



VS Lab

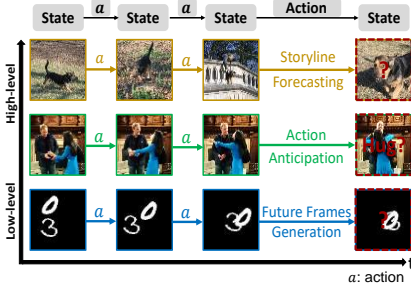
# Visual Forecasting by Imitating Dynamics in Natural Sequences

Kuo-Hao Zeng<sup>++</sup>, William B. Shen<sup>+</sup>, De-An Huang<sup>+</sup>, Min Sun<sup>+</sup>, Juan Carlos Nieves<sup>+</sup>  
<sup>+</sup>Stanford University <sup>++</sup>National Tsing Hua University



## Visual Forecasting

Forecast future by visual input at multiple levels.

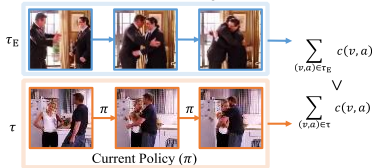


- General framework: Single approach applies to multiple tasks with different levels of abstractions.
- Expert trajectories: Natural sequence. E.g., natural videos, photo streams.
- Visual GAIL: Visual Generative Adversarial Imitation Learning.
- Efficient sampling: Eschew expensive sampling for possible future states.

## Problem Formulation

Model visual forecasting as an imitation learning task.

Training



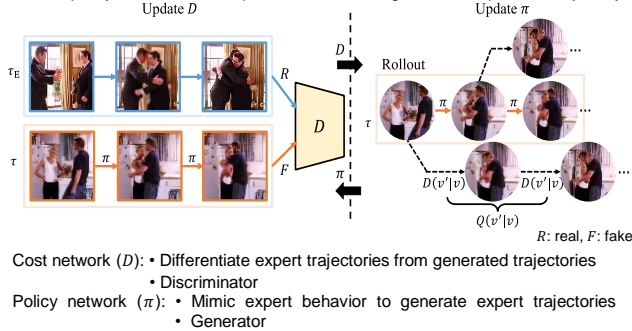
Testing



Training: Learn expert behavior from expert trajectories.  
Testing: Evaluate multiple tasks on expert trajectories.

## Visual Imitation Learning

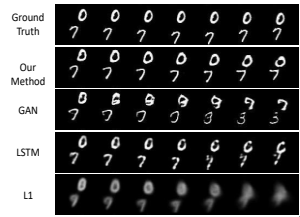
Learn a policy which mimics expert's behavior and generates expert-like trajectory.



## Experiments

### Future Frame Generation

Human Evaluation (%)	Real sequence
Real sequence	50.0
Our method	<b>15.1</b>
GAN	10.2
LSTM [1]	1.4
L1	0.3



Acknowledgement:



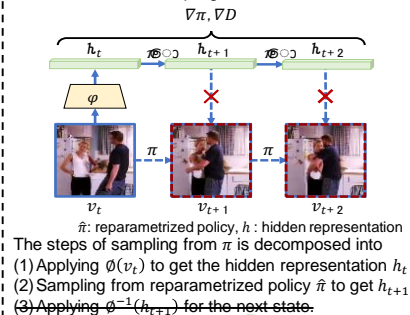
### Action Prediction

Model	Accuracy (%)
Random	25.0
SVM static	36.2
SVM	35.8
Deep Regression (K = 1) [2]	37.5
LSTM [1]	40.5
Pixel Synthesis	26.1
Ours	<b>45.8</b>
Deep Regression (K = 3) [2]	44.0



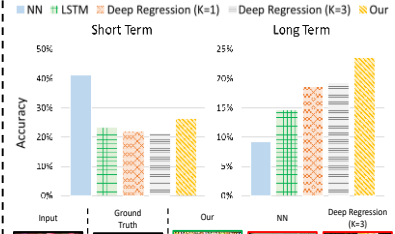
## Efficient Sampling

Make future state sampling efficient.



Not required!

### Storyline Forecasting



[1] N. Srivastava, E. Mansimov, and R. Salakhutdinov. Unsupervised learning of video representations using lstms. In *ICML*, 2015.  
 [2] C. Vondrick, H. Pirsiavash, and A. Torralba. Anticipating visual representations from unlabeled video. In *CVPR*, 2016.