Perception and prediction are important components in the autonomous driving stack
Standard Perception and Prediction Pipeline

Sensor Data

3D Object Detection

3D Multi-Object Tracking

Trajectory Forecasting

Perception

Prediction
Standard Perception and Prediction Pipeline

Sensor Data

3D Object Detection

3D Multi-Object Tracking

Trajectory Forecasting

LiDAR

RGB

Standard Perception and Prediction Pipeline
Standard Perception and Prediction Pipeline

1. Sensor Data
2. 3D Object Detection
3. 3D Multi-Object Tracking
4. Trajectory Forecasting

Detection results
Standard Perception and Prediction Pipeline

Sensor Data

3D Object Detection

3D Multi-Object Tracking

Trajectory Forecasting

Tracking results
Standard Perception and Prediction Pipeline

1. Sensor Data
2. 3D Object Detection
3. 3D Multi-Object Tracking
4. Trajectory Forecasting

Forecasting results
Standard Pipeline Remains the Same

Sensor Data

3D Object Detection

3D Multi-Object Tracking

Trajectory Forecasting

Significant progress on each individual component

But the pipeline is the same.

Any potential improvement at the pipeline level?
Standard Perception and Prediction Pipeline

Sensor Data

3D Object Detection

3D Multi-Object Tracking

Trajectory Forecasting

Is this really the best place to perform prediction?
Standard Perception and Prediction Pipeline

Sensor Data

3D Object Detection

3D Multi-Object Tracking

Trajectory Forecasting

Can we do prediction here?
Limitation of the Standard Pipeline

• Pipeline in a sequential order
  • Downstream module takes the outputs of its upstream module as inputs

• Limitation?
  • Errors from upstream are propagated to downstream

• Can we go beyond the sequential pipeline?
Our contribution

A parallelized tracking and prediction (PTP) framework

→ Alleviate that errors in 3D MOT affect prediction
→ Enable joint optimization
PTP: Parallelized Tracking and Prediction

Sensor Data

3D Object Detection

3D Multi-Object Tracking

Feature Extraction

Matching

Trajectory Prediction

Feature Extraction

Trajectory Decoder

Sensor Data

3D Object Detection

Shared Feature Learning

3D Multi-Object Tracking

Matching

Trajectory Prediction

Trajectory Decoder

Similar components, which aims to encode object features from past information

Module-specific components

Sequential Pipeline

Parallelized Tracking and Prediction (PTP)
Our Method Can Jointly Track and Predict
Parallelized Tracking and Prediction

- Advantages
  - Reduce error propagation from tracking to prediction
  - Share feature learning -> efficient and improves performance

- Overview

Shared Feature Learning

3D MOT

Forecasting

- Feature extraction
- GNN for feature interaction
- Edge features
- Node features
- Feature extraction
- Diversity sampling
- Trajectory prediction head

Sensor Data
3D Object Detection
Shared Feature Learning
3D Multi-Object Tracking
Trajectory Prediction

Objects trajectories in past H frames
Detected objects in current frame
Last frame
Current frame
Predicted trajectories in future T frames

3D MOT

Feature extraction

Diversity sampling

Trajectory prediction head
Parallelized Tracking and Prediction

- Shared feature learning
  - Use LSTM/MLP to learn motion features from objects’ box trajectories
  - Encode contextual / relative features from nearby objects by modeling interaction with GNNs
Parallelized Tracking and Prediction

- 3D multi-object tracking
  - MLP takes edge features as inputs to regress the similarity scores between every pair of objects
  - During training, estimated affinity matrix is supervised with GT
  - During testing, estimated affinity matrix is fed to Hungarian algorithm

\[ \mathcal{L}_{\text{aff}} = \mathcal{L}_{\text{bce}} + \mathcal{L}_{\text{ce}} \]

**Diagram:**
- **Sensor Data**
- **3D Object Detection**
- **Shared Feature Learning**
- **3D Multi-Object Tracking**
- **Trajectory Prediction**
Parallelized Tracking and Prediction

- Trajectory prediction
  - Conditional VAE is used to predict future trajectories
  - Diversity sampling technique that maps node feature to a set of latent codes covering various modes of future trajectories
Diversity Sampling | Limitation of Random Trajectory Prediction

Learn a generative model $p_{\theta}(x|\psi)$

Low sample efficiency!
Our Approach | Diversity Sampling Function

Diversity Sampling Function (DSF)

\[ S_\gamma(\psi) \]

Latent codes \( \{z_1, \ldots, z_N\} \)

Generator \( G_\theta(x|z, \psi) \)

Data

Context feature \( \psi \)

Future trajectories \( x \)

Diversity loss on samples \( \{x_1, \ldots, x_N\} \)

Latent space

Trajectory Space

Optimize
Parallel Tracking Benefits Prediction

• Is the parallel pipeline and joint optimization effective?
  • How does adding 3D MOT affect performance of prediction?
  • Add 3D MOT branch improves performance on prediction

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Metrics</th>
<th>w/o MOT+DSF</th>
</tr>
</thead>
<tbody>
<tr>
<td>KITTI-1.0s</td>
<td>ADE↓</td>
<td>0.663</td>
</tr>
<tr>
<td></td>
<td>FDE↓</td>
<td>1.121</td>
</tr>
<tr>
<td></td>
<td>ASD↑</td>
<td>1.796</td>
</tr>
<tr>
<td></td>
<td>FSD↑</td>
<td>3.168</td>
</tr>
<tr>
<td></td>
<td>ADE↓</td>
<td>1.729</td>
</tr>
<tr>
<td>KITTI-3.0s</td>
<td>FDE↓</td>
<td>3.086</td>
</tr>
<tr>
<td></td>
<td>ASD↑</td>
<td>3.196</td>
</tr>
<tr>
<td></td>
<td>FSD↑</td>
<td>5.776</td>
</tr>
</tbody>
</table>

Performance improved after adding MOT!
Parallel Prediction Benefits Tracking

- Is the parallel pipeline and joint optimization effective?
  - Add MOT is useful to prediction

- How does adding prediction affect performance of 3D MOT?
  - Add prediction branch improves performance on tracking

<table>
<thead>
<tr>
<th>Metrics</th>
<th>w/o forecasting</th>
</tr>
</thead>
<tbody>
<tr>
<td>sAMOTA(%)↑</td>
<td>91.31</td>
</tr>
<tr>
<td>AMOTA(%)↑</td>
<td>43.68</td>
</tr>
<tr>
<td>AMOTP(%)↑</td>
<td>76.94</td>
</tr>
<tr>
<td>MOTA(%)↑</td>
<td>83.51</td>
</tr>
<tr>
<td>MOTP(%)↑</td>
<td>78.11</td>
</tr>
<tr>
<td>IDS↓</td>
<td>5</td>
</tr>
</tbody>
</table>

Improvement on 5 out of 6 entries!

3D MOT evaluation without forecasting module
Parallelized Tracking and Prediction

• For more details in this work
  • Scan the QR code for the paper
Parallelized Tracking and Prediction

Sensor Data

3D Object Detection

3D Multi-Object Tracking

Trajectory Prediction

Sequential Pipeline

Sensor Data

3D Object Detection

Shared Feature Learning

3D Multi-Object Tracking

Trajectory Prediction

PTP: Parallelized Tracking and Prediction