JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

AdamsFormer for Spatial Action Localization in the Future

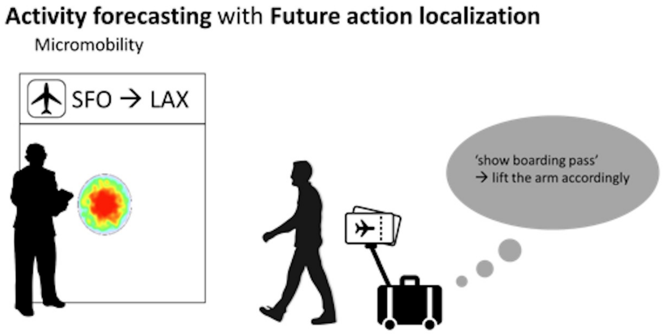
Hyung-gun Chi[2] Kwonjoon Lee[1] Nakul Agarwal[1] Yi Xu[3] Karthik Ramani[2] Chiho Choi[4]

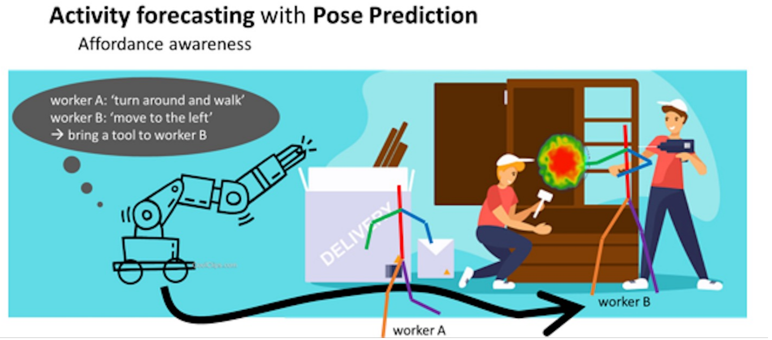
[1]Honda Research Institute USA [2]Purdue University [3]Northeastern University [4]Samsung Semiconductor US

Introduction

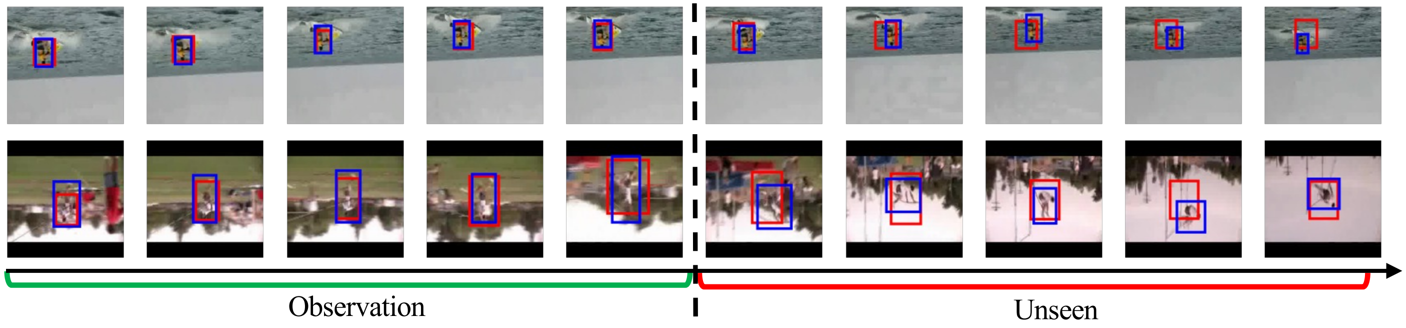
- Look for a location where current actions appear in the future.

- I.e., By understanding the exact location of future activities, the robot agent can provide more comfortable cooperation form the end application.





Spatial Action Localization in the Future



- We introduce a new task that aims to localize action in both observation and unseen frames.

Initial Value Problem and Nueral ODE

$$z'(t) = \frac{dz}{dt} = f(t, z), z(t_0) = z_0,$$

- IVP: Ordinary differential equation with an initial condition.
- Neural ODE (NeurIPS18) : it models $f(t, z)$ with a neural network.

Single-step VS Multi-Step

- Single-Step
  - i.e. Euler method

$$z_{n+1} = z_n + hf(t_n, z_n).$$

- Multi-Step
  - i.e. Adams method

N=2

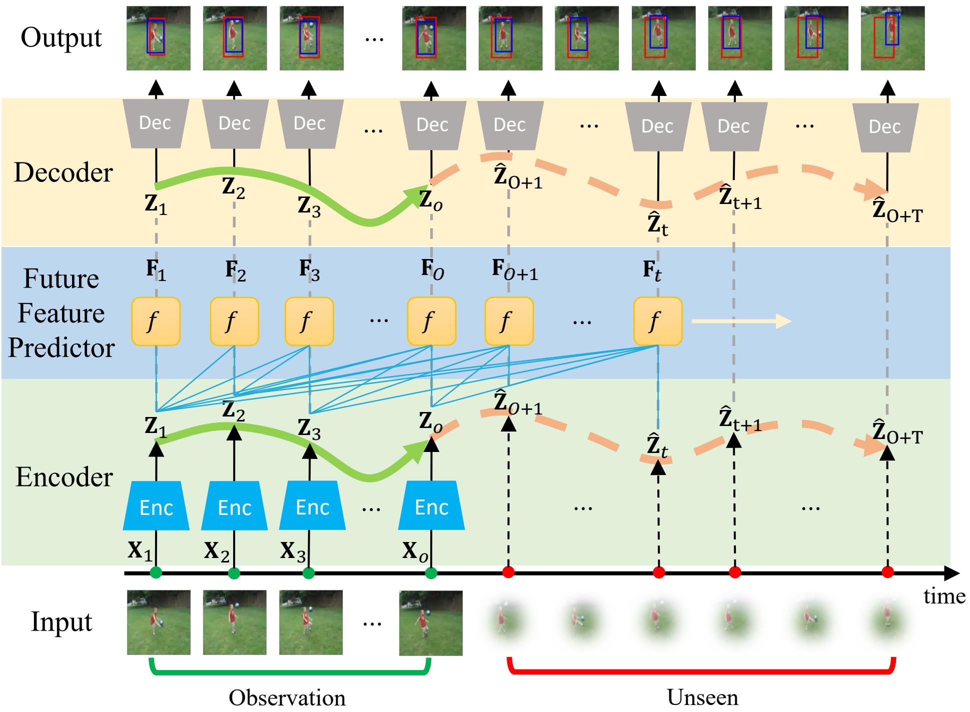
$$z_{n+1} = z_n + \frac{h}{2}[3f(t_n, z_n) - f(t_{n-1}, z_{n-1})].$$

N=3

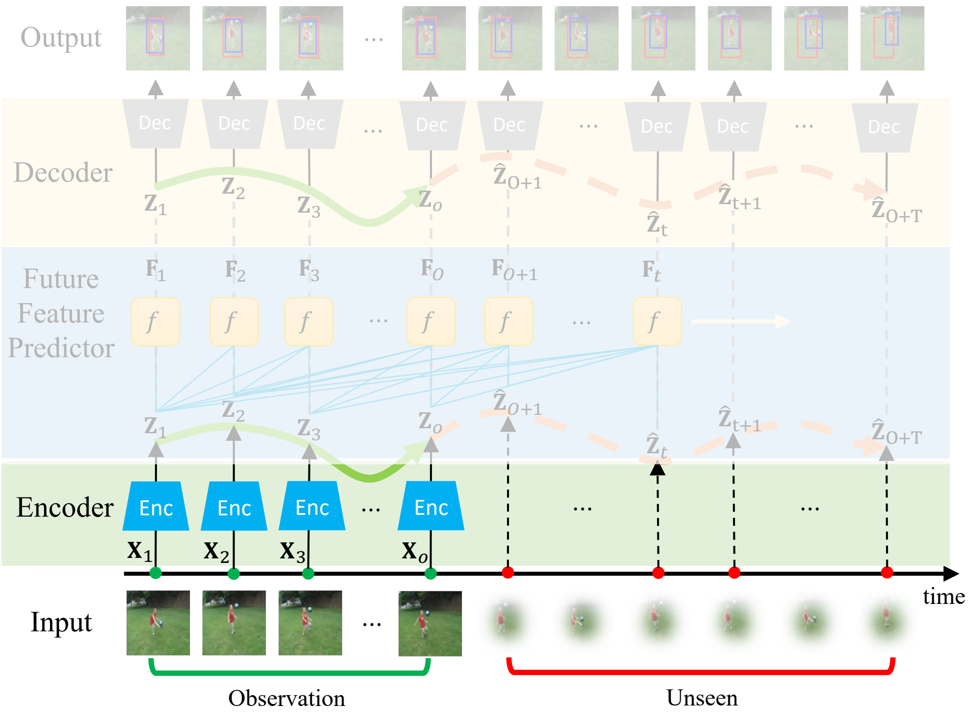
$$z_{n+1} = z_n + h[\frac{12}{23}f(t_n, z_n) - \frac{16}{23}f(t_{n-1}, z_{n-1}) + \frac{5}{12}f(t_{n-2}, z_{n-2})].$$

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

AdamsFormer - Overview



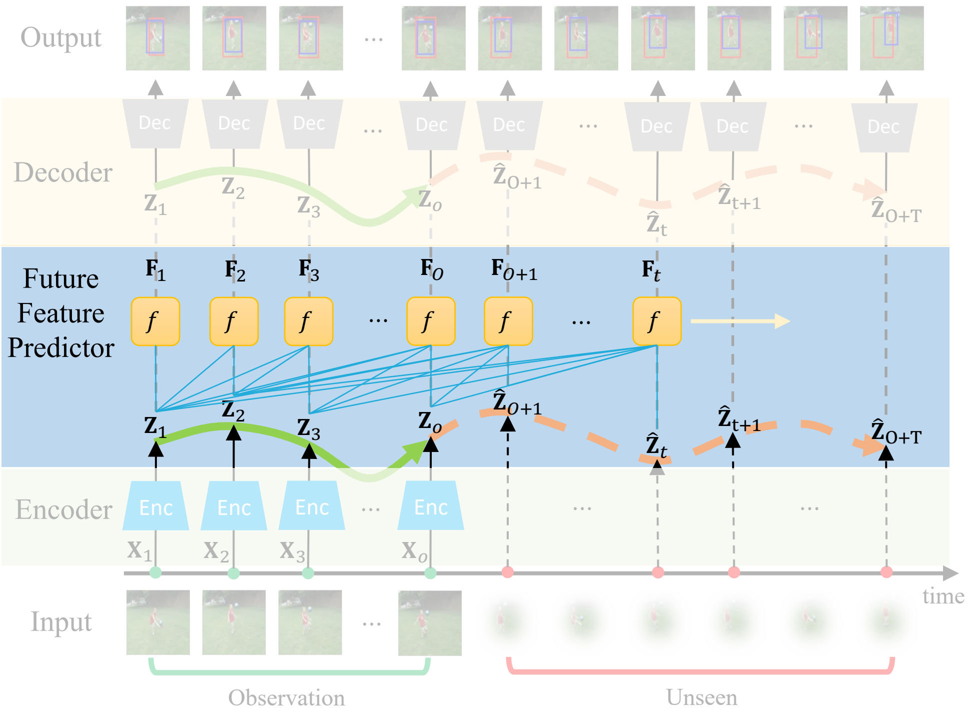
Encoder



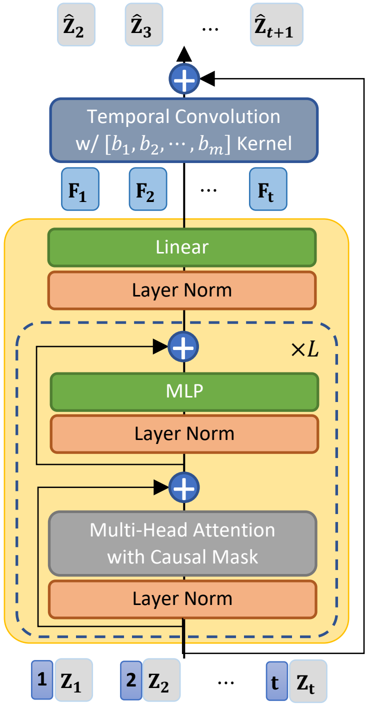
- Combination of 2D-CNN and 3D-CNN to fully utilize temporal information.

Future Feature Predictor

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA



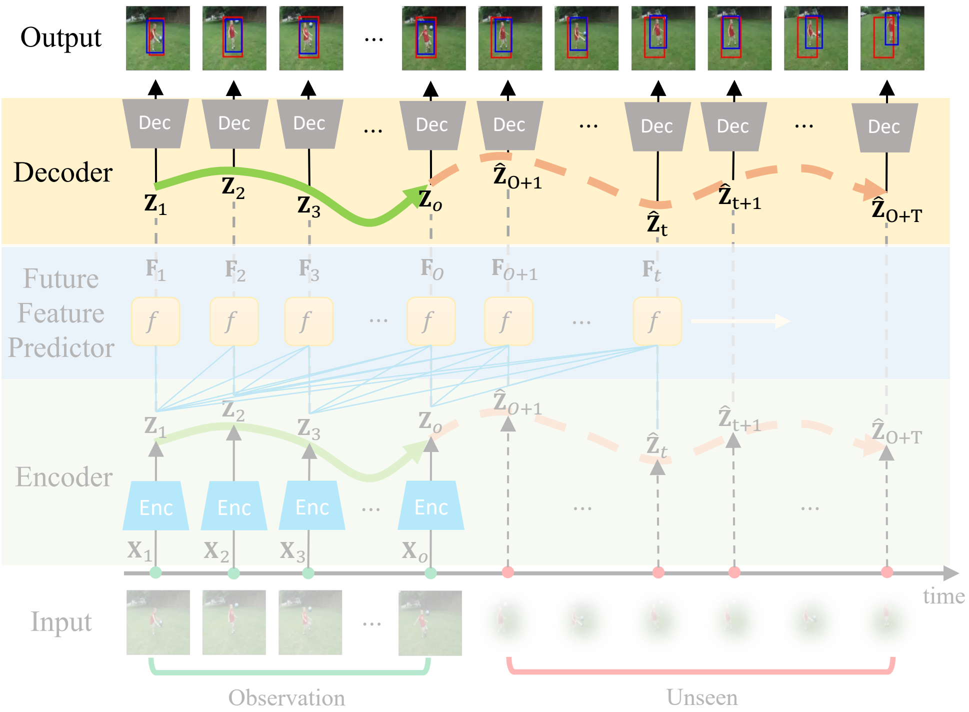
$$\mathbf{Z}_{t+1} = \mathbf{Z}_t + h \sum_{j=1}^{m} b_j \mathbf{F}_{t-j},$$



$$\mathbf{F} = f(\mathbf{t}_i, \mathbf{Z}_i)$$

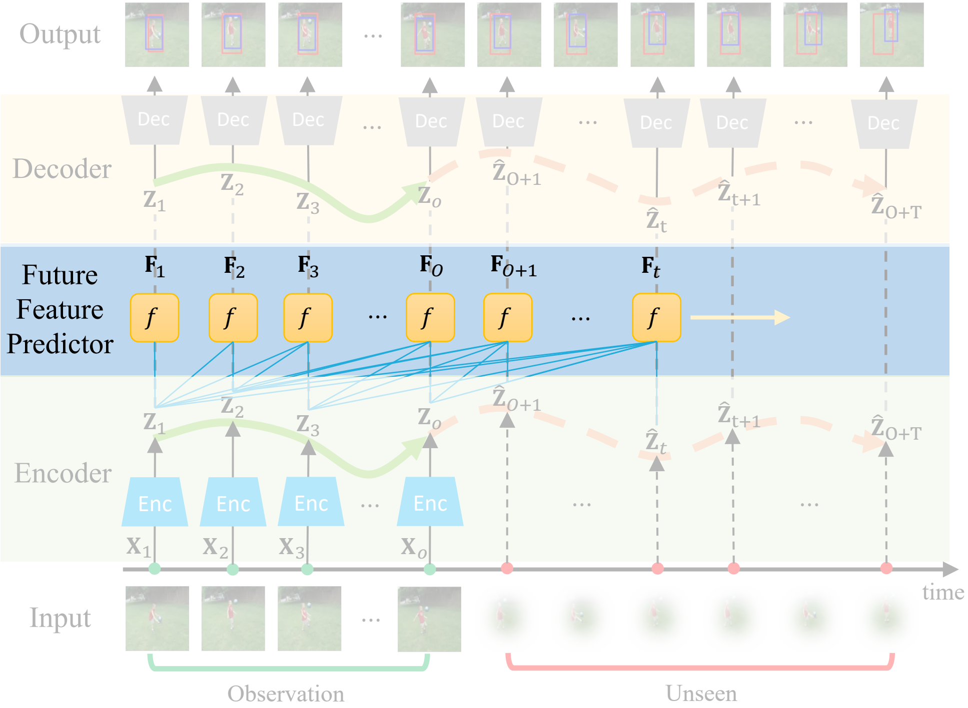
$$\mathbf{F} = f(\mathbf{t}_{1:i}, \mathbf{Z}_{1:i})$$

Decoder



- Decoder regresses the tensor to action location and category.

Experiments



- Setup
  - Replace future feature predictors with long-range temporal modeling methods.
- Baselines
  - RNN
  - ODE-RNN
  - PhyDNet
  - Anticipative Transformer

Comparison with baselines

Datasets	Methods	Observation Ratio									
		10%		20%		30%		40%		50%	
		OBS	UNSEEN	OBS	UNSEEN	OBS	UNSEEN	OBS	UNSEEN	OBS	UNSEEN
UCF101-24	RNN [44]	71.46	32.18	66.75	37.30	67.53	39.29	67.71	42.32	64.41	41.38
	ODE-RNN [6]	-	31.56	-	34.84	-	35.59	-	37.71	-	39.70
	PhyDNet [17]	65.86	29.90	67.16	37.22	67.43	39.69	67.69	41.44	66.42	42.47
	Transformer [14]	59.66	34.21	66.20	37.85	63.62	41.06	63.95	43.73	65.34	44.87
	AdamsFormer	**72.01**	**37.86**	**77.91**	**41.00**	**70.34**	**42.92**	**71.00**	**45.25**	**73.39**	**48.74**
JHMDB21	RNN [44]	40.64	10.85	40.61	24.76	42.62	32.06	39.95	29.82	38.19	31.19
	ODE-RNN [6]	-	19.99	-	21.63	-	24.57	-	28.86	-	31.69
	PhyDNet [17]	4.09	0.38	34.69	22.22	35.28	29.74	32.31	28.85	33.58	29.41
	Transformer [14]	38.46	35.17	38.61	40.24	41.65	44.09	46.87	50.66	45.34	50.45
	AdamsFormer	**51.26**	**49.39**	**51.59**	**49.55**	**51.21**	**51.72**	**51.84**	**53.28**	**50.19**	**52.81**

Advantage of the multi-step method

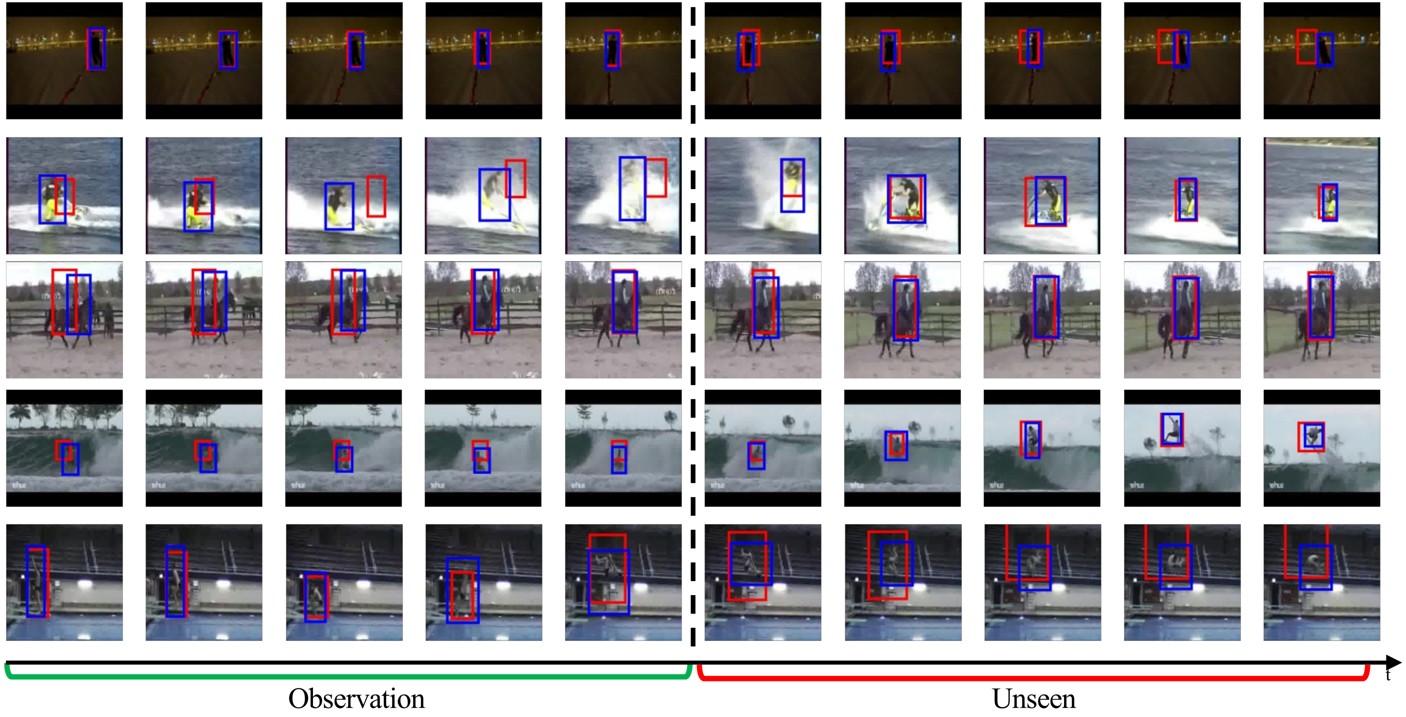
Methods	Observation Ratio									
	10%		20%		30%		40%		50%	
	OBS	UNSEEN	OBS	UNSEEN	OBS	UNSEEN	OBS	UNSEEN	OBS	UNSEEN
Single-step (m=1)	69.27	36.81	70.79	39.54	67.33	42.75	68.49	44.41	70.16	47.39
Multi-step (m=2)	72.01	37.86	74.11	40.04	68.61	42.87	70.91	45.32	72.59	47.19
Multi-step (m=4)	-	-	**77.91**	**41.00**	70.34	**42.92**	71.00	**45.25**	73.39	**48.74**
Multi-step (m=6)	-	-	-	-	**72.52**	42.34	**72.83**	42.44	**75.14**	48.00

- The multi-step method outperforms the single-step method.

Effect of order of the multi-step method

Methods	Observation Ratio									
	10%		20%		30%		40%		50%	
	OBS	UNSEEN	OBS	UNSEEN	OBS	UNSEEN	OBS	UNSEEN	OBS	UNSEEN
Single-step (m=1)	69.27	36.81	70.79	39.54	67.33	42.75	68.49	44.41	70.16	47.39
Multi-step (m=2)	72.01	37.86	74.11	40.04	68.61	42.87	70.91	45.32	72.59	47.19
Multi-step (m=4)	-	-	**77.91**	**41.00**	70.34	**42.92**	71.00	**45.25**	73.39	**48.74**
Multi-step (m=6)	-	-	-	-	**72.52**	42.34	**72.83**	42.44	**75.14**	48.00

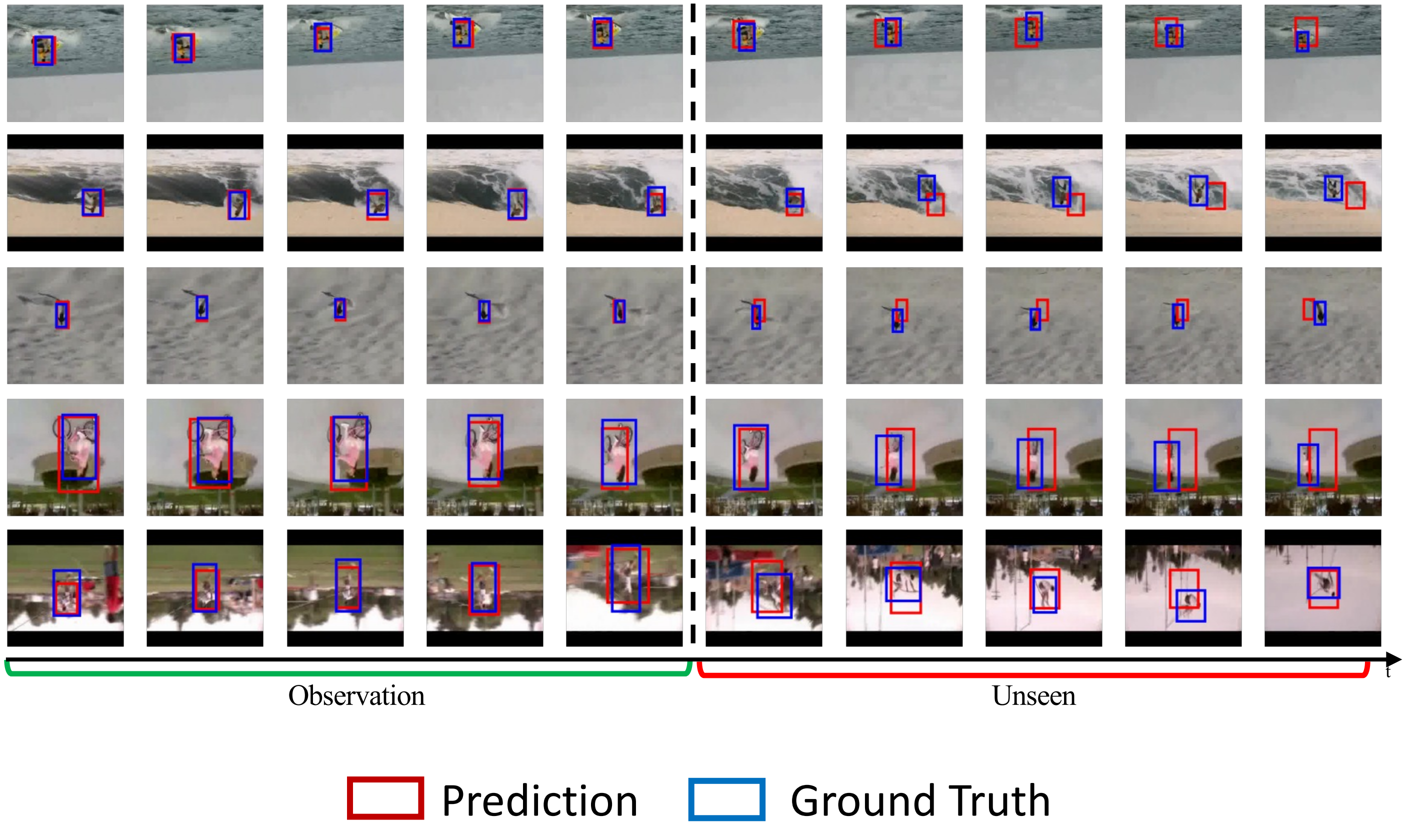
Qualitative Results



Prediction Ground Truth

Qualitative Results

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA



Thank you for listening!