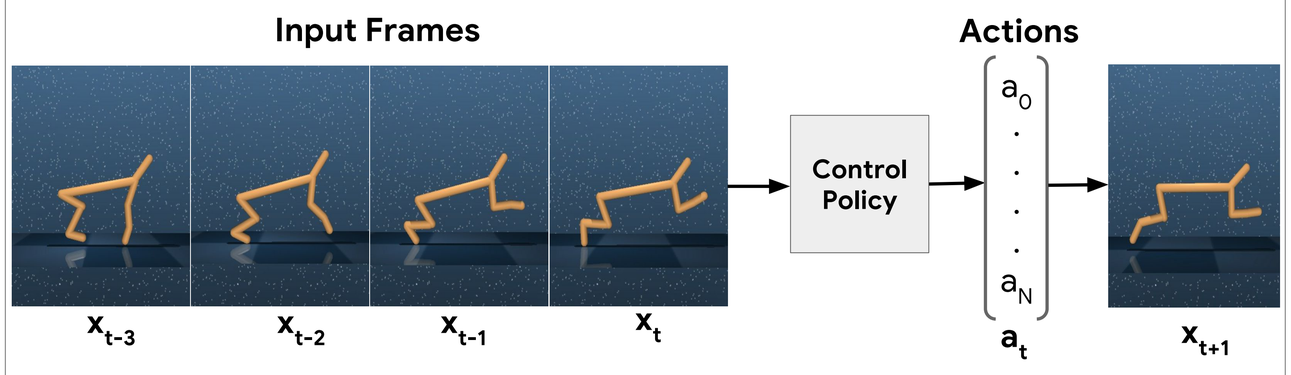
Self-Supervised Representation Learning for Continuous Control

Debidatta Dwibedi, Jonathan Tompson, Corey Lynch, Pierre Sermanet

Problem Setting

Continuous Control using Visual Input



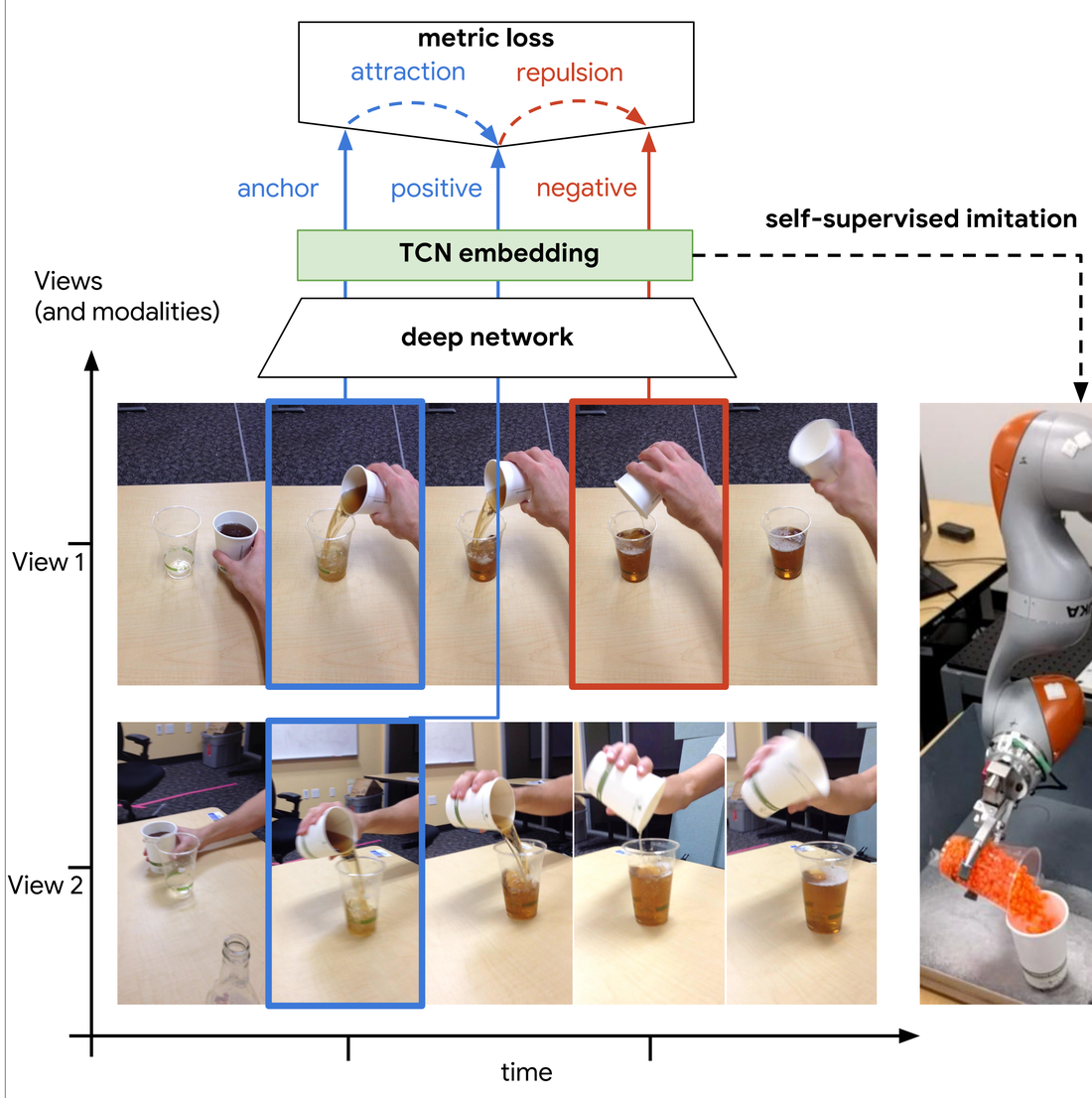
Research Questions

1. How can we learn good representations from video for model free RL?
2. How do we encode motion features in our representations?

Related Work

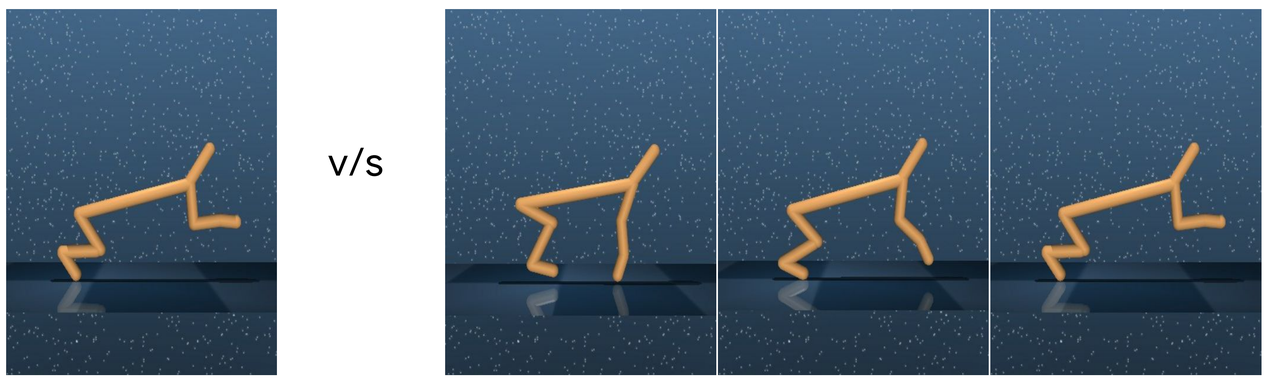
Previous work[1,2,3,4] on learning good visual representations for continuous control via self-supervised approaches have focused on inducing desirable properties like local linear dynamics or view invariance on the learned embeddings. In this work, we enable learning of representations that encode better motion features.

Time Contrastive Networks (TCN) [4]



Why Encode Motion in Embeddings?

At inference time, multiple frames encode motion features



v/s

Difficult for single frame visual embeddings to encode motion features like velocity

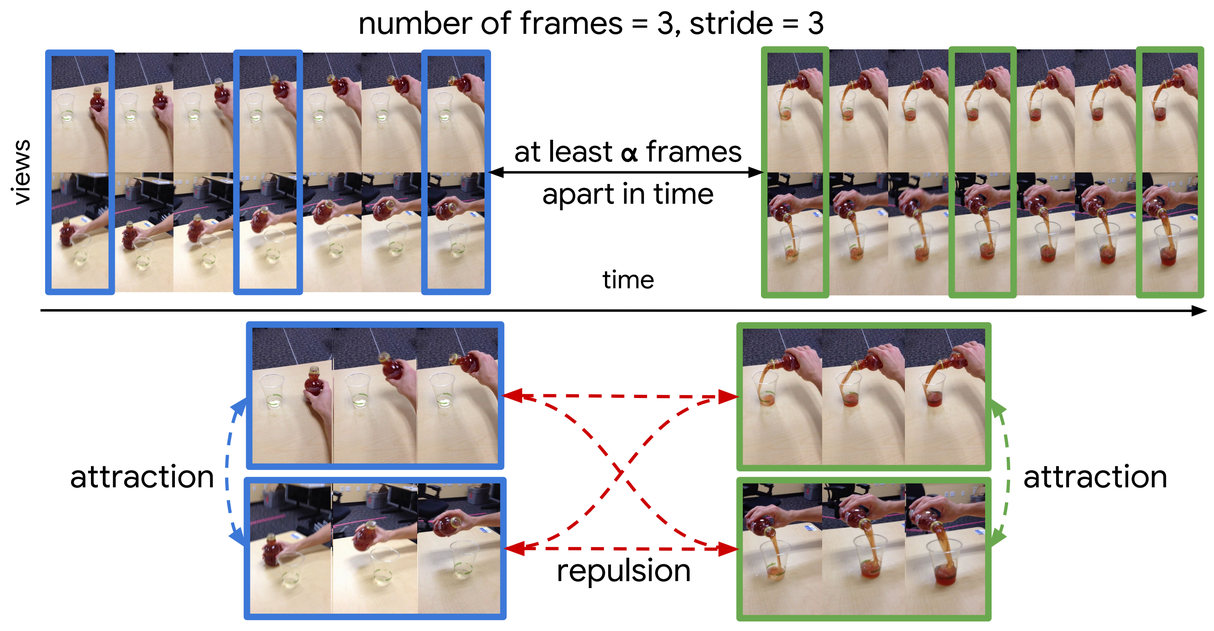
During training, multiple frames provide more context

v/s

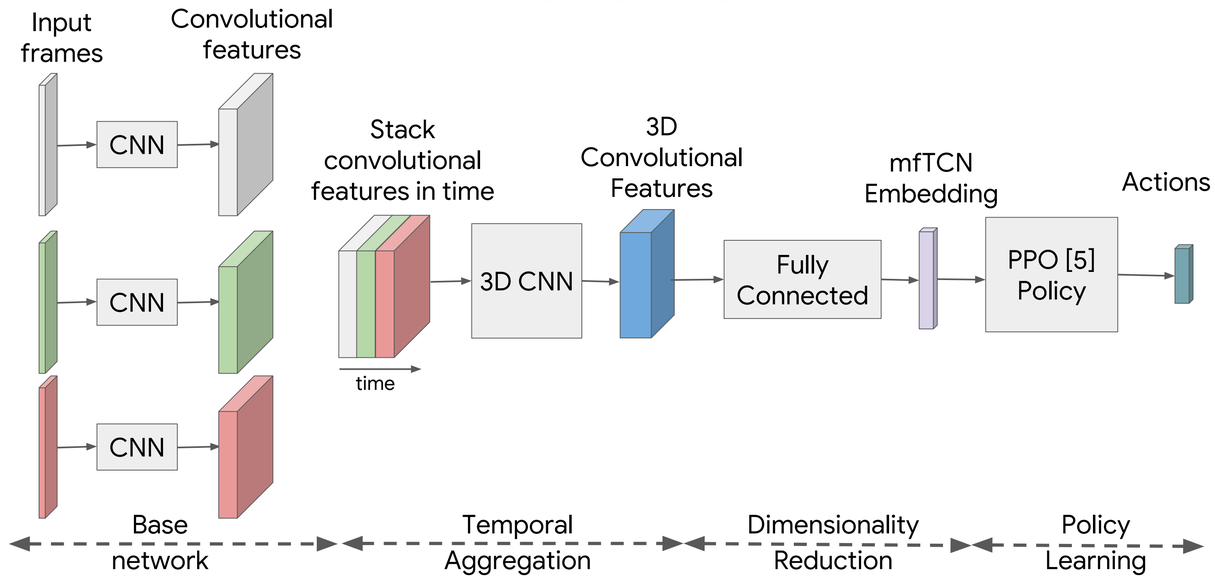
Frame pairs on the left are ambiguous as they are visually similar but at different time steps. Neighbouring frames not only help the model disambiguate between such frames, but also reason about occlusion and motion cues.

Our Solution: Multi-frame TCN (mfTCN)

Sampling Anchor, Positive, and Negative Tuples



Architecture



Results

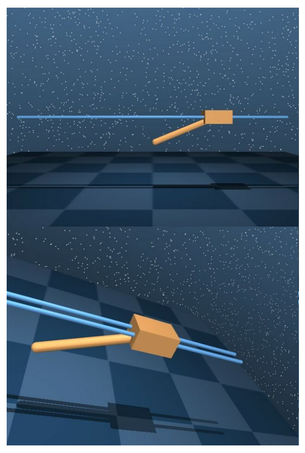
Pouring dataset: Alignment between Views / Attributes Classification

Method	# of frames	Alignment Error (%)	Static Error (%)	Motion Error (%)
TCN [4]	1	16.21	18.92	30.17
mfTCN (ours)	8	14.27	17.25	24.83
mfTCN (ours)	16	11.29	**16.79**	**18.30**
mfTCN (ours)	32	**8.86**	19.36	20.88

Regression to Position and Velocity of Cartpole from Embeddings

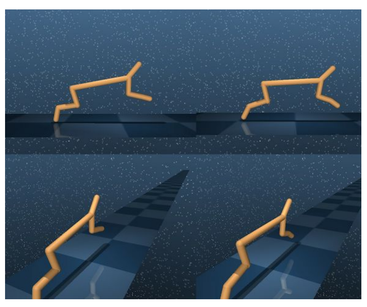
Method	# of frames	Position MSE (x Std. Dev.)	Velocity MSE (x Std. Dev.)
TCN [4]	1	0.0052	0.2201
mfTCN (ours)	2	0.0019	0.0974
mfTCN (ours)	3	0.0014	0.0550
mfTCN (ours)	4	**0.0013**	**0.0476**

Performance on Controlling CartPole



Input to PPO	Avg of 100 runs
Random State	121.5
True State	861.4
Raw Pixels	283.8
PVE [2]	457.3
TCN	759.3
TCN (Moving Cam)	691.7
mfTCN	787.5
mfTCN (Moving Cam)	811.1

Performance on Controlling Cheetah



Input to PPO	Avg of 100 runs
Random State	28.3
True State	390.2
Raw Pixels	146.1
mfTCN	360.5

References

[1] M. Watter, J. Springenberg, J. Boedecker, and M. Riedmiller, “Embed to control: A locally linear latent dynamics model for control from raw images”
[2] R. Jonschkowski, R. Hafner, J. Scholz, and M. Riedmiller, “Pves: Position-velocity encoders for unsupervised learning of structured state representations”
[3] C. Finn, X. Y. Tan, Y. Duan, T. Darrell, S. Levine, and P. Abbeel, “Deep spatial autoencoders for visuomotor learning”
[4] P. Sermanet, C. Lynch, Y. Chebotar, J. Hsu, E. Jang, S. Schaal, and S. Levine, “Time-contrastive networks: Self-supervised learning from video”
[5] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms”