OBJECT-ORIENTED MODEL LEARNING THROUGH MULTI-LEVEL ABSTRACTION

Anonymous authors
Paper under double-blind review

1 MODULAR TEST

We conduct modular test to better understand the contribution of each abstraction level. First, we investigate whether the level of dynamics learning can learn the accurate dynamics model when the coarse region proposals of dynamic instances are given. We remove the other two levels and replace them by the artificially synthesized coarse proposals of dynamic instances to test the independent performance of the dynamics learning level. Specifically, the synthesized data are generated by adding standard Gaussian or Poisson noise on ground-true dynamic instance masks (Figure 1). As shown in Table 1, the level of dynamics learning can learn accurate dynamics of all dynamic objects given coarse proposals of dynamic instances. Similarly, we also test the independent performance of the dynamics instance segmentation level. We replace the foreground proposal generated by the motion detection level with the artificially synthesized noisy foreground proposal. Figure 2 shows cases to demonstrate our learned dynamic instances in the level of dynamic instance segmentation, which demonstrates the competence of the dynamic instance segmentation level. Taken together, the modular test shows that each level of MAOP can independently perform well and has a good robustness to the proposals generated by the more abstracted level.

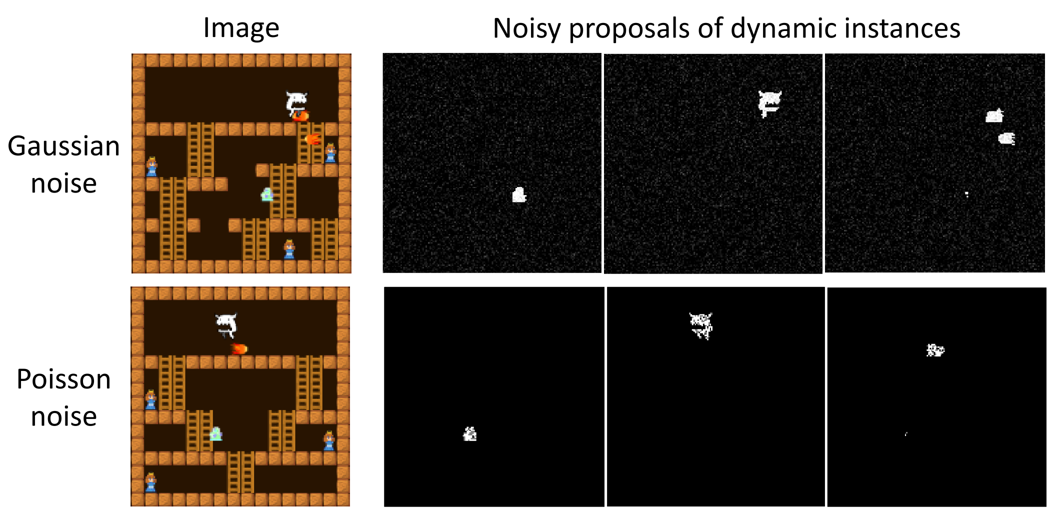

Figure 1: Noisy region proposals of dynamic instances. Zoom in to see the details.

2 FREEWAY DOMAIN

We also test our model on *Freeway* from Atari games, which has a large number of dynamic objects. To test generalization ability, we use first 1800 frames for training and the last 200 frames for testing. As shown in Table 2, our model outperforms the existing modeling methods in this domain. Note that only the ground-true location of the agent is accessible in Arcade Learning Environment, so we just show the quantitative prediction performance of the agent's dynamics. Actually, we observe that the predictions of other dynamic objects are also accurate by comparing the predicted with the

Noise type of proposals	Training environments						Unseen environments					
	0-acc		1-acc		2-acc		0-acc		1-acc		2-acc	
	Agent	All	Agent	All	Agent	All	Agent	All	Agent	All	Agent	All
Computed by DIS	0.80	0.83	0.96	0.93	0.98	0.94	0.80	0.85	0.95	0.93	0.98	0.95
Gaussian	0.63	0.57	0.94	0.89	0.99	0.95	0.60	0.57	0.93	0.89	0.98	0.95
Poisson	0.93	0.91	0.98	0.95	0.99	0.96	0.93	0.91	0.99	0.96	0.99	0.96

Table 1: Prediction performance of the dynamic instance level with different region proposals of dynamic instances. ”All” represents all dynamic objects. ”Computed by DIS” refers to using the proposal regions of dynamic instances computed from the level of dynamic instance segmentation in MAOP.

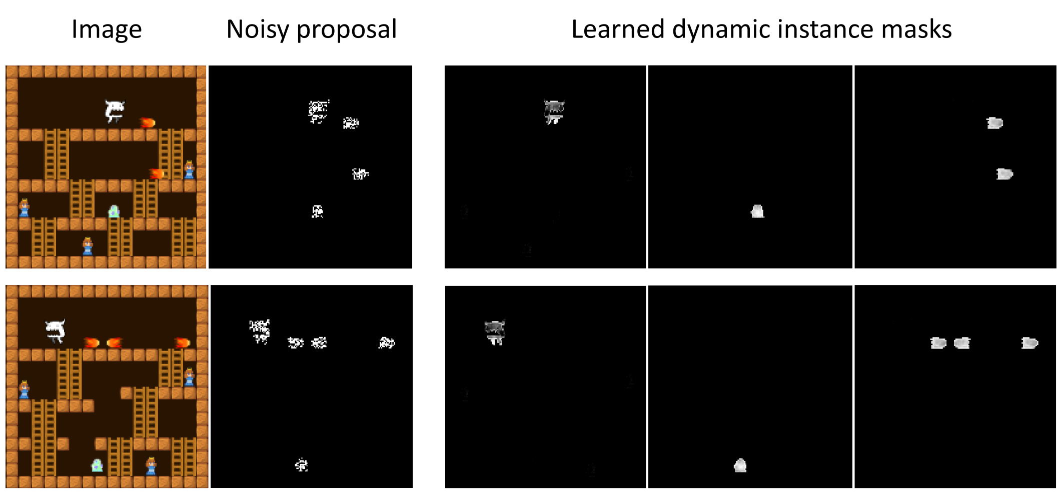

Figure 2: The learned dynamic instance masks in the level of dynamic instance segmentation with noisy foreground proposals.

ground-true images, as shown in Figure 3. The validation results on *Freeway* demonstrate that our model is effective for the concurrent dynamics prediction of a large number of objects.

Model	Training environments			Unseen environments		
	0-acc	1-acc	2-acc	0-acc	1-acc	2-acc
MAOP	**0.80**	**0.91**	**0.94**	**0.79**	**0.89**	**0.94**
CDNA	0.46	0.65	0.73	0.46	0.66	0.74
AC Model	0.59	0.74	0.78	0.63	0.76	0.80

Table 2: Prediction performance of the agent's dynamics on *Freeway*. n-acc means the n-error accuracy.

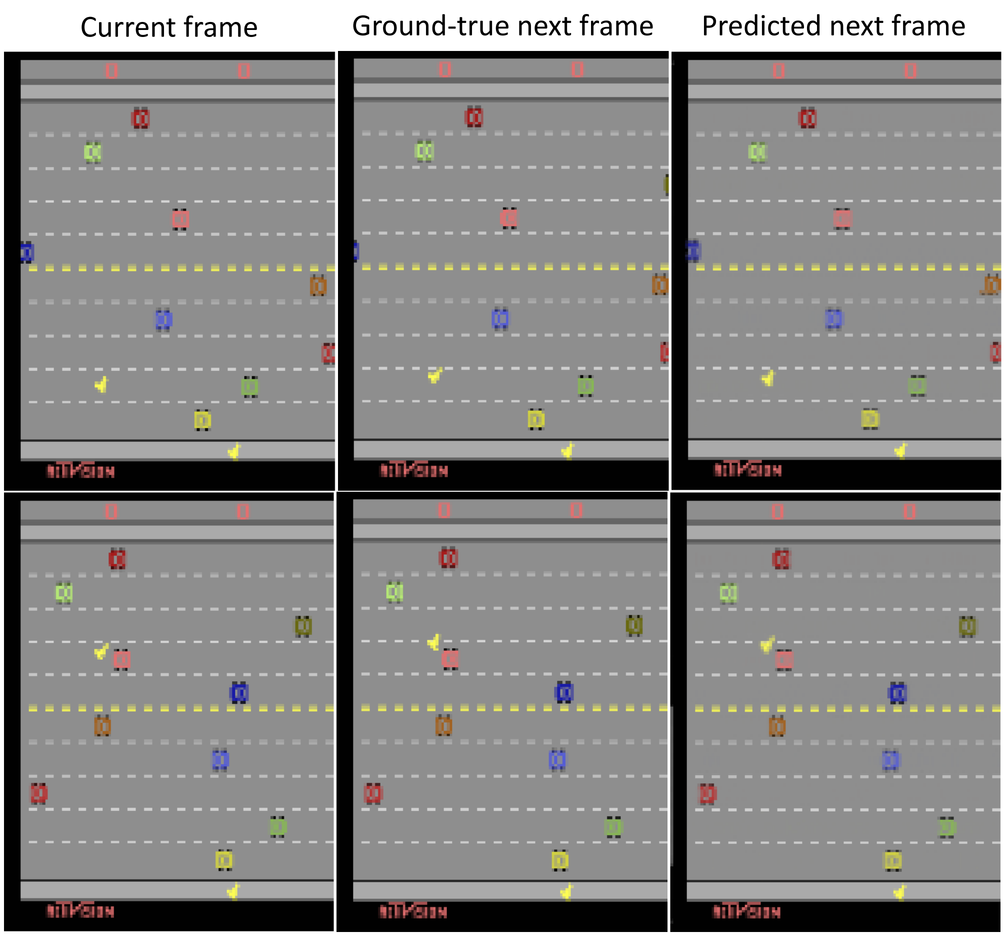

Figure 3: Image predictions in testing environments on *Freeway*.