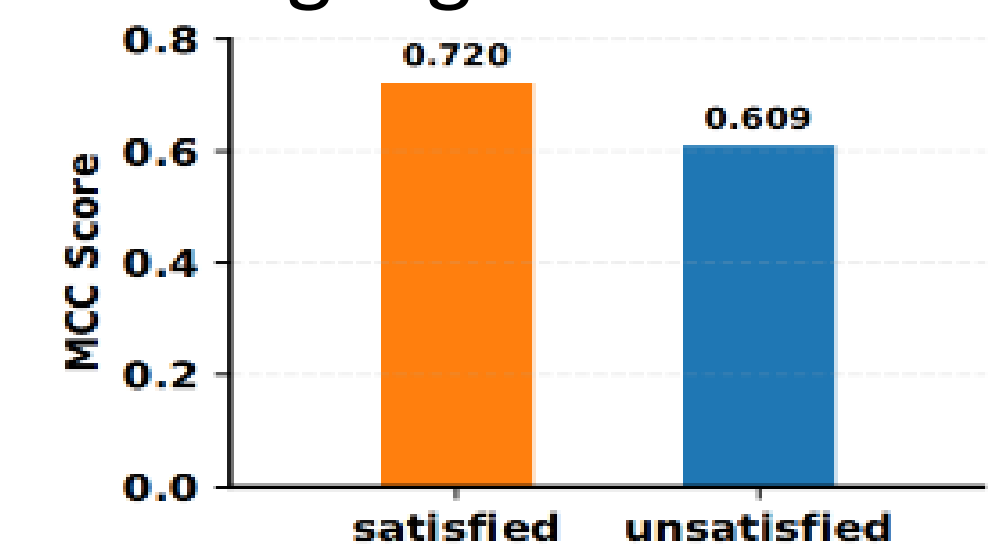Achieving high MCC is difficult

Figure 6: MCC comparison of sufficient change in satisfied and unmet dataset.

Configurable evaluation