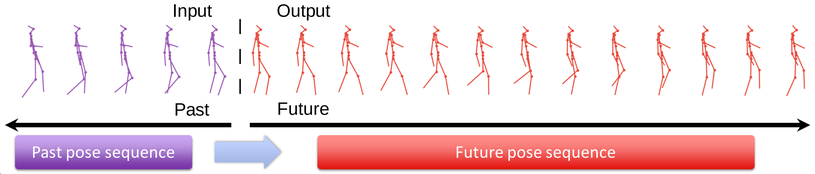
Action-Agnostic Human Pose Forecasting

Hsu-kuang Chiu, Ehsan Adeli, Borui Wang, De-An Huang, Juan Carlos Niebles
Stanford University

IEEE 2019 **W**inter Conference on **A**pplications of **C**omputer **V**ision

Problem Statement

Given the past pose sequence as the input, we would like to forecast the future pose sequence.



Motivations

Unlike previous works, we would like to build a model that does not require action labels while performing well for both short-term and long-term forecasting simultaneously.

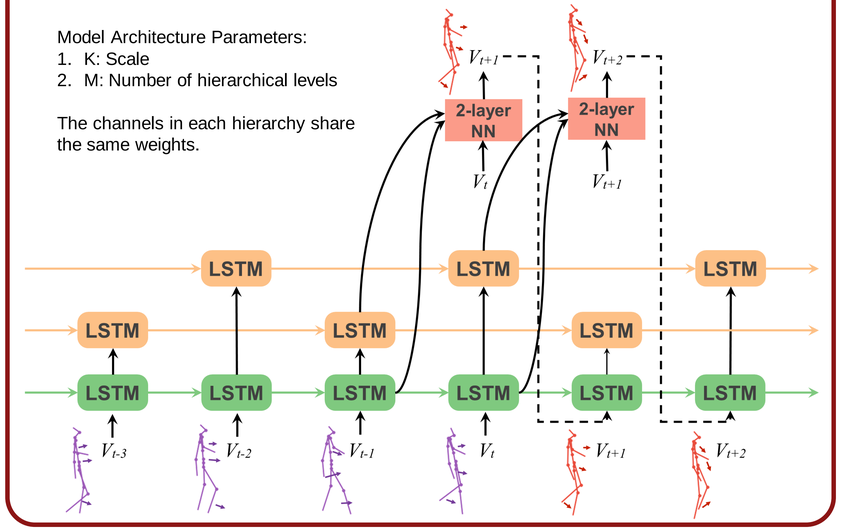
No dependence on action labels, which are unavailable in real-world scenarios

Perform well on both short-term and long-term forecasting

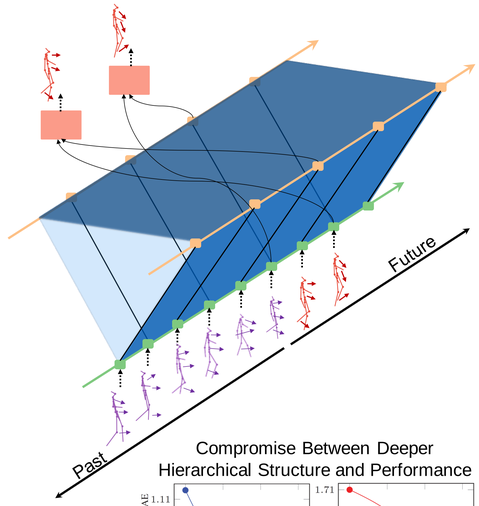
Triangular-Prism RNN (TP-RNN) Model

Model Architecture Parameters:
1. K: Scale
2. M: Number of hierarchical levels

The channels in each hierarchy share the same weights.



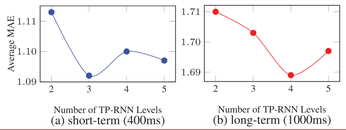
Triangular-Prism RNN (TP-RNN) Model Design



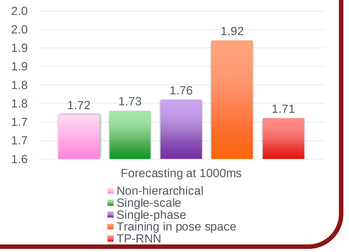
Design Intuitions:
- Human motions have hierarchical and multi-scale structures.
- Velocity information can be used to better predict the future.

Model Design Components:
- Hierarchical structures
- Multi-scale structures
- Multi-phase structure for data augmentation
- Velocity space operations

Compromise Between Deeper Hierarchical Structure and Performance



Importance of each component
Average MAE over all 15 actions of Human 3.6M Dataset

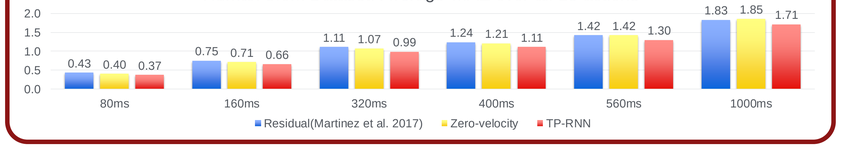


Results on Human 3.6M Dataset

Human 3.6M Dataset: Evaluated Using the Mean Angle Errors (MAE), AA: Action-Agnostic

	AA	80ms	160ms	320ms	400ms	560ms	1000ms
Residual (short)	×	**0.36**	0.67	1.02	1.15	-	-
Residual (short)	✓	0.39	0.72	1.08	1.22	-	-
Residual (long)	×	0.43	0.75	1.11	1.24	1.42	1.83
Residual (long)	✓	0.42	0.73	1.09	1.23	1.42	1.84
Zero-velocity	-	0.40	0.71	1.07	1.21	1.42	1.85
TP-RNN (Ours)	✓	0.37	**0.66**	**0.99**	**1.11**	**1.30**	**1.71**

Human 3.6M Dataset: Average MAE over all 15 actions

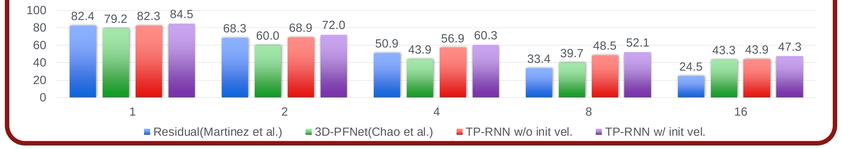


Results on Penn Action Dataset

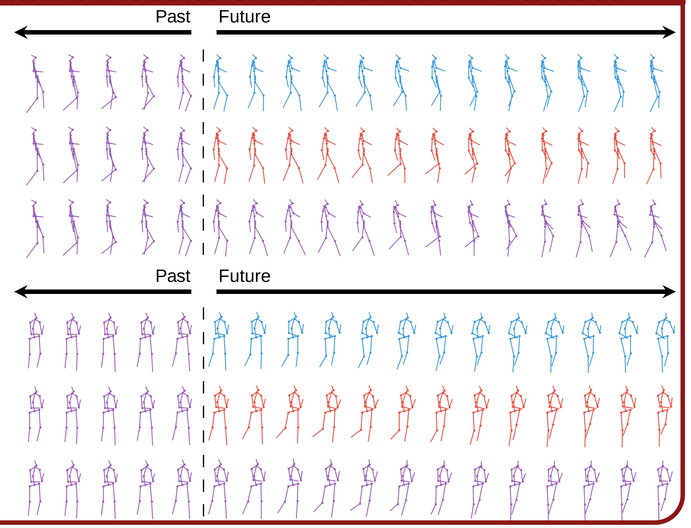
Penn Action Dataset: Evaluated Using the Percentage of Correct Keypoints (PCK)

Future Frame	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Residual	82.4	68.3	58.5	50.9	44.7	40.0	36.4	33.4	31.3	29.5	28.3	27.3	26.4	25.7	25.0	24.5
3D-PFNet	79.2	60.0	49.0	43.9	41.5	40.3	39.8	39.7	40.1	40.5	41.1	41.6	42.3	42.9	43.2	43.3
TP-RNN w/o init vel.	82.3	68.9	61.5	56.9	53.9	51.7	50.0	48.5	47.3	46.2	45.6	45.0	44.6	44.3	44.1	43.9
TP-RNN w/ init vel.	**84.5**	**72.0**	**64.8**	**60.3**	**57.2**	**55.0**	**53.4**	**52.1**	**50.9**	**50.0**	**49.3**	**48.7**	**48.3**	**47.9**	**47.6**	**47.3**

Penn Action Dataset: PCK@0.05 at future timesteps



Qualitative Results on Human 3.6M Dataset



Conclusion

- We proposed an **action-agnostic** model for **both short-term and long-term** forecasting.
- Our model was trained in the **velocity space**.
- Our Triangular-Prism RNN outperforms the previous state-of-the-art.

We thank **Panasonic** for the support