Carnegie Mellon University
The Robotics Institute

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

# Inverting the Pose Forecasting Pipeline with SPF$^2$: Sequential Pointcloud Forecasting for Sequential Pose Forecasting

Xinshuo Weng[1], Jianren Wang[1], Sergey Levine[2], Kris Kitani[1], Nick Rhinehart[2]

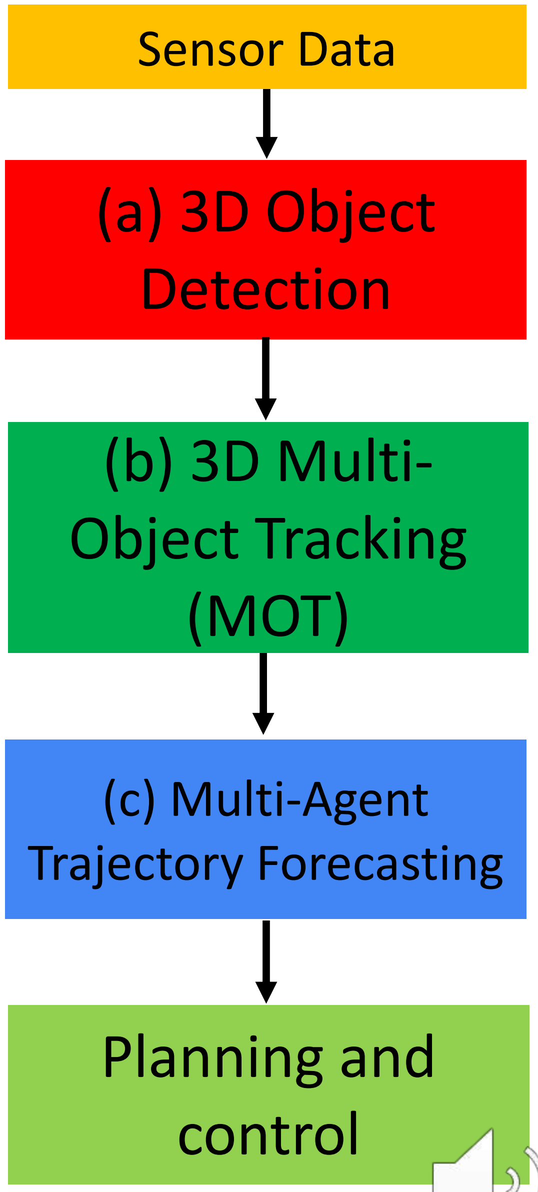
[1] Robotics Institute, Carnegie Mellon University

[2] Berkeley Artificial Intelligence Research Lab, University of California, Berkeley

*4th Conference on Robot Learning (**CoRL**)*, 2020

# Standard Pose (Trajectory) Forecasting Pipeline

- (a) Detection -> (b) MOT -> (C) Trajectory Forecasting
- Is this detect-then-forecast pipeline the only option?
- Any limitation?
  - Expensive to scale as pose forecasting algorithms typically require labeled sequences of object poses
  - Expensive to annotate in 3D space



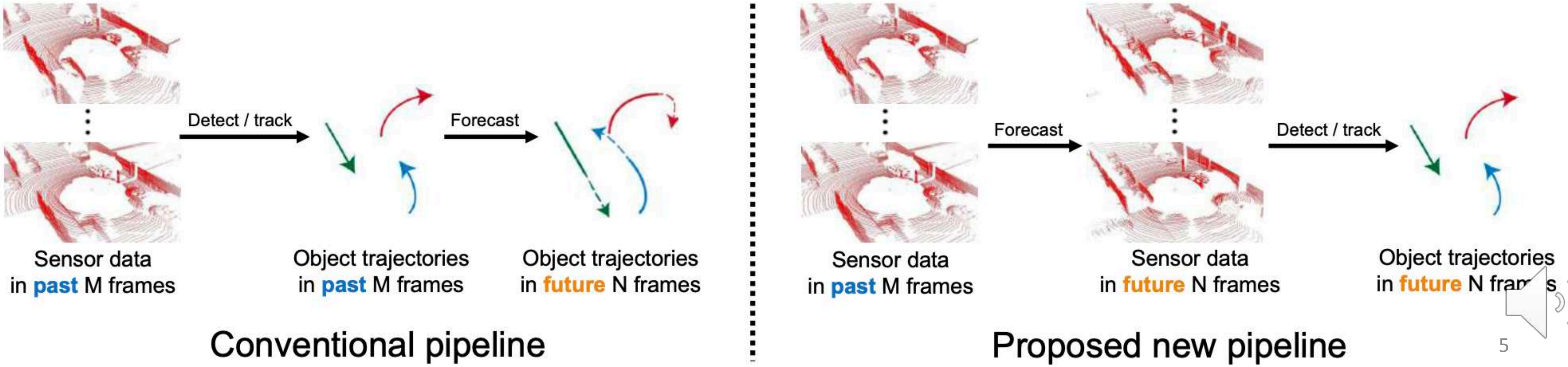
Can we scale pose forecasting performance without requiring additional labels?

We hypothesize yes!

A forecast-then-detect pipeline
that inverts the order of forecasting and makes
it less expensive to scale pose forecasting

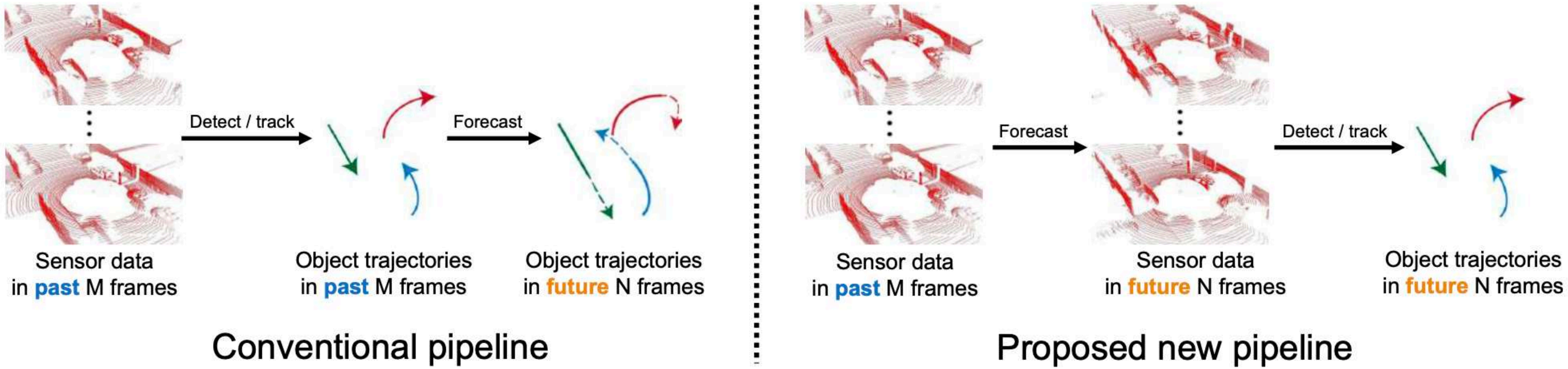
# SPF$^2$: Sequential Pointcloud Forecasting for Sequential Pose Forecasting

- 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



# SPF$^2$: Sequential Pointcloud Forecasting for Sequential Pose Forecasting

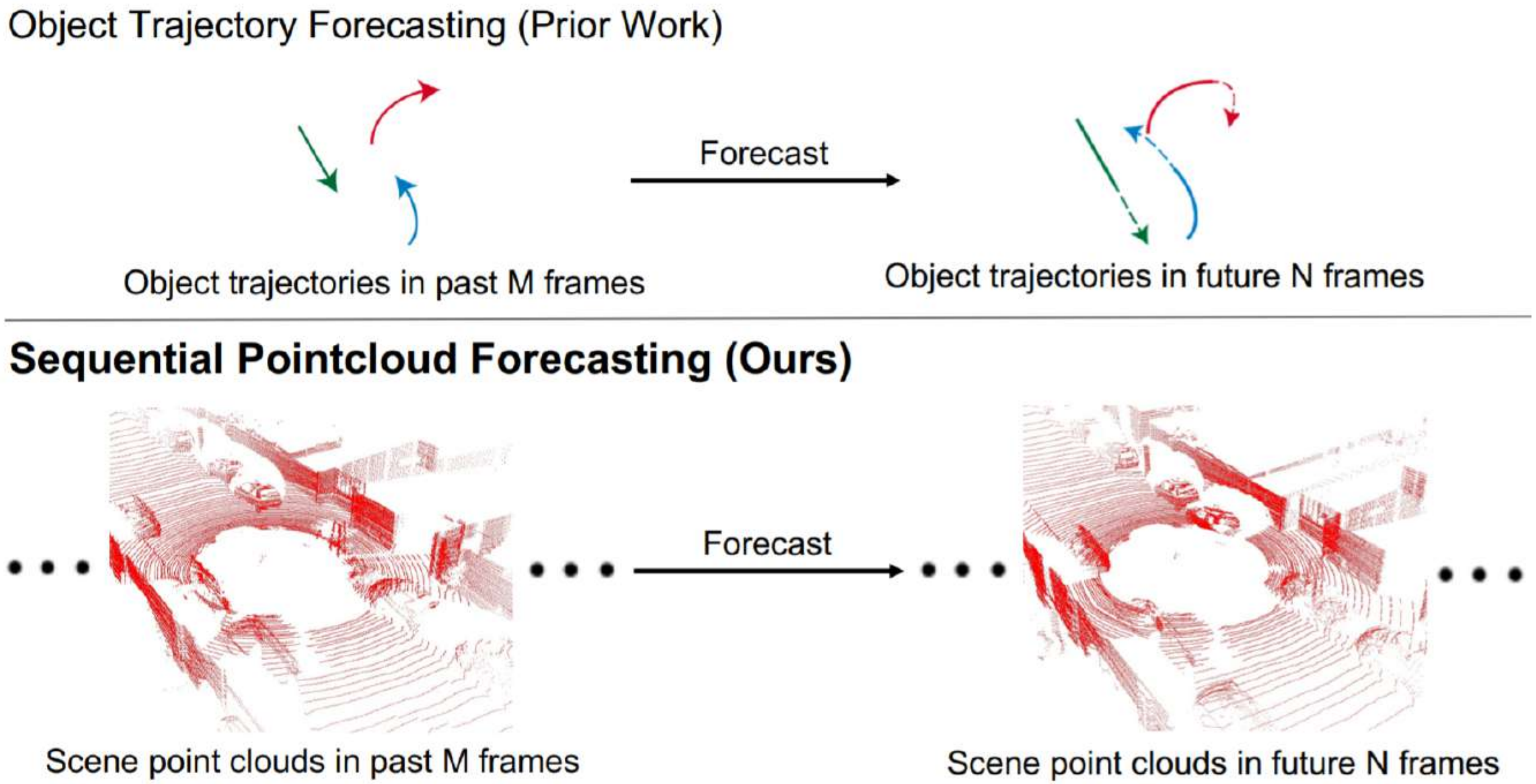
- Any advantage of our pipeline?
  - The forecasting module does not require human annotation
  - Forecasting performance can scale with addition of unlabeled point cloud data
- Inherited from the Sequential Pointcloud Forecasting (SPF)



# The challenging first step -- Sequential Pointcloud Forecasting (SPF)

# SPF: Sequential Pointcloud Forecasting

- Advantages:
  - Does not require expensive-to-collect labels for training / evaluation
  - Prediction represents the entire scene, including information about both scene background and foreground objects



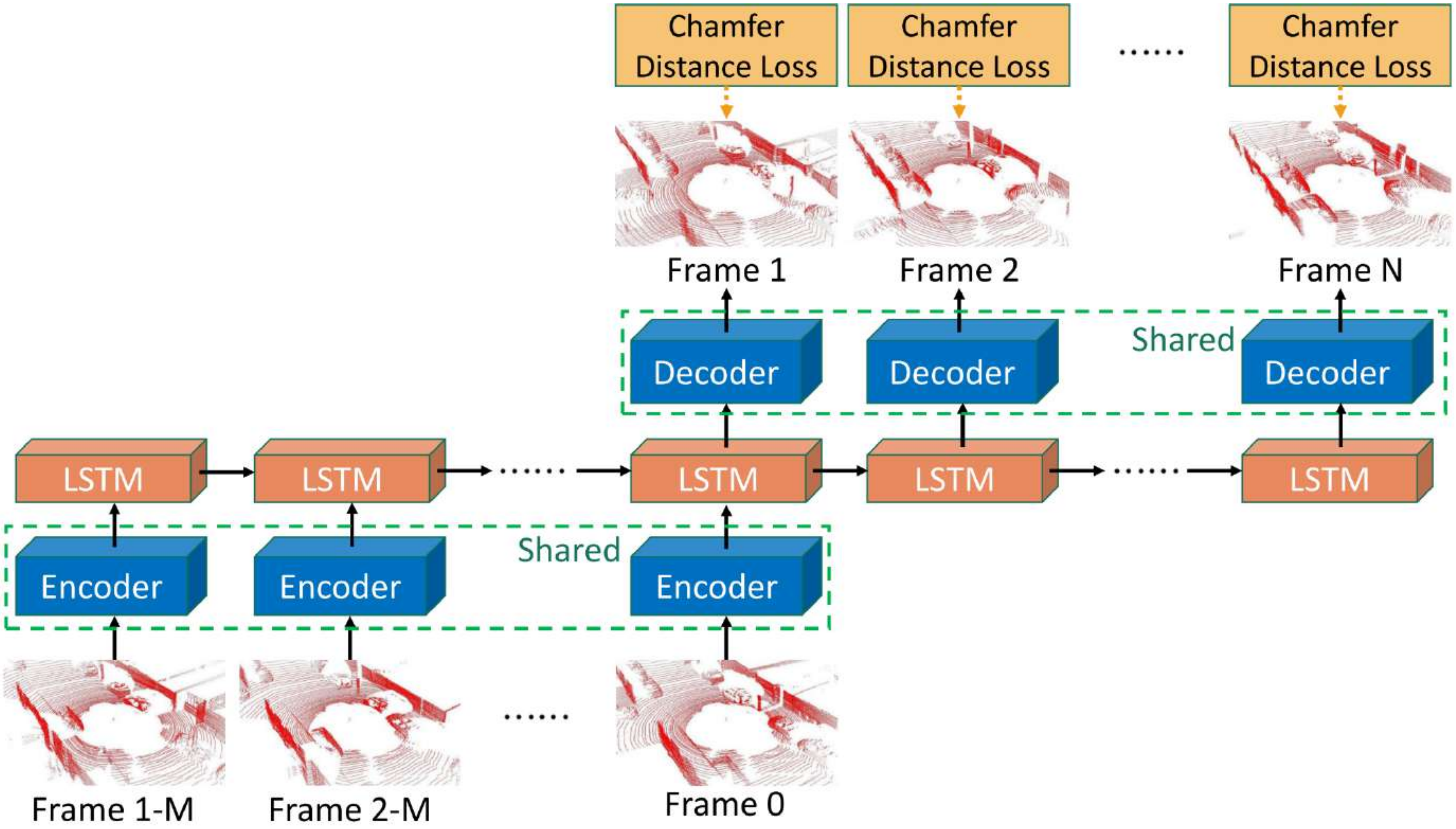
# An effective approach for SPF deemed SPFNet

# 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.

| Datasets | Metrics | Identity | GT-Ego | Est-Ego | Align-ICP | Align-[46] | SceneFlow | **Ours+Point** | **Ours+RM** |
|---|---|---|---|---|---|---|---|---|---|
| KITTI-1.0s | CD↓ | 12.82 | 5.47 | 9.18 | 6.13 | 6.02 | 3.15 | 2.37 | 0.92 |
| | EMD↓ | 526.87 | 391.03 | 495.21 | 418.25 | 439.17 | 291.63 | 211.47 | 128.81 |
| KITTI-3.0s | CD↓ | 13.31 | 7.91 | 11.31 | 9.14 | 9.57 | 5.08 | 3.91 | 1.57 |
| | EMD↓ | 602.89 | 452.81 | 502.83 | 470.25 | 493.26 | 351.46 | 267.42 | 175.54 |
| nuScenes-1.0s | CD↓ | 8.42 | 2.16 | 4.91 | 4.04 | 3.50 | 1.93 | 1.25 | 0.40 |
| | EMD↓ | 461.63 | 168.37 | 299.13 | 281.53 | 270.81 | 117.41 | 135.94 | 78.37 |
| nuScenes-3.0s | CD↓ | 10.16 | 2.85 | 6.52 | 7.13 | 5.27 | 3.81 | 2.97 | 0.69 |
| | EMD↓ | 494.81 | 190.14 | 370.91 | 419.37 | 332.97 | 294.53 | 128.26 | 91.83 |

# Evaluation of the $SPF^2$ Pipeline on KITTI and nuScenes

- Is our new pose forecasting 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

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

# Inverting the Pose Forecasting Pipeline with SPF$^2$: Sequential Pointcloud Forecasting for Sequential Pose Forecasting

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

*4th Conference on Robot Learning (**CoRL**)*, 2020