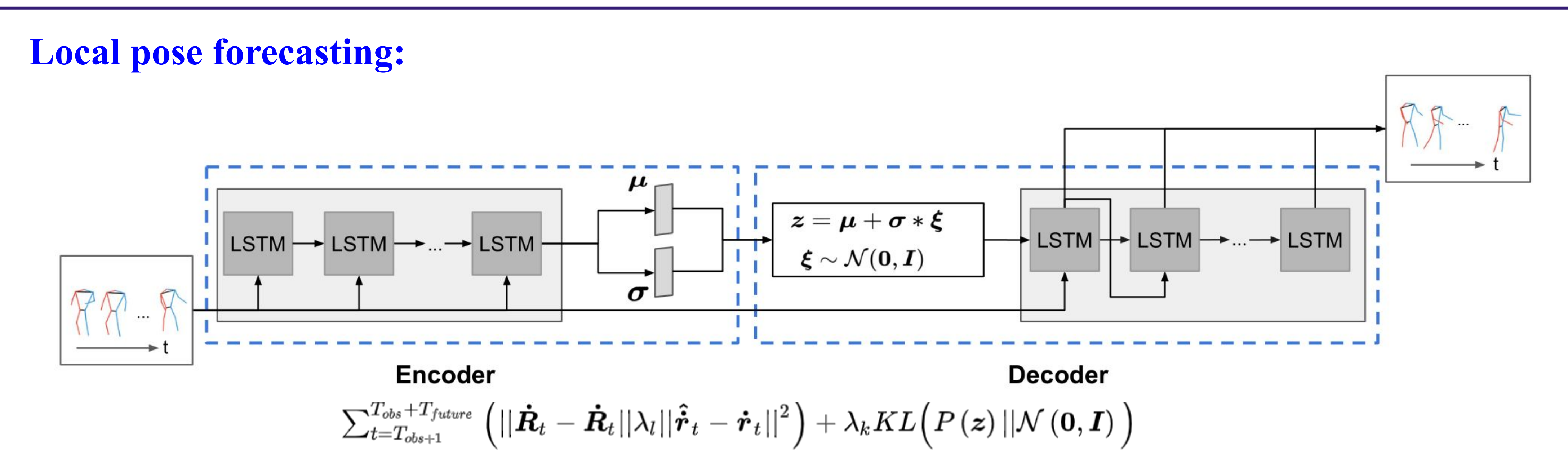
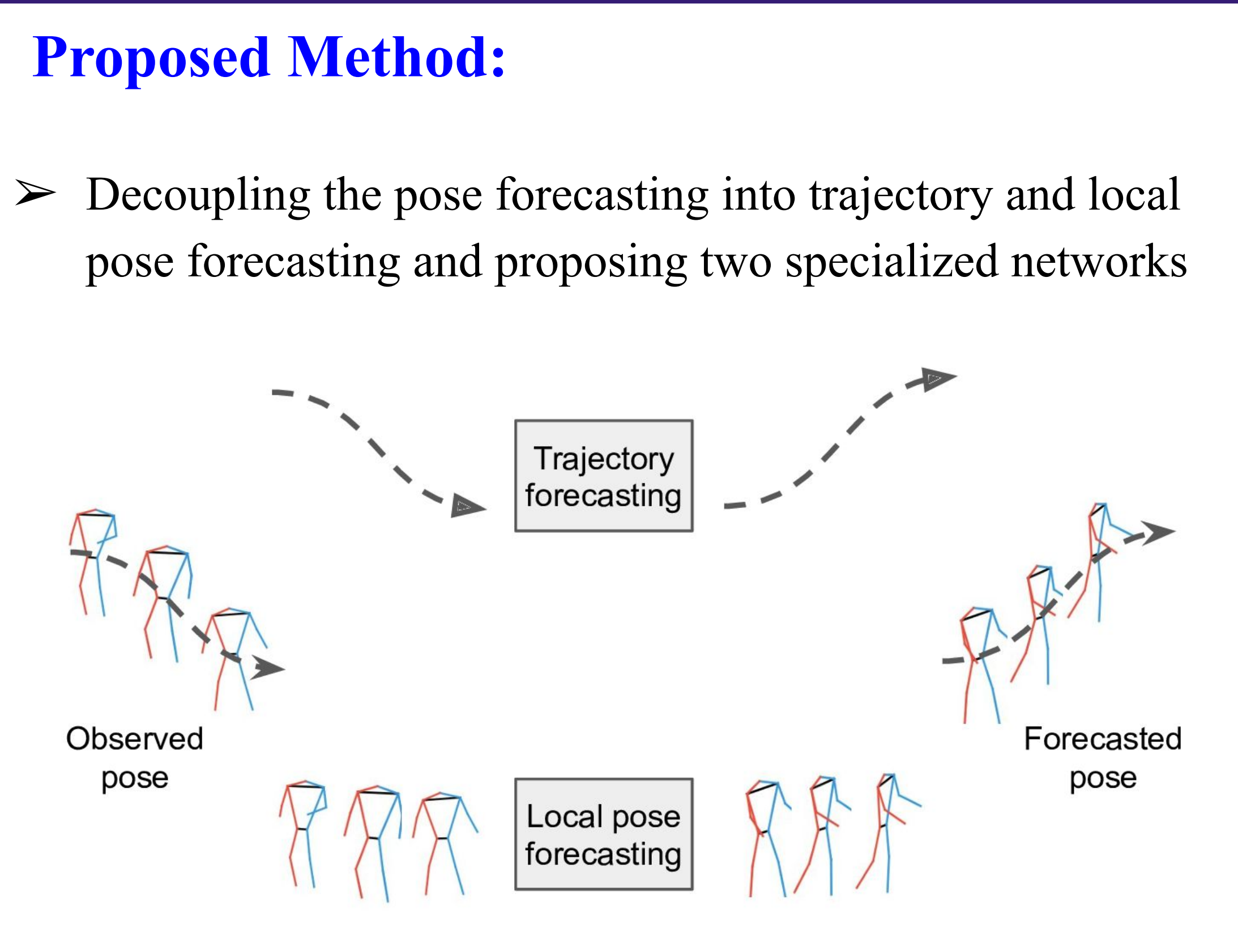


Introduction:

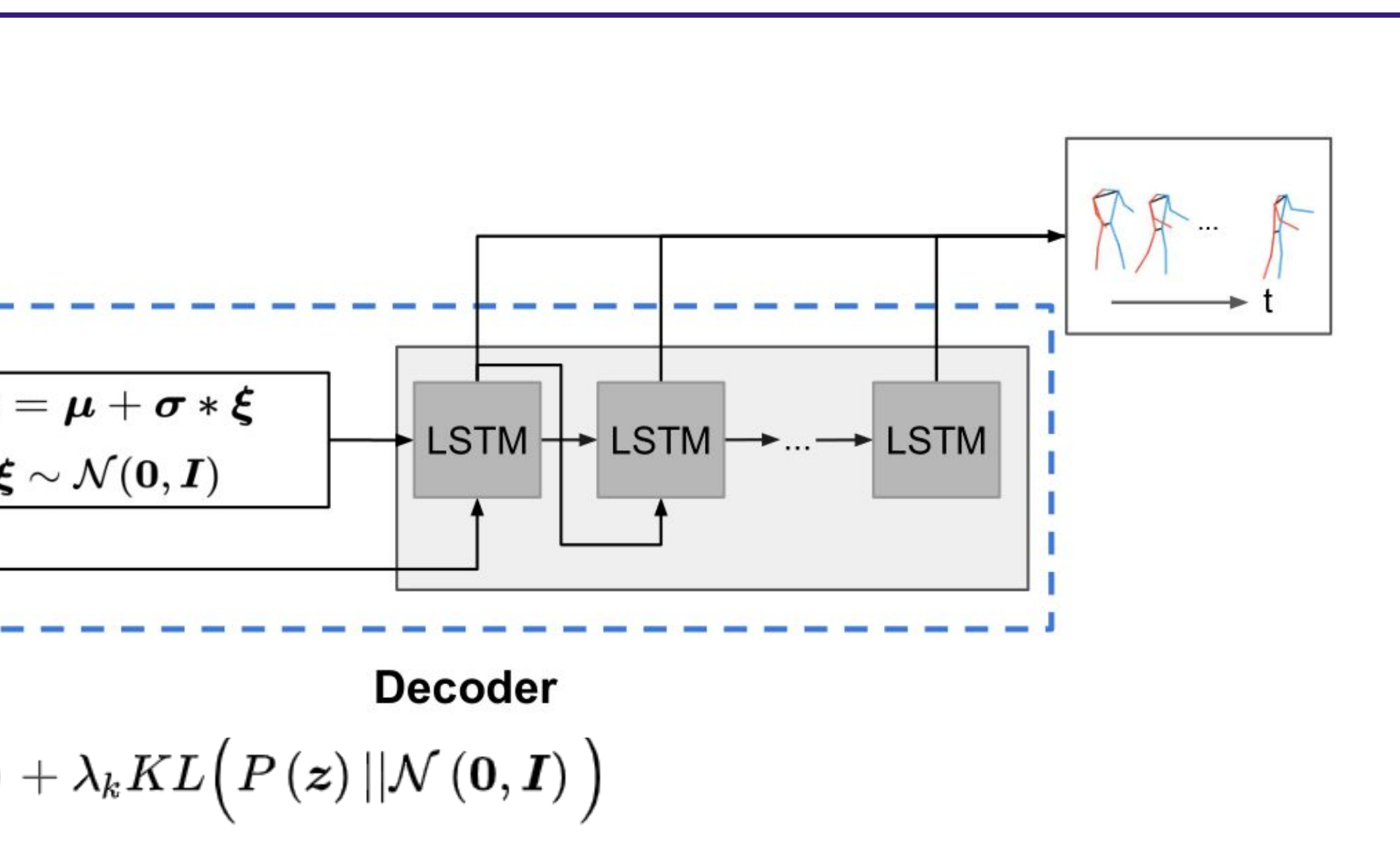
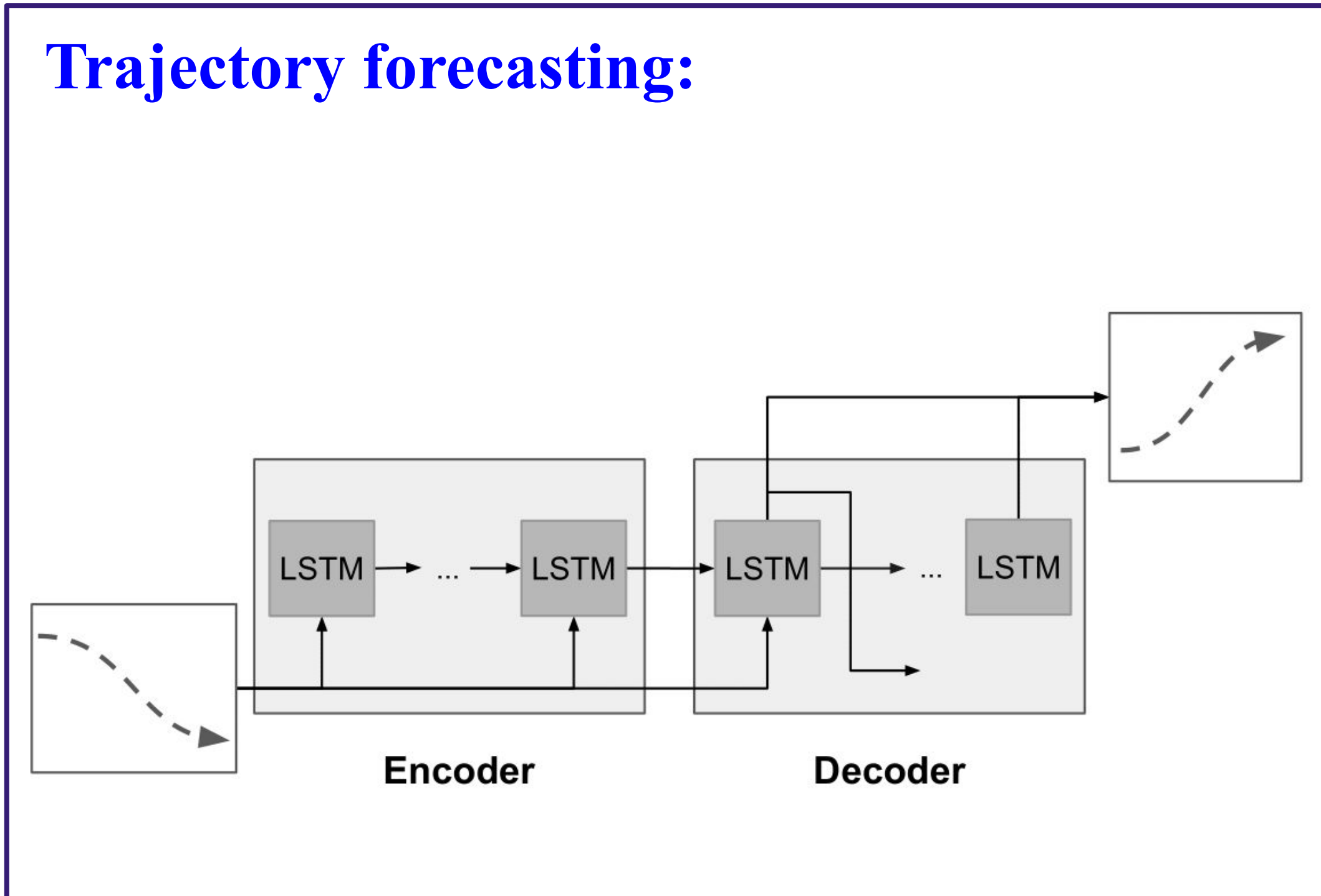
- **Human pose forecasting:** forecasting humans' keypoints locations (pose dynamics) given the keypoints locations of previous frames
- **Applications:** self-driving cars, healthcare, detection of behavioral patterns in surveillance systems, etc.
- **Challenges:** High uncertainty in human motion, hard when both global (trajectory) and local pose change



Results:

method	Prediction time				
	100 ms	240 ms	500 ms	640 ms	900 ms
PF-RNN [35]+S-LSTM [3]	73.82	127.23	179.07	202.78	277.55
MoAtt [34]+S-LSTM [3]	64.64	111.67	168.67	202.16	267.65
PF-RNN [35]+S-GAN [19]	83.35	138.48	182.84	204.84	291.96
MoAtt [34]+S-GAN [19]	66.36	112.18	166.48	209.53	277.85
PF-RNN [35]+ST-GAT [22]	66.95	117.77	165.99	190.52	252.23
MoAtt [34]+ST-GAT [22]	62.15	97.74	155.23	184.96	250.98
SC-MPF [1]	46.28	73.88	130.23	160.83	208.44
TRiPOD [2]	30.26	51.84	85.08	104.78	146.33
Nearest neighbour	27.34	51.68	97.75	121.40	168.27
Zero velocity	29.35	53.56	94.52	112.68	143.10
DeRPoF (ours)	19.53	36.89	68.29	85.45	118.21

3DPW



method	Prediction time				
	80 ms	160 ms	320 ms	400 ms	560 ms
PF-RNN [35]+S-LSTM [3]	87.02/89.44	103.22/111.11	129.21/136.43	138.77/145.72	160.95/157.04
MoAtt [34]+S-LSTM [3]	82.06/86.76	100.99/109.02	121.43/130.82	132.73/142.35	156.15/155.21
PF-RNN [35]+S-GAN [19]	84.40/87.23	99.24/106.23	130.21/131.12	130.21/139.94	150.03/150.44
MoAtt [34]+S-GAN [19]	82.45/85.82	98.76/104.13	119.38/128.97	127.98/139.07	149.53/151.45
PF-RNN [35]+ST-GAT [22]	82.06/86.76	94.25/102.61	117.70/127.87	126.71/137.87	148.65/150.80
MoAtt [34]+ST-GAT [22]	80.60/86.29	93.43/100.92	115.68/125.32	129.54/137.50	141.13/147.92
SC-MPF [1]	22.01/78.36	37.99/ 99.80	64.62/124.38	75.84/138.52	93.54/147.93
TRiPOD [2]	15.21/30.00	26.79/ 49.66	48.12/ 80.32	58.68/ 93.32	74.11/110.40
Nearest neighbour	11.75/24.62	21.35/ 42.05	41.15/ 70.95	50.99/ 82.76	66.80/ 99.91
Zero velocity	13.17/26.57	24.06/ 45.17	43.31/ 72.92	52.17/ 83.87	65.63/ 97.34
DeRPoF (ours)	10.20/22.05	18.56/37.29	34.89/62.01	42.76/73.10	54.62/88.12

PoseTrack

Ablation study:

method	Prediction time				
	100 ms	240 ms	500 ms	640 ms	900 ms
DeRPoF	19.53	36.89	68.29	85.45	118.21
w/o Early stop	19.53	36.89	70.70	89.02	126.19
w/o Decoupling	19.27	36.84	71.02	89.76	127.73
w/o VAE, Decoupling	20.50	37.95	72.68	91.94	131.99

Ablation study on 3DPW

Conclusion:

- We proposed a Decoupled Representation for Pose Forecasting (DeRPoF)
 - An LSTM-encoder-decoder for trajectory forecasting
 - A VAE-encoder-decoder for local pose forecasting
- We stop the prediction once the uncertainty in the human motion is high
- Our proposed method outperforms the baselines in SoMoF challenge

Code and more information:
github.com/vita-epfl/decoupled-pose-prediction