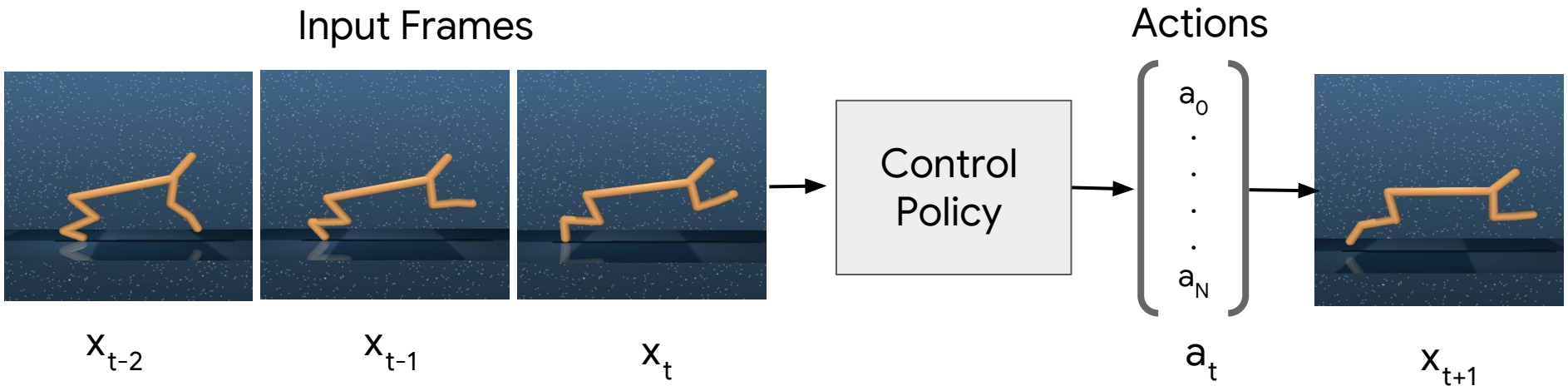
Self-Supervised Representation Learning for Continuous Control

Debidatta Dwibedi
Jonathan Tompson
Corey Lynch
Pierre Sermanet

Google

Problem Setting

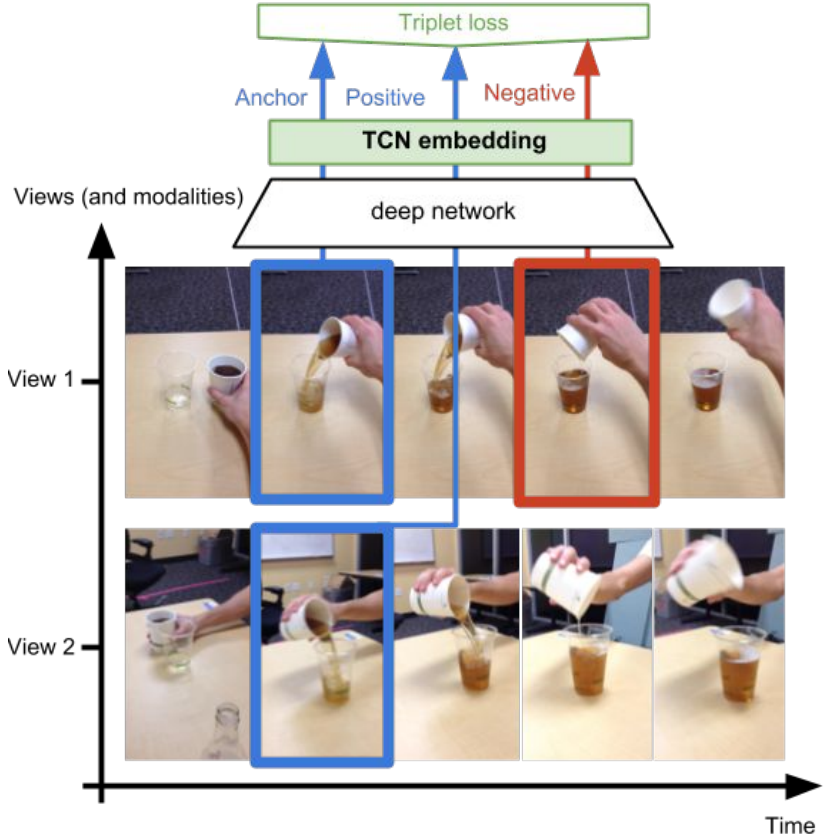


Continuous Control from Visual Input

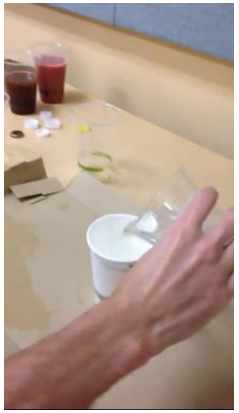
Visual Representation Learning	Control Policy Learning
Time-Contrastive Networks [1]	Proximal Policy Optimization [2]

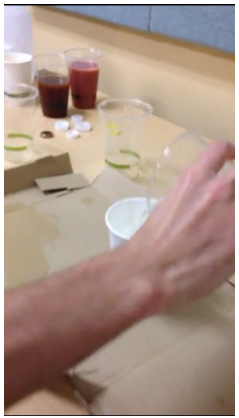
[1] P. Sermanet, C. Lynch, Y. Chebotar, J. Hsu, E. Jang, S. Schaal, and S. Levine, “Time-contrastive networks: Self-supervised learning from video”
[2] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms”

Time Contrastive Networks



Can we differentiate reliably based on only one frame?



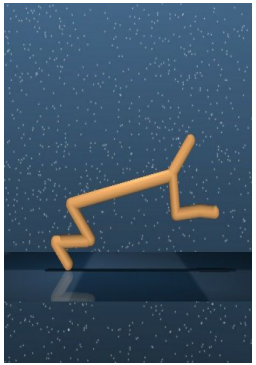


Visually similar frames may be separated in time

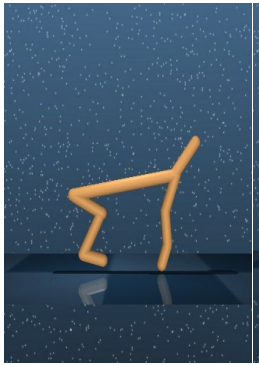
Can we differentiate reliably based on only one frame?

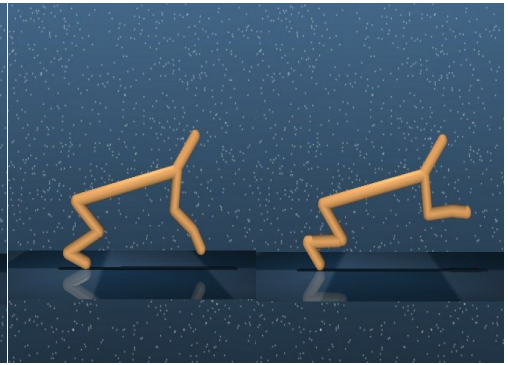
Much easier to differentiate short clips

Multiple frames enable learning of motion cues



v/s

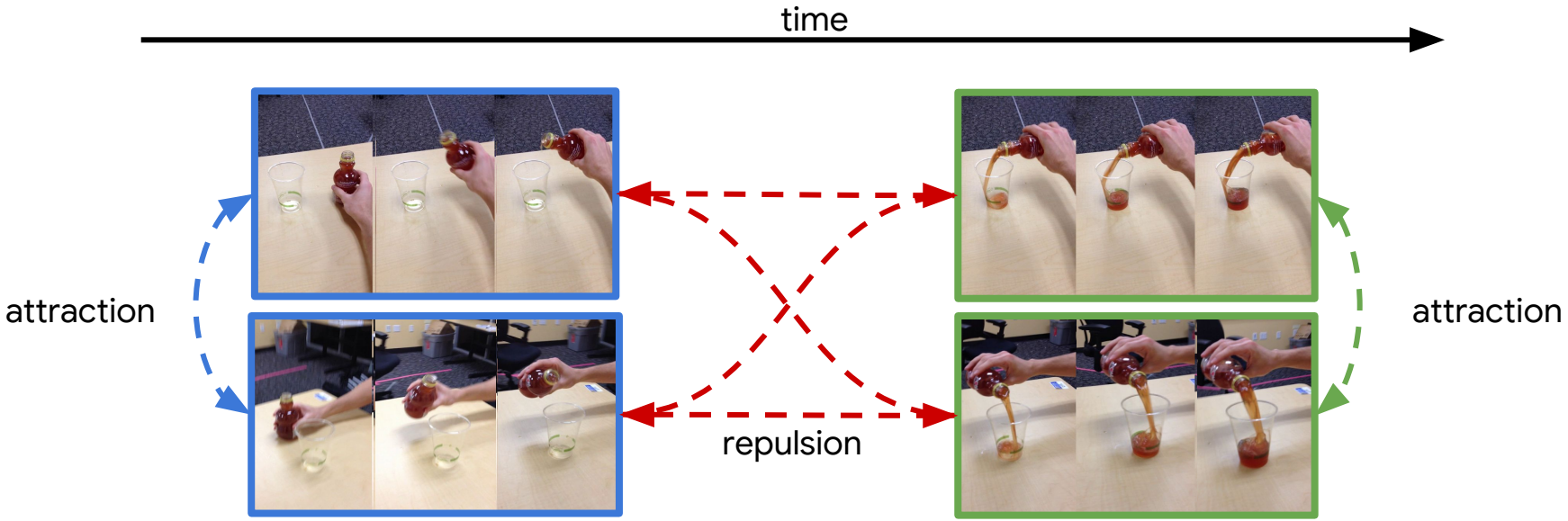




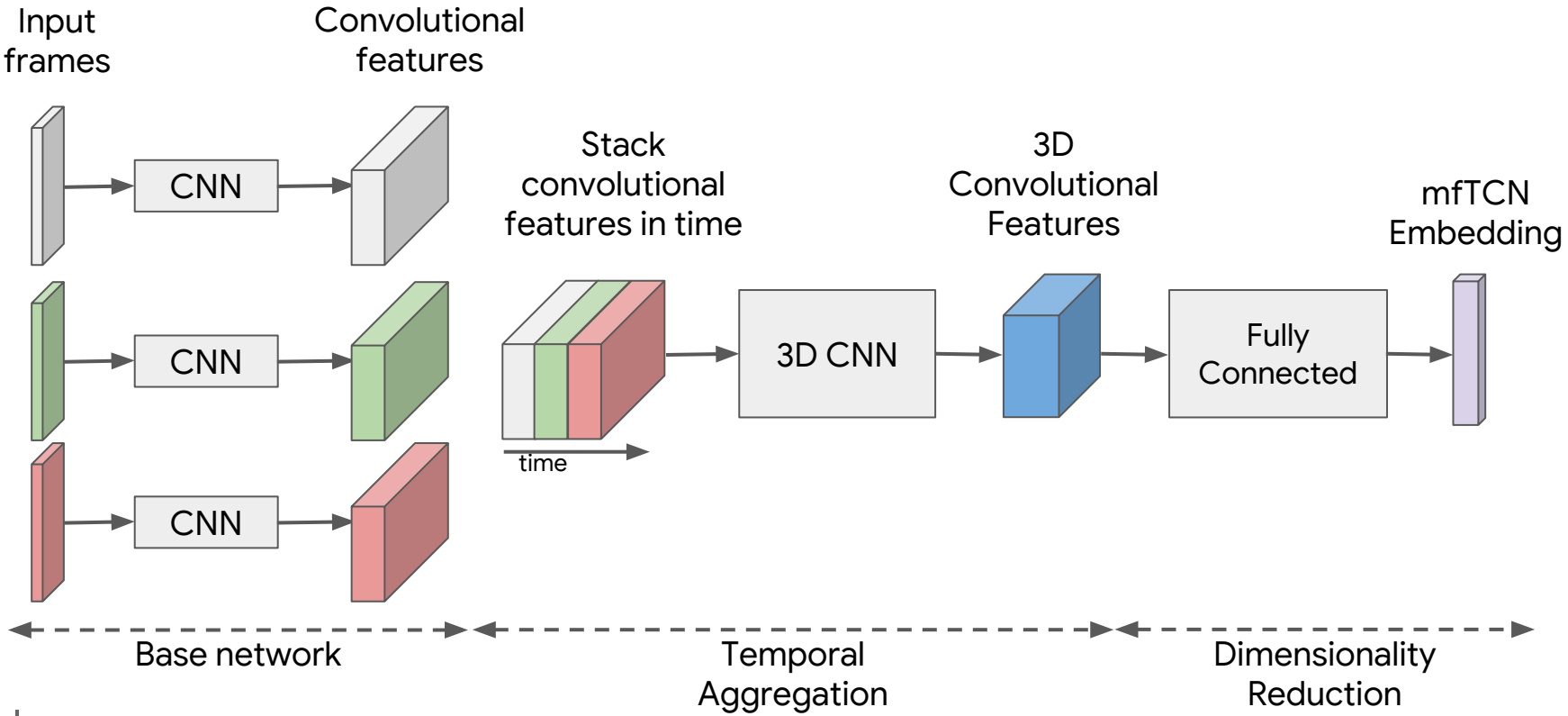
Additional context useful in encoding motion cues like velocity

Our Solution: Multi-frame TCN

Multi-frame TCN: Sampling Training Tuples

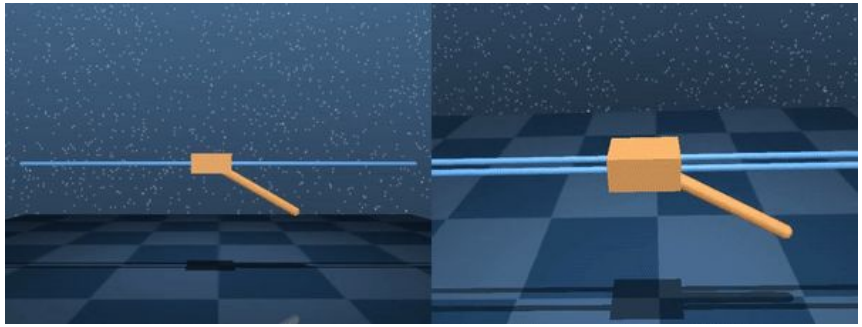


Multi-frame TCN: Architecture



Results

CartPole Dataset



View 1 View 2

Position Attributes:

1. Position of Cart
2. Angle of Pole with the horizontal line

Velocity Attributes:

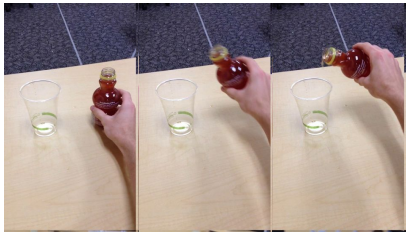
1. Velocity of Cart
2. Angular velocity of Pole

Effect of Embedding Multiple Frames

Number of frames	Position MSE (x Std. Dev.)	Velocity MSE (x Std. Dev.)
1	0.0052	0.2201
2	0.0019	0.0974
3	0.0014	0.0550
4	0.0013	0.0476

Results on the CartPole dataset

Pouring Dataset



1st Person View

3rd Person View

Static Attributes:

1. Is there liquid in cup?
2. Is hand within pouring distance?
3. What is the tilt angle of container?
4. Is liquid flowing?
5. Is hand in contact with container?

Motion Attributes:

1. Is the hand reaching towards the container?
2. Is the hand receding away?
3. Is the bottle going up?
4. Is the bottle coming down?

Effect of Embedding Multiple Frames

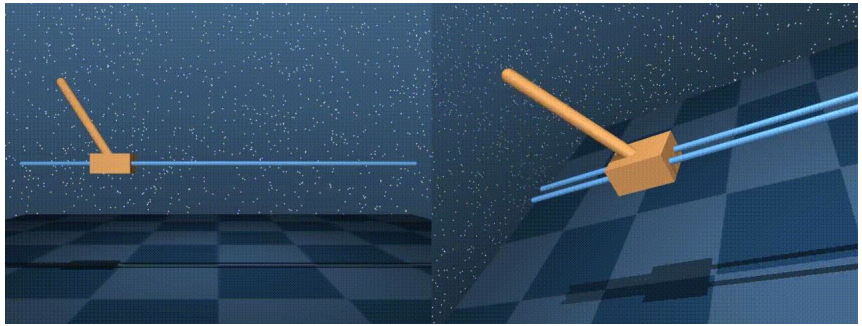
Number of frames	Alignment Error (%)	Static Error (%)	Motion Error (%)
1	16.21	18.92	30.17
8	14.27	17.25	24.83
16	11.29	16.79	18.30
32	8.86	19.36	20.88

Results on the Pouring dataset

Can we learn to control using these embeddings?

CartPole Environment

Agent takes random actions and observes itself



View 1 View 2

View 1 View 2

CartPole Results

PPO on true state

PPO on learned visual representations

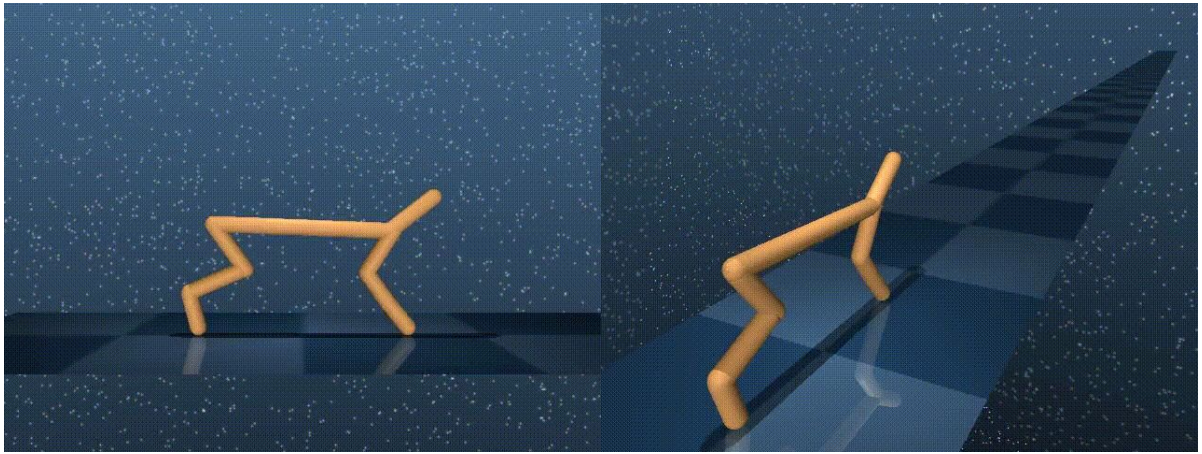
Quantitative Results

Input to PPO	Avg of 100 runs
Random State	121.45
True State	861.41
Raw Pixels	283.82
Position Velocity Encoders	457.27
mfTCN	787.47
mfTCN (Moving Cam)	811.10

Results on the CartPole dataset

Cheetah Environment

Agent observes another agent demonstrating an action



View 1 View 2

Cheetah Results

PPO on true state

PPO on learned visual representations

Quantitative Results

Input to PPO	Avg of 100 runs
Random State	28.31
True State	390.16
Raw Pixels	146.14
mfTCN	360.50

Results on the Cheetah dataset

Takeaways

1. TCN embeddings can be used to perform continuous control from pixels
2. Embedding multiple frames helps in reasoning about motion cues

More details at the poster