Supplementary Material: Task-Agnostic Dynamics Priors for Deep Reinforcement Learning

Yilun Du [1] **Karthik Narasimhan** [2]

1 Additional Dynamic Prediction Experiments

ϵ	RCNet	ConvLSTM	SpatialNet (ours)
0	0.0061	0.0026	**0.0024**
0.1	0.0078	0.0030	**0.0026**
0.5	0.0268	0.0072	**0.0062**

Table 1. MSE loss on physics prediction data-set on on single-step prediction with test inputs corrupted with Gaussian noise of magnitude ϵ (model trained with no corruption). Due to its local nature, SpatialNet suffers less form errors in inputs and is able to maintain object numbers/dynamics more consistently even with domain shift.

1.1 Sensitivity to Corruption of Inputs

We investigate the effects of noisy observations in the input domain at test time on both SpatialNet and RCNet, by adding different amounts of Gaussian random noise to input images (Table 1). We find that SpatialNet is more resistant to noise addition. SpatialNet predictions are primarily local, preventing compounding of error from corrupted pixels elsewhere in the image whereas RCNet compresses all pixels into a latent space, where small errors can easily escalate.

1.2 Qualitative visualizations of Generalization Predictions

We provide visualizations of video prediction on each of the generalization datasets in Figure 1 and Figure 2.

1.3 Dataset Generalization.

We test generalization by evaluating on two unseen datasets. For the first, we create a test set where objects are half the size of the training set and initialized randomly with approximately twice the starting velocity. In this new regime, we found that RCNet had a MSE of 0.0115, ConvLSTM has a MSE of 0.0067, while SpatialNet had a MSE of 0.0039. We find RCNet is unable to maintain shapes of the smaller objects, sometimes omitting them, while ConvLSTM maintains shape but is unable to adapt to new dynamics as seen in Figure 1. In contrast, SpatialNet local structure allows it to generate new shapes, and its dynamic seperation allows better generalization. In the second dataset, we explore input size invariance. We create a second testing data-set consisting 16-32 random circles and squares and input images of size 168x168x3 (the density of objects per area is conserved). On this dataset, we obtained a MSE of 0.0042 compared to ConvLSTM of 0.0060, which is comparable to the MSE on the original test dataset of 0.0024, with qualitative images in Figure 2 showing that the spatial memories local structure allows to easily generalize to different input image sizes.

2 PhysWorld

We provide a description of the three games environments in PhysWorld:

PhysGoal: In this environment, an agent has to navigate to a large red goal. Each successful navigation (+1 reward) respawns the red goal at a random location while collision with balls or boxes terminates the episode (-1 reward).

PhysForage: Here, an agent has to collect moving balls while avoiding moving boxes. Each collected ball (+1 reward) will randomly respawn at a new location with a new velocity. Collision with boxes lead to termination of episode (-1 reward).

PhysShooter: In PhysShooter, the agent is stationary and has to choose an angle to shoot bullets. Each bullet travels through the environment until it hits a square (+1 reward) or circle (-1 reward) or leaves the screen. If a moving ball or box hits the agent (-1 reward), the episode is terminated. After firing a bullet, the agent cannot fire again until the bullet disappears.

Examples of agents playing the PhysWorld environments are given in Figure 3.

[1]Massachusetts Institute of Technology [2]Princeton University. Correspondence to: Yilun Du <yilundu@gmail.com>, Karthik Narasimhan <karthikn@cs.princeton.edu>.

Proceedings of the 36^{th} International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019.

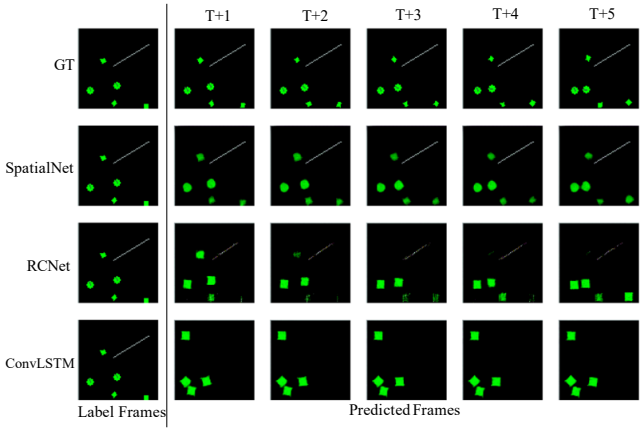

Figure 1. Predictions of SpatialNet, RCNet on test data-set with objects twice as small and with twice the movement speed as trained on. All shown frames are one step predictions. SpatialNet is able to accurately generalize to smaller, faster objects while RCNet is unable to generate the shapes of the smaller objects and suffers from background degradation and ConvLSTM is unable to maintain shapes and dynamics.

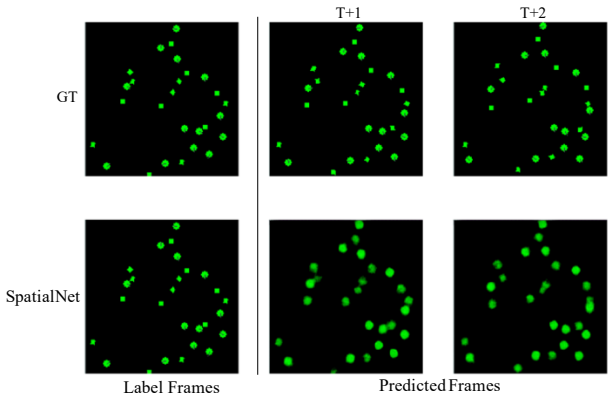

Figure 2. Predictions of SpatialNet on input images of 168 x 168 when SpatialNet was trained on 84 x 84 images. Prediction shown are 1 step future predictions. SpatialNet is able to maintain physical consistency in at large input sizes.

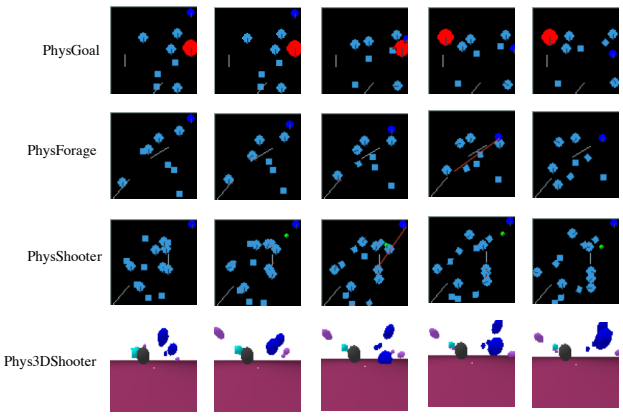

Figure 3. Example agent game-play in each of the PhysWorld environments. In PhysGoal the dark blue agent attempts to reach a red goal while avoiding moving objects. In PhysForage the dark blue agent attempts to gather light blue circles while avoiding squares. In PhysShooter, the dark blue agent is immobile and chooses to fire bullet a green bullet at squares while avoiding circles. In Phys3DShooter, the grey fires turqoise bullets at purple spheres while avoiding blue spheres.

2.1 SpatialNet Predictions

Figure 4 shows the qualitative next 3 frame predictions of SpatialNet on each of the different PhysWorld environment with the first frame being the current observation. In PhysGoal, SpatialNet is able to infer the movement of the obstacles, the dark blue agent, and the red goal after agent collection. In PhysGather, SpatialNet is able to infer movement of obstacles as well as the gather of a circle. In PhysShooter, SpatialNet is able to anticipate a collision of the bullet with a moving obstacle and further infer the shooting of a green bullet by the agent.

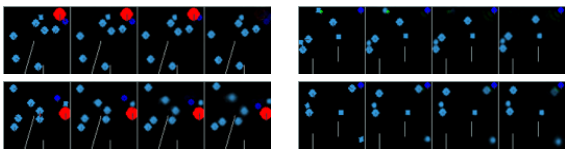

Figure 4. Future image prediction on PhysGoal (left) and PhysShooter (right). First image is current observation, the next three are predicted. SpatialNet is able to predict future dynamics of boxes and balls and anticipate agent movement (PhysGoal) and agent shooting (PhysShooter).

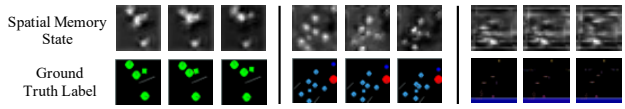

Figure 5. Visualization of SpatialNet hidden state on PhysVideos (left), PhysGoal (middle) and Atari DemonAttack (right). Hidden state has high activations for moving objects while background objects such as walls (left), red goals (middle) and platforms (right) are not attended to as much.

2.2 Visualization of Spatial Memory

We provide visualization of the values of spatial memory hidden state while predicting future frames. We visualize the values of spatial memory on PhysVideos, PhysGoal and the Atari environment Demon Attack in Figure 5. To visualize, we take the mean across the channels of each grid pixel in the spatial memory hidden state. We find strong correspondence between high activation regions in the spatial memory and dynamic objects in the associated ground label of the dynamic objects. We further find that static background, such as walls in the input, goals and platforms appear to be passed along in input features.

3 Additional Atari experiments

We provide plots of training curves on all Atari environments in Figure 6 on provide on quantitative numbers in Figure 2.

Predictions on Atari We also investigate the benefits (in terms of MSE) of initializing SpatialNet pretrained on the physics dataset compared to training with scratch in Figure 3. We evaluate the MSE error at 1 million frames and find that initializing with the physics dataset provides a 12.9% decrease in MSE error. We find that pretraining helps on 7 of the 10 Atari environments, with the most negatively impacted environment being Enduro, a 3D racecar environment in which the environmental prior encoded by the physics dataset may be detrimental. More significant gains in transfer may be achievable by using a large online database of 2D YouTube videos which cover even more of diversity of games.

SpatialNet Predictions We further visualize qualitative results on SpatialNet on training Atari in Figure 7. In general, across the Atari Suite, we found that SpatialNet is able to accurately model both the environment and agents behavior. In the figure, we seed that SpatialNet is able to accurately predict agent movement and ice block movement in Frostbite. On DemonAttack, SpatialNet is able to infer the falling of bullets. On Asteroid, SpatialNet is able to infer the movement of asteroids. Finally, on FishingDerby, SpatialNet is able to the right player capturing a fish and also predict that the left player is likely to catch a fish (indicated by the blurriness of the rod). We note that any blurriness in predicted output may in fact even be beneficial to the policy, since a policy can learn to interpret the input.

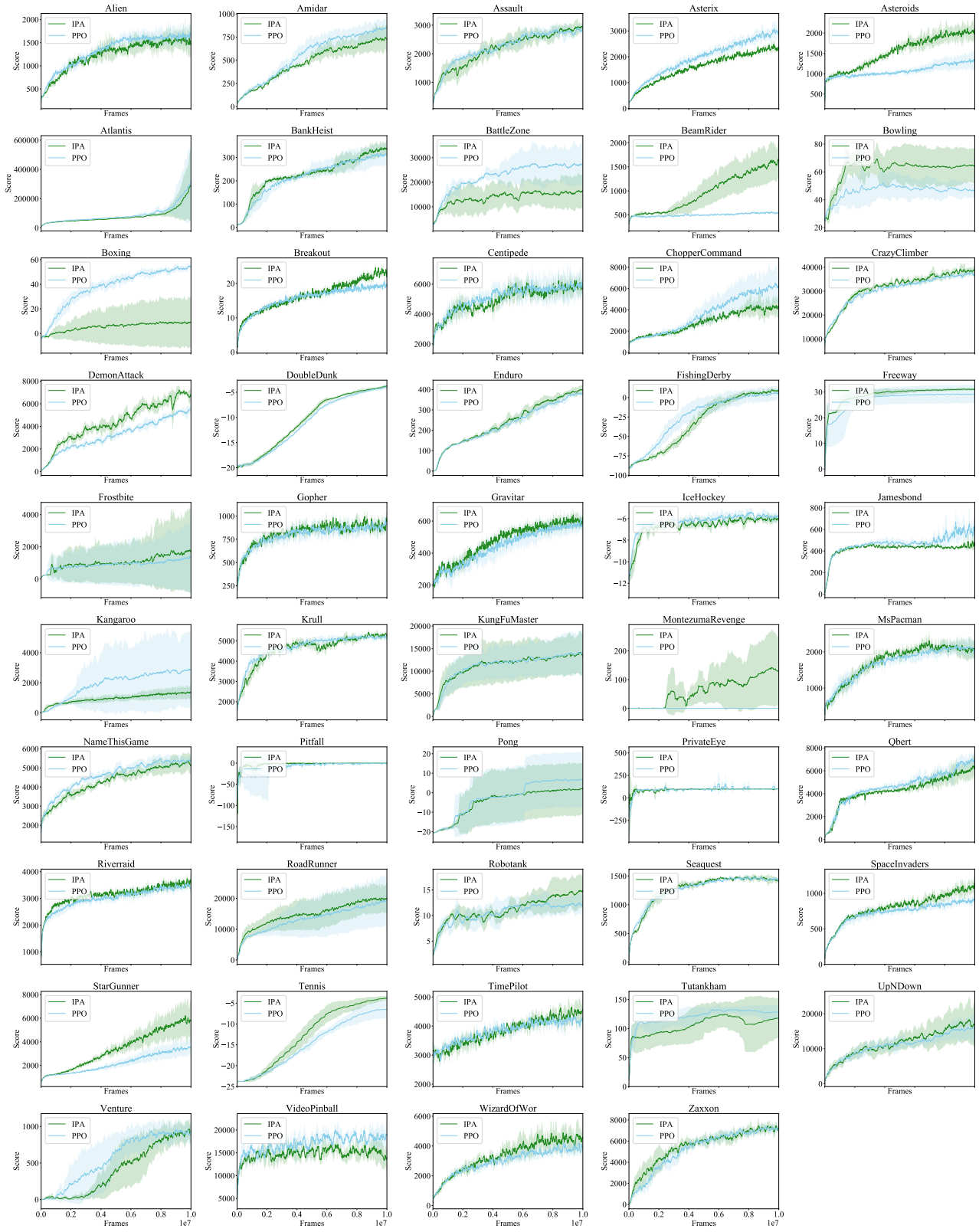

Figure 6. Plots of policy performance trained with either PPO or IPA on all Atari environments on 5 different seeds. IPA sometimes leads to low learning early on the training due to rapid change of 3 predicted future frames. However, later on in training in many different environments, IPA provides performance gains by giving policies future trajectories.

Environment	PPO	D2A
Alien	**1668.6** ± 224.3	1485.5 ± 281.0
Amidar	**855.9** ± 98.6	725.5 ± 135.0
Assault	2939.2 ± 153.2	**2968.4** ± 124.0
Asterix	**2920.8** ± 287.3	2334.0 ± 184.0
Asteroids	1321.0 ± 233.5	**2098.4** ± 102.0
Atlantis	**323205.4** ± 277643.2	289369.8 ± 239469.0
BankHeist	310.4 ± 44.0	**334.3** ± 29.0
BattleZone	**26828.0** ± 8472.0	16526.7 ± 6986.0
BeamRider	553.1 ± 28.4	**1630.3** ± 400.0
Bowling	46.6 ± 5.2	**64.3** ± 13.0
Boxing	**54.3** ± 2.5	8.9 ± 20.0
Breakout	19.7 ± 0.9	**23.4** ± 1.0
Centipede	**6043.7** ± 990.6	6032.5 ± 199.0
ChopperCommand	**6549.4** ± 1779.1	4112.0 ± 1024.0
CrazyClimber	36893.2 ± 463.9	**38499.0** ± 1221.0
DemonAttack	5510.9 ± 412.5	**6793.6** ± 558.0
DoubleDunk	-4.0 ± 0.5	**-3.8** ± 0.0
Enduro	376.7 ± 10.5	**398.6** ± 23.0
FishingDerby	6.7 ± 10.1	**9.3** ± 3.0
Freeway	29.2 ± 3.6	**31.2** ± 1.0
Frostbite	1342.5 ± 2154.5	**1701.1** ± 2485.0
Gopher	904.0 ± 42.3	**941.1** ± 56.0
Gravitar	574.9 ± 36.2	**627.2** ± 25.0
IceHockey	**-5.9** ± 0.3	-6.1 ± 0.0
Jamesbond	**598.9** ± 112.1	454.3 ± 34.0
Kangaroo	**2842.4** ± 2461.2	1373.0 ± 445.0
Krull	5178.9 ± 205.1	**5219.3** ± 129.0
KungFuMaster	**13831.6** ± 4483.6	13358.5 ± 4352.0
MontezumaRevenge	0.0 ± 0.0	**129.7** ± 122.0
MsPacman	1990.1 ± 227.9	**2097.3** ± 259.0
NameThisGame	**5406.4** ± 278.0	5131.3 ± 427.0
Pitfall	-0.1 ± 0.3	**0.0** ± 0.0
Pong	**6.6** ± 14.1	2.2 ± 13.0
PrivateEye	95.6 ± 5.4	**99.6** ± 0.0
Qbert	**6981.0** ± 548.0	6331.4 ± 769.0
Riverraid	3411.0 ± 201.9	**3612.4** ± 130.0
RoadRunner	19329.6 ± 8472.6	**20041.8** ± 4906.0
Robotank	11.9 ± 1.8	**14.9** ± 3.0
Seaquest	**1426.0** ± 43.5	1408.7 ± 51.0
SpaceInvaders	902.4 ± 66.0	**1132.6** ± 101.0
StarGunner	3450.0 ± 801.5	**5778.5** ± 1584.0
Tennis	-6.5 ± 2.1	**-3.8** ± 1.0
TimePilot	4281.8 ± 126.6	**4580.0** ± 314.0
Tutankham	**128.5** ± 12.3	118.2 ± 35.0
UpNDown	15872.3 ± 3995.3	**16913.7** ± 6344.0
Venture	930.2 ± 137.9	**946.7** ± 167.0
VideoPinball	**18878.1** ± 1251.7	13981.2 ± 2136.0
WizardOfWor	3835.6 ± 404.7	**4629.8** ± 662.0
Zaxxon	7197.4 ± 220.6	**7271.0** ± 264.0

Table 2. Scores obtained on Stochastic Atari Environments with *sticky actions* (actions repeated with 50% probability at each step). Scores are average performance over 100 episodes after 10M training frames, over 5 different random seeds.

Environment	MSE PD	MSE DN	Percent Advantage
Assault	0.00477	0.00522	9.4%
Asteroids	0.002506	0.002518	4.7%
Breakout	0.000417	0.000423	1.4%
DemonAttack	0.00433	0.00562	29.8%
Enduro	0.00576	0.00411	-28.7%
FishingDerby	0.00183	0.00192	4.9%
Frostbite	0.000965	0.00107	10.8%
IceHockey	0.000614	0.0013	111.7%
Pong	0.00636	0.00584	-8.2%
Tennis	0.00142	0.00132	-7.1%

Table 3. MSE on Stochastic Atari Environments (a action is repeated with a geometric distribution with p=0.5) at 1 million training frames. MSE PD is trained with a model from physics dataset while MSE DN is trained with a model from scratch. We evaluate percentage advantage for initializing with a physics dataset as compared to from scratch. We average 12.9% decrease in MSE error using a initialization from pretraining on a physics dataset. The most negative environment, Enduro, involves a 3D landspace which initializing from model trained on a physics data set may be detrimental.

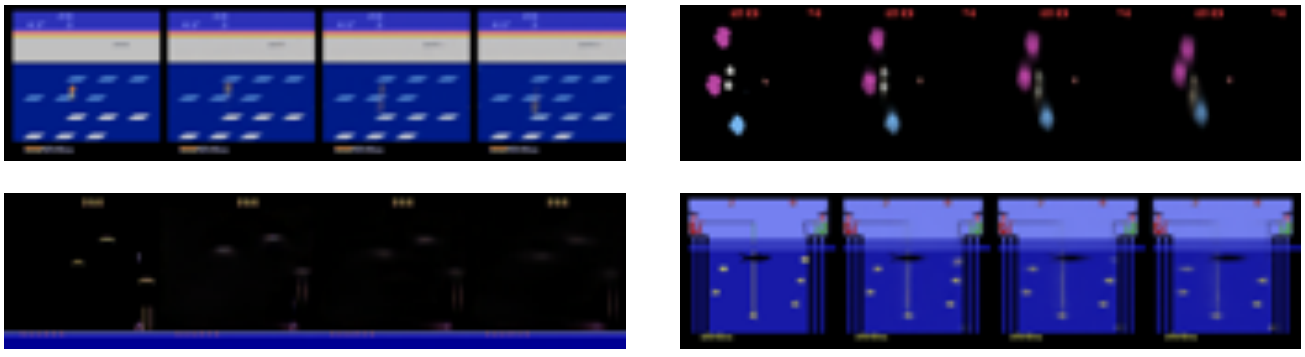

Figure 7. Visualization of model future state prediction on 4 games in Atari (Frostbite - upper left, DemonAttack - lower left, Asteroids - upper right, FishingDerby - lower right). SpatialNet is able to predict falling of bullets, the catching of fish, movement of asteroids, and the movement of tiles/future agent movement in different environments. First frame visualized is ground truth observation, next 3 frames are model future frame predictions.