ECCV'20 ONLINE

23-28 AUGUST 2020

16TH EUROPEAN CONFERENCE ON

COMPUTER VISION

WWW.ECCV2020.EU

4D Forecasting: Sequential Forecasting of 100,000 Points

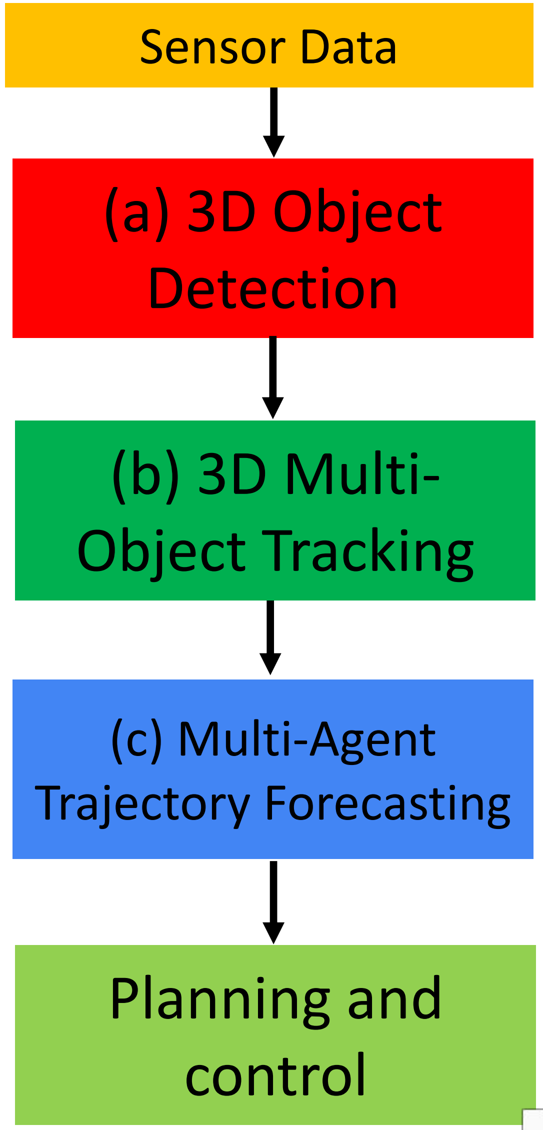
Xinshuo Weng[1], Jianren Wang[1], Sergey Levine[2], Kris Kitani[1], Nick Rhinehart[2]
[1] Robotics Institute, Carnegie Mellon University
[2] Berkeley Artificial Intelligence Research Lab, University of California, Berkeley

European Conference on Computer Vision (ECCV) Workshops

Standard Perception and Prediction Pipeline

- (a) Detection -> (b) MOT -> (C) Trajectory Forecasting

- Is this pipeline the best?
- Any limitation?
  - Requires instance-level object labels to train (a)
  - Requires sequence-level object labels to train (b)(c)
  - Expensive to obtain in 3D space

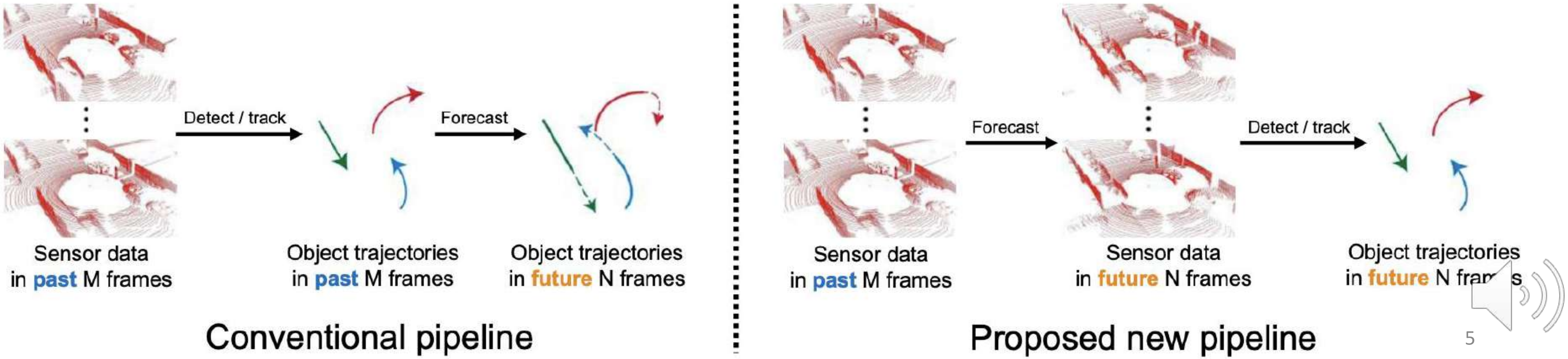


Our Contributions

1. A novel pipeline that inverts the order of forecasting and reduces labeling requirement
2. A new task, Sequential Pointcloud Forecasting (SPF), predicting a 3D representation of the future of the scene

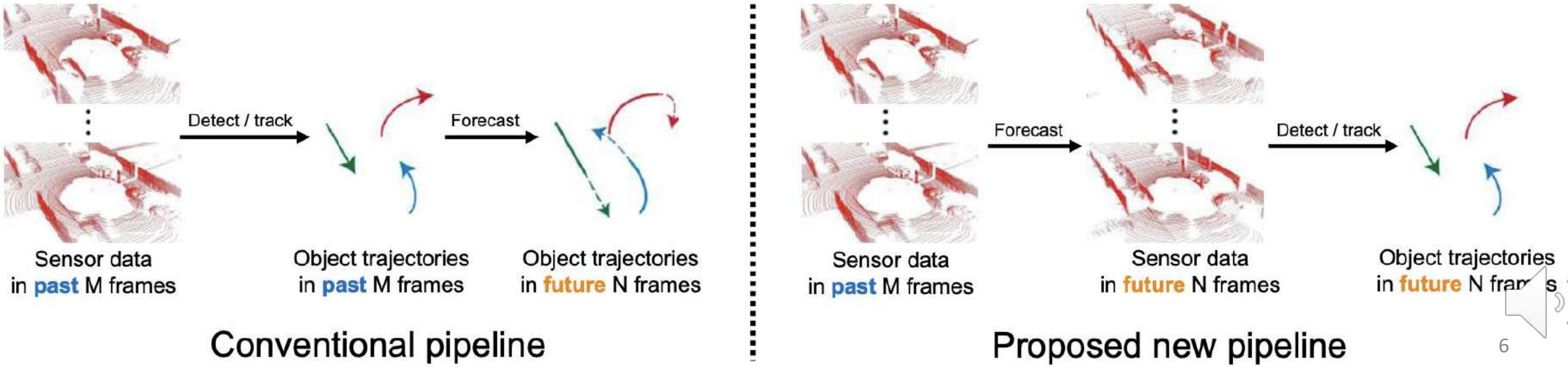
SPF2: Sequential Pointcloud Forecasting for Sequential Pose Forecasting

- Traditional pipeline:
  - Detection -> MOT -> Trajectory Forecasting
- Our new pipeline
  - Sequential Pointcloud Forecasting -> Detection -> MOT
- Differences
  - Invert the order of forecasting
  - Forecast at the sensor level, instead of at the object level



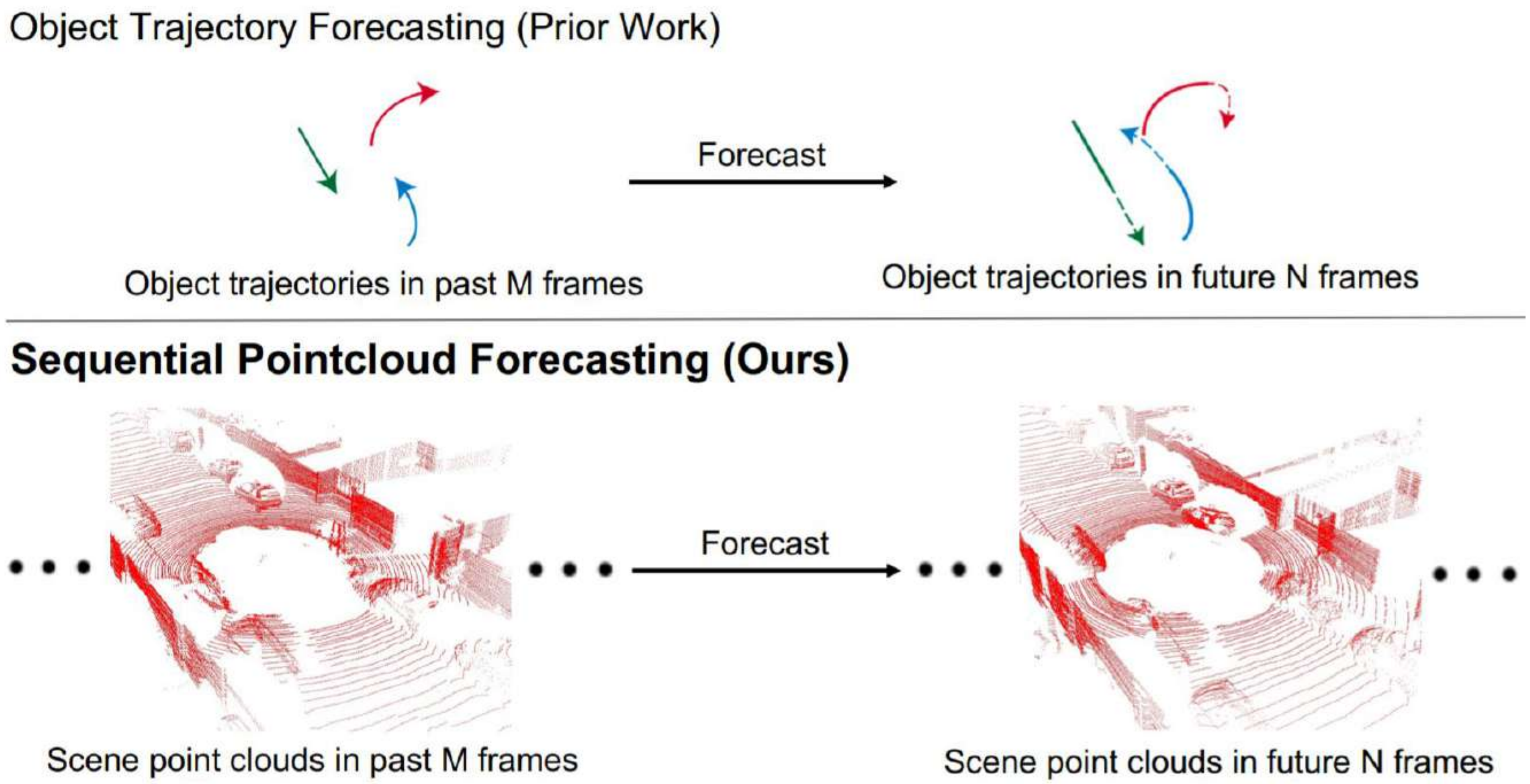
SPF2: Sequential Pointcloud Forecasting for Sequential Pose Forecasting

- Any advantage of our pipeline?
  - The forecasting module does not require human annotation
  - If using filter-based 3D MOT methods with S.O.T.A. performance, the labeling requirement can reduce to instance-level labels
    - Sequence-level labels are not required anymore



SPF: Sequential Pointcloud Forecasting

- Advantages:
  - Remove the need of labels for training
  - Prediction represents the entire scene, including information in the background



Our Contributions

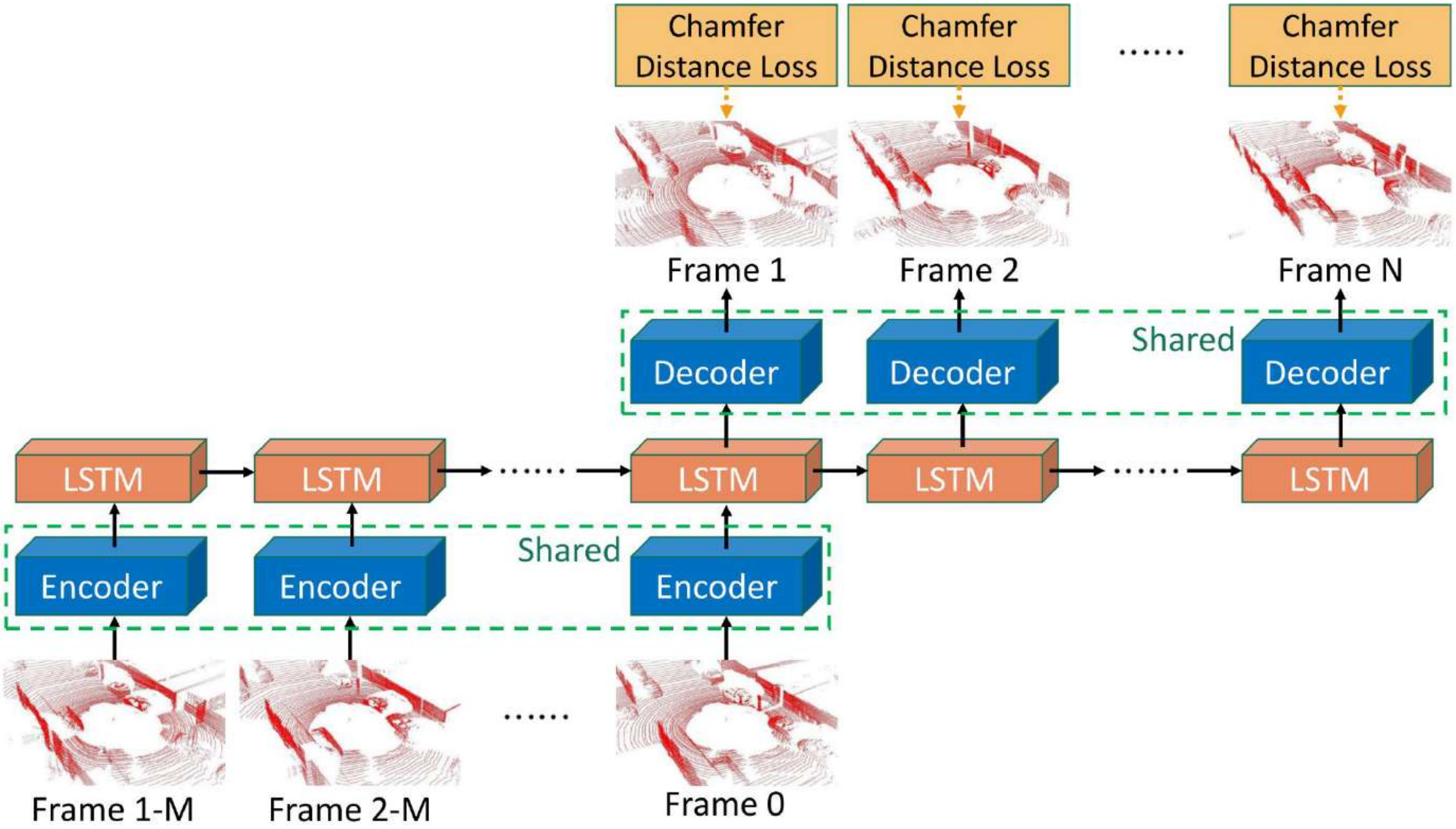
1. A novel pipeline that inverts the order of forecasting and reduces labeling requirement
2. A new task, Sequential Pointcloud Forecasting (SPF), predicting a 3D representation of the future of the scene
3. An effective approach for SPF, deemed SPFNet

SPFNet

- Four modules
  - (a) Shared point cloud encoder
  - (c) Shared point cloud decoder

(b) LSTM for temporal modeling

(d) Losses



Quantitative Results

Evaluation of the SPFNet on KITTI and nuScenes

- Is our SPFNet effective to the proposed SPF task?
  - Outperform baselines that we have devised using existing techniques

Table 1: Quantitative evaluation for the proposed SPF task on the KITTI and nuScenes datasets. Ours+Point and Ours+RM denote that we use point-based and range map-based encoder-decoder in our SPFNet.

Datasets	Metrics	**Ours+Point**	**Ours+RM**
KITTI-1.0s	CD↓	1.71	0.89
	EMD↓	211.47	128.81
KITTI-3.0s	CD↓	1.95	0.94
	EMD↓	267.42	175.54
nuScenes-1.0s	CD↓	1.03	0.35
	EMD↓	135.94	78.37
nuScenes-3.0s	CD↓	1.37	0.41
	EMD↓	128.26	91.83

Evaluation of the SPF2 Pipeline on KITTI and nuScenes

- Is our new perception and prediction pipeline competitive?

Table 3: Evaluation for the perception and trajectory forecasting pipeline on the KITTI and nuScenes datasets.

Datasets	Metrics	Samples	Conv-Social [16]	Social-GAN [5]	Social-BiGAT [6]	TraPHic [43]	**Ours**
KITTI-1.0s	AADE↓	1	0.792	0.524	1.099	0.470	0.317
		20	0.623	0.340	0.443	0.382	–
	AFDE↓	1	1.285	0.886	1.708	0.889	0.405
		20	1.152	0.511	0.546	0.613	–
KITTI-3.0s	AADE↓	1	1.692	1.362	2.720	1.432	0.408
		20	1.593	0.984	1.231	0.725	–
	AFDE↓	1	2.670	2.267	3.938	2.536	0.504
		20	2.385	1.512	1.405	1.118	–
nuScenes-1.0s	AADE↓	1	1.186	1.117	2.030	1.214	0.821
		20	0.907	0.762	0.826	0.881	–
	AFDE↓	1	1.490	1.310	2.337	1.563	0.825
		20	1.231	0.763	0.849	1.197	–
nuScenes-3.0s	AADE↓	1	1.794	2.224	4.954	2.417	1.044
		20	1.658	1.426	1.760	1.938	–
	AFDE↓	1	2.850	3.224	6.765	3.479	1.043
		20	2.538	1.652	1.845	2.766	–

Carnegie Mellon University
The Robotics Institute

BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

4D Forecasting: Sequential Forecasting of 100,000 Points

Xinshuo Weng[1], Jianren Wang[1], Sergey Levine[2], Kris Kitani[1], Nick Rhinehart[2]
[1] Robotics Institute, Carnegie Mellon University
[2] Berkeley Artificial Intelligence Research Lab, University of California, Berkeley

European Conference on Computer Vision (ECCV) Workshops