Hochschule München
University of Applied Sciences

Fakultät für Informatik und Mathematik
Department of Computer Sciences and Mathematics

Multimodal Trajectory Prediction in Multi-Agent Scenarios

Seminar: Video Analysis & Object Tracking

16 April 2025

Lukas Röß, Jan Duchscherer

Structure

- Objective Recap and Revision
- The UniTraj Framework
- The Dataset
- Metrics and Optimization
- MTR Training
- MTR Prediction
- Current Challenges & Issues
- Roadmap

Objective Recap and Revision

- ~~Joint Multi-Agent Trajectory Prediction~~ $\Longrightarrow$ Single-Agent Trajectory Prediction
- ~~Query Centric~~ $\Longrightarrow$ Agent Centric
- ~~Smol-CASPFormer~~ $\Longrightarrow$ Smol-LM-Former-alike-model
  - Transformer Encoder on Vector Embeddings
  - DAB-alike-Decoder (deformable attention w/ grounding)
  - NO lame RNNs
  - Maybe non-recurrent decoding w/ causality masking
- ~~Use fully functional framework~~ $\Longrightarrow$ Refactor everything

The UniTraj Framework

Dataset Fusion

- Standardized Multi-Dataset Training and Evaluation via **ScenarioNet**
  - ArgoverseV2, NuScenes, Waymo: different map and agent features, data formats, map resolutions & semantic annotations

[UniTraj]

⟹ Unified data features

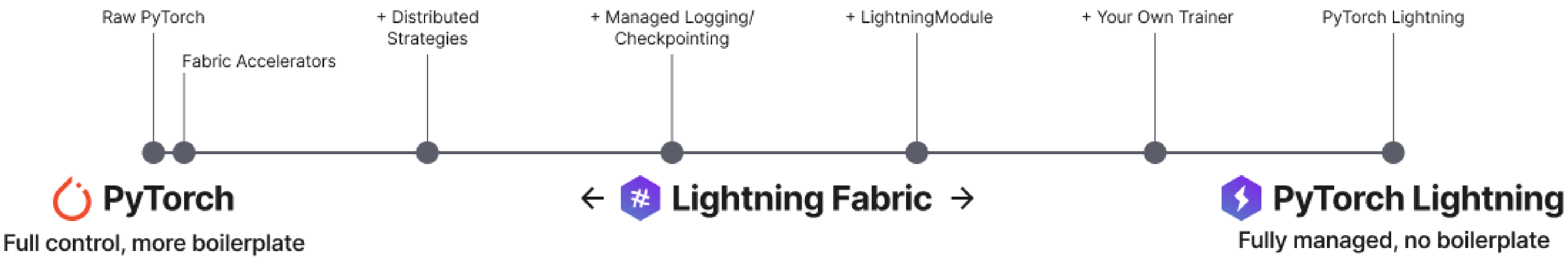
⟹ Improved inter-dataset comparability of generalization capabilities

⟹ Combination into largest Motion-Forecasting DS (**2M+ samples**)

The UniTraj Framework

Unified Training and Evaluation Suite

- Training, Evaluation and Logging via **PyTorch Lightning** and **WandB**



[PyTorch Lightning]

- Config and HParam handling via **Pydantic**
- Unified evaluation metrics and loss functions for different models
- Easily swap models / datasets / losses via *Config-as-Factory* pattern

The Data Processing Pipeline

AV2 $\Longrightarrow$ ScenarioNet $\Longrightarrow$ UniTraj

- **Agent Selection:**
  - Type $\in$ {VEH, PED, CYCL}
  - Movement: $\Delta d_i = \left\| p_i(T_p - 1) - p_i(0) \right\|_2 \geq d_{\min}$
  - Visibility: $\rho_i = \frac{1}{T_p} \sum_{t=0}^{T_p-1} \mathbb{1}[\text{valid}_{i,t}] \geq \rho_{\min}$
  - Kalman difficulty in specified range.

(1 Sample $\Longrightarrow$ N_c Samples)

Kalman difficulty	Easy	Medium	Hard
	$\in [0, 30[$	$\in [30, 50[$	$\in [50, 100[$

- **Coordinate Normalization:**

$$p_t^{(i),a} = R_z(-\theta_c)\big(p_t^{(i),w} - p_c\big), \quad R_z(-\theta_c) = \begin{pmatrix} \cos\theta_c & \sin\theta_c \\ -\sin\theta_c & \cos\theta_c \end{pmatrix}$$

(Scene Centric $\Longrightarrow$ Agent Centric)

- **Feature & Mask Assembly:**

$$\mathbf{X}_d \in \mathbb{R}^{N_{\max} \times T_p \times F_{\text{ap}}}, \quad \mathbf{M}_d \in \{0,1\}^{N_{\max} \times T_p}$$

$$\mathbf{X}_s \in \mathbb{R}^{K \times L \times F_{\text{map}}}, \quad \mathbf{M}_s \in \{0,1\}^{K \times L}$$

(padding and masking)

The Dataset

```
def __getitem__(self, idx: int) -> DatasetItem
```

- Agent-Centric Samples
  - All static & dynamic features are transformed into the center agent’s frame
  - Original scenario has 5 eligible agents $\Longrightarrow$ 5 distinct DatasetItems
    - *Single Agent Trajectory Prediction* (can be adapted easily to Join Multi Agent Prediction)
- Efficient Batched Loading via HDF5
  - Randomly partition the full sample index set into 32 shards
  - Assign one shard per DataLoader worker
  - Ensures balanced, parallel prefetching for high throughput

The Dataset

```
def __getitem__(self, idx: int) -> DatasetItem
```

- Agent-Centric Samples
- Efficient Batched Loading via HDF5
- Rich Metadata for Analysis and Filtering

same original Argoverse2 scenario

Dataset ($\mathcal{D}$)	# Agents ($N_{\max}$)	# Interest Agents (N_c)	Future Duration (T_f)	# Polylines (K)	Kalman Difficulty	Trajectory Type
av2	15	4	60	256	Easy	Straight
av2	15	4	60	256	Moderate	Straight
av2	15	4	60	256	Hard	Turning
av2	15	4	60	256	Moderate	Turning
av2	22	4	60	256	Easy	Straight

The Dataset

```python
def __getitem__(self, idx: int) -> DatasetItem
```

$$\textbf{obj\_trajs} \quad \in \mathbb{R}^{N_{\text{max}} \times T_p \times F_{\text{ap}}}$$

$$\textbf{center\_gt\_trajs} \quad \in \mathbb{R}^{T_f \times F_{\text{af}}}$$

T_p past timesteps

K # map polylines

$$\textbf{map\_polylines} \quad \in \mathbb{R}^{K \times L \times F_{\text{map}}}$$

N_{max} # agents per sample

T_f future timesteps

L # points per polyline

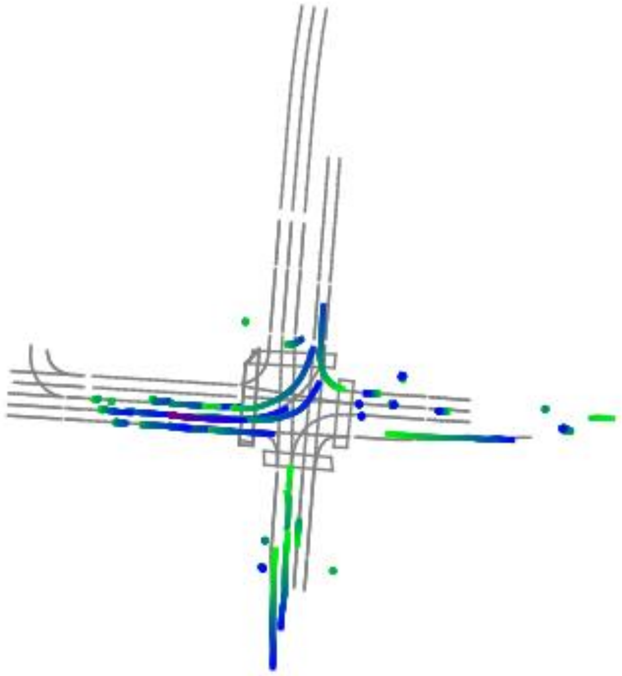
```
The Fap dimension (e.g., 39) consists of:
- [0:3] Relative Position (x, y, z)
- [3:6] (length, width, height)
- [6:11] Object Type one-hot encoding
- [11:11+Tp] One-hot Time encoding
- [A:A+2] Heading Embedding
- [A+2:A+4] Relative Velocity (vx, vy)
- [A+4:Fap] Relative Acceleration (ax, ay)
```

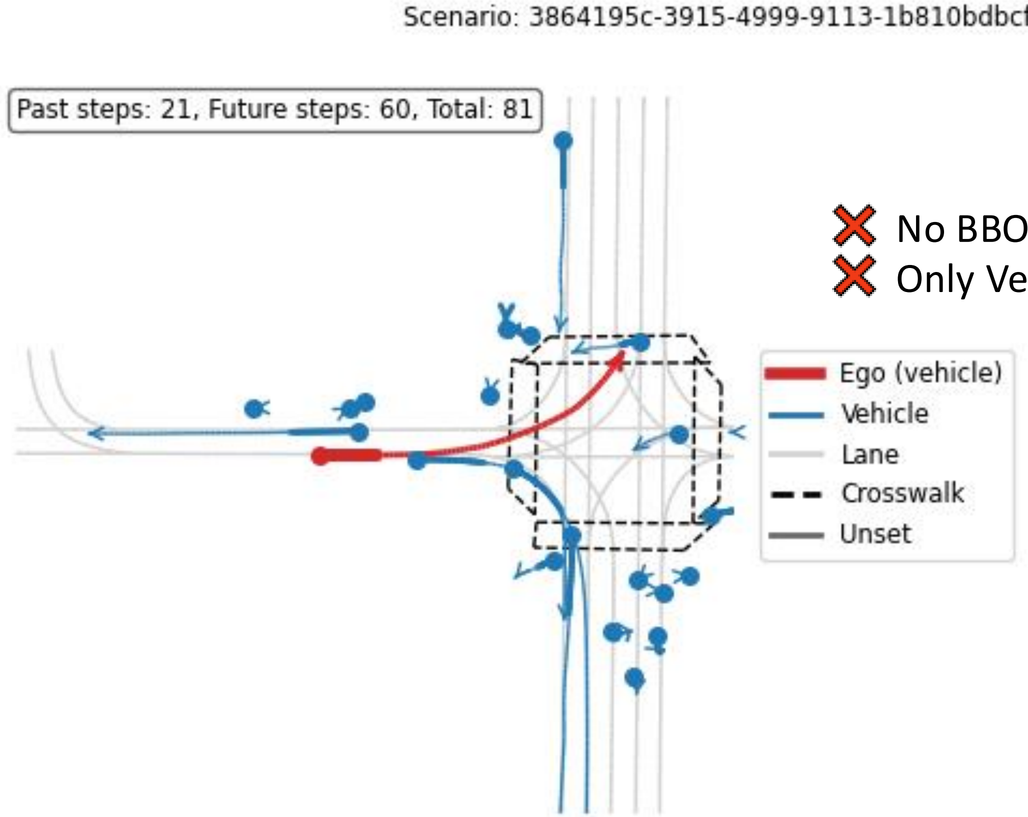
```
The Fmap dimension consists of:
- [0:3] Position (x, y, z)
- [3:6] Direction (x, y, z)
- [6:9] Previous point position (x, y, z)
- [9:29] Lane type one-hot encoding
```

The Dataset

DatasetItem Visualization

<DatasetItem '3864195c-3915-4999-9113-1b810bdbcf48' @ 'av2': Agents=32/64, Traj(P=21, F=60, D_past=39, D_future=4), Map(R=256, L=20, D_map=29), kd=(3,), traj_type=7>





❌ No BBOXs from AV2
❌ Only Vehicles?

Metrics and Optimization

- **Average Displacement Error (ADE):**

$$\text{ADE} = \mathbb{E}_t \left[\|\hat{y}_t - y_t\|_2 \right]$$

- **Final Displacement Error (FDE):**

$$\text{FDE} = \|\hat{y}_T - y_T\|_2$$

- **Miss Rate (MR):**

$$\text{MR} = \mathbb{E}_k \left[\mathbb{1}\left\{ \left\| \hat{y}_T^{(k)} - y_T \right\|_2 > d_{\text{thresh}} \right\} \right]$$

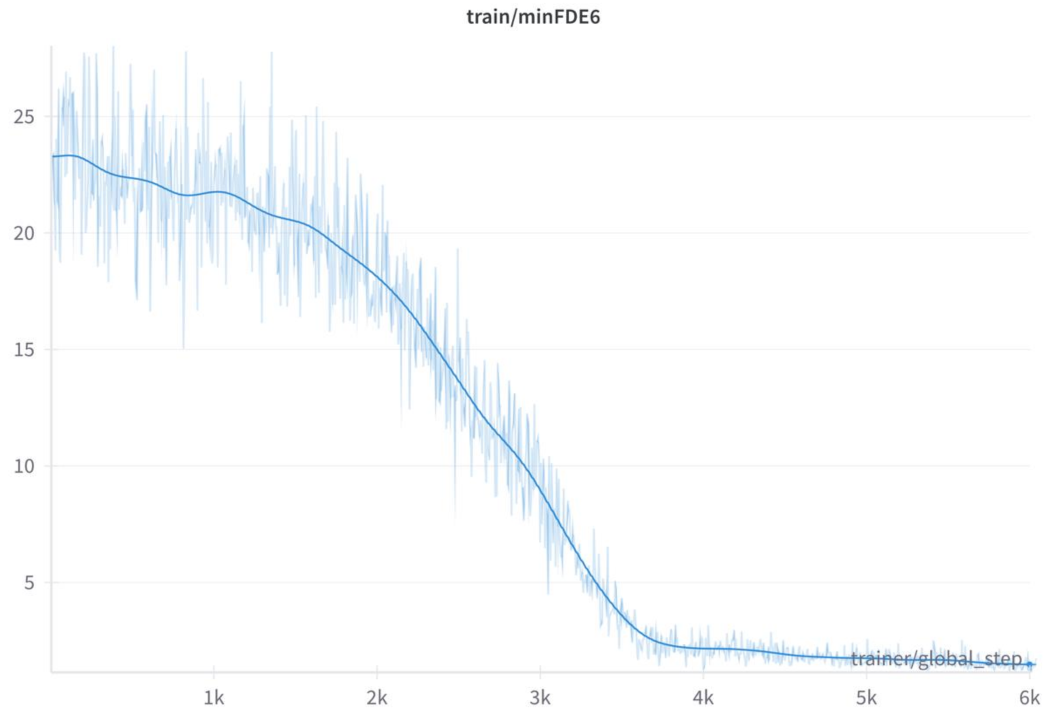
- **Brier Final Displacement Error (Brier FDE):**

$$\text{BrierFDE} = \mathbb{E}_k \left[p_k \cdot \left\| \hat{y}_T^{(k)} - y_T \right\|_2^2 \right]$$

Here, $\hat{y}_T^{(k)}$ denotes the final position of the trajectory of the k-th mode, p_k its predicted probability (after applying *softmax* to all *logits*), and d_{thresh} the miss threshold distance (e.g., 2.0 m).

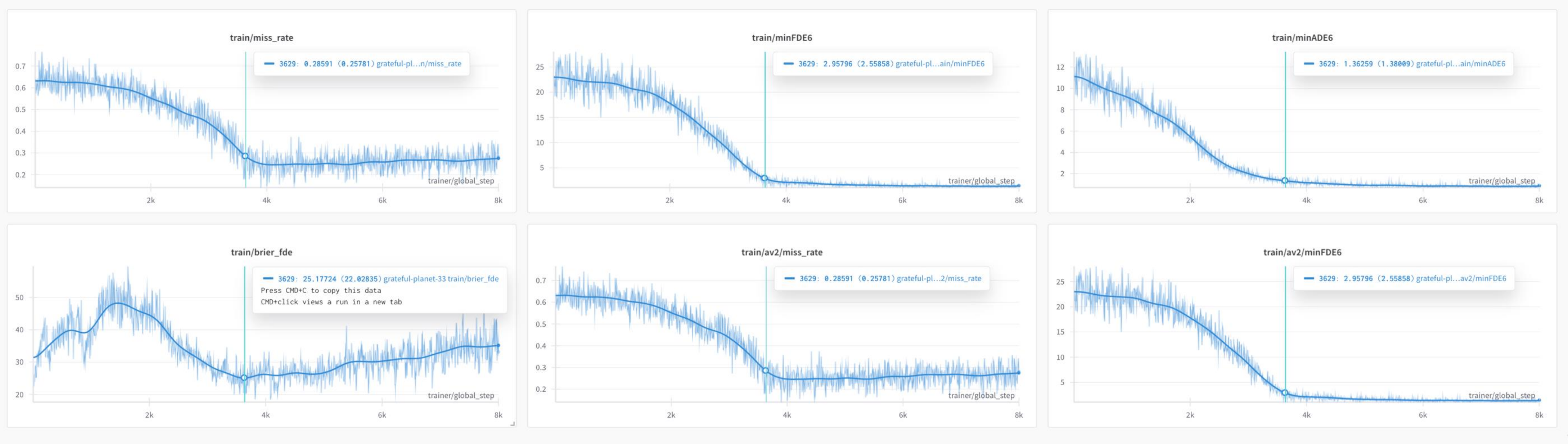
MTR Training

- No pre-trained Model
- Adaption to new UniTraj data parsing and config handling
- Validation results:
  - brier-minFDE: 1.98 (1.98) [1] (2.08) [2]
  - minFDE: 1.6655 (1.3650) [1]
  - minADE: 0.86294 (0.6697) [1]
  - Miss Rate: 0.30141 (0.2111) [1]

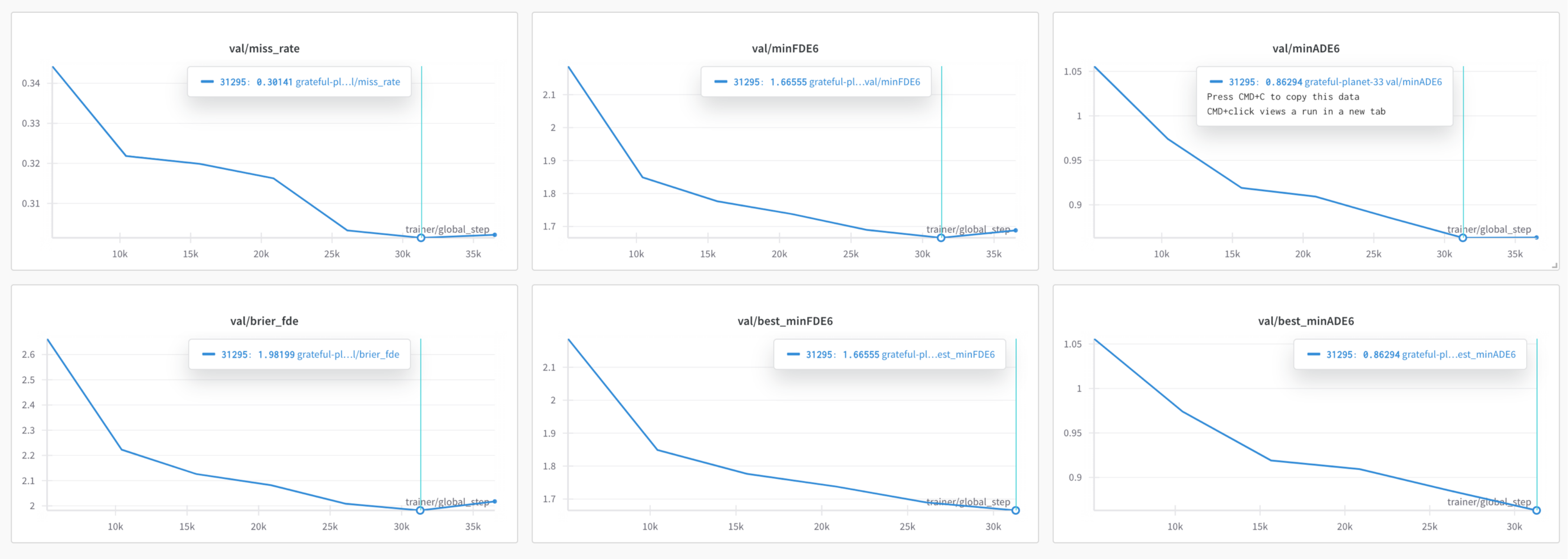


[1]Shaoshuai Shi et al., "Motion Transformer with Global Intention Localization and Local Movement Refinement," arXiv:2209.13508 [cs.CV], 2022, Appendix D
[2] Lan Feng, Mohammadhossein Bahari, Kaouther Messaoud Ben Amor, Éloi Zablocki, Matthieu Cord und Alexandre Alahi. (2024). *UniTraj: A Unified Framework for Scalable Vehicle Trajectory Prediction*. arXiv:2403.15098v3 [cs.CV]

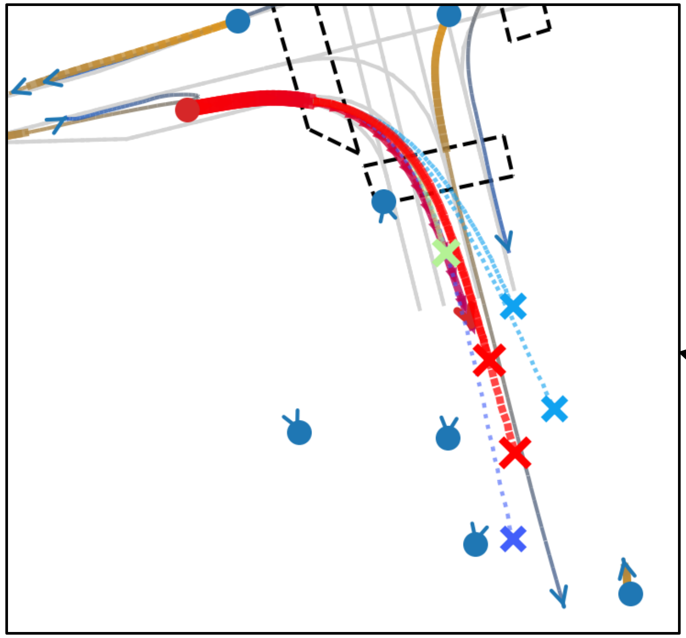
MTR Training

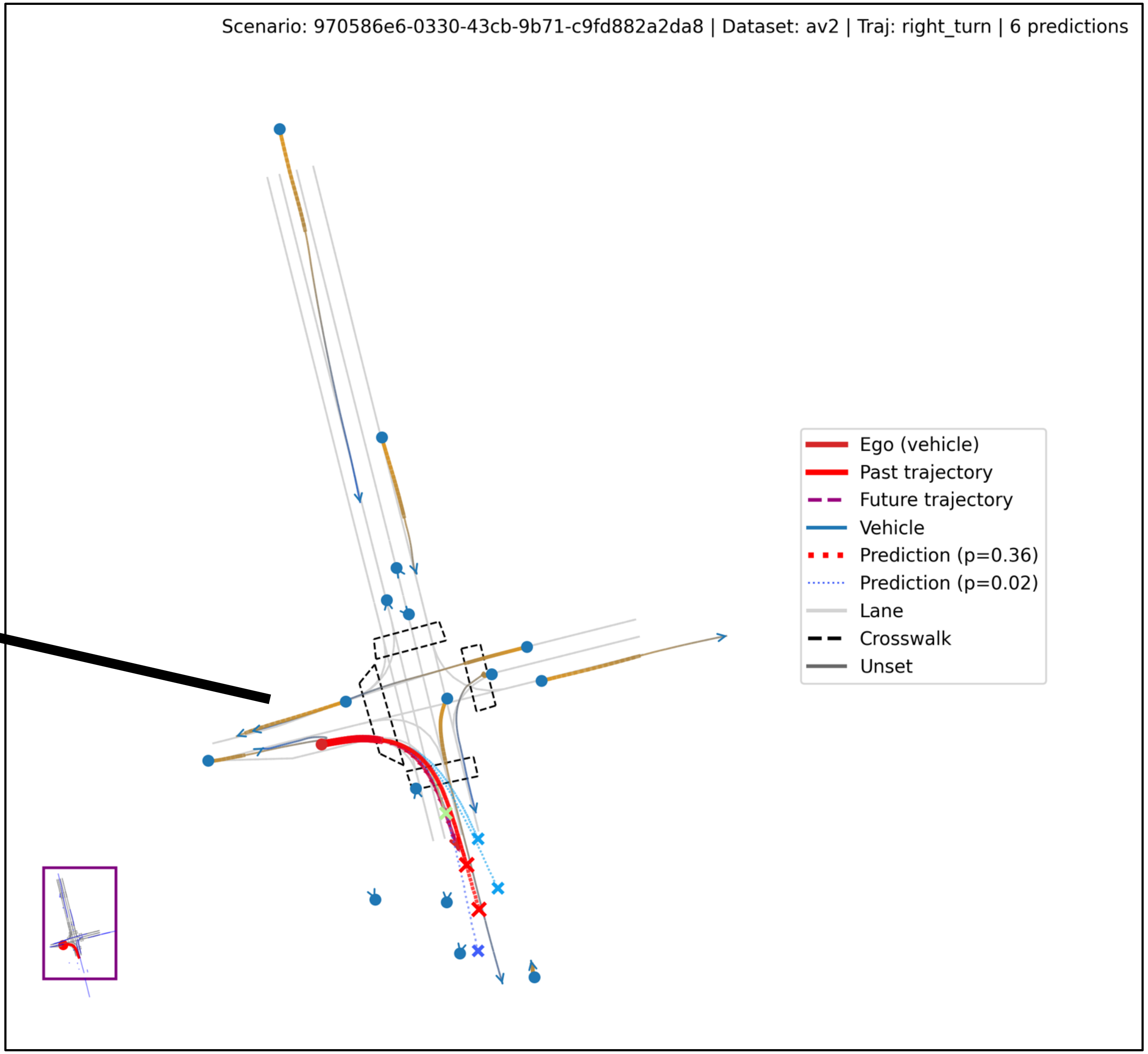


MTR Evaluation



MTR Prediction

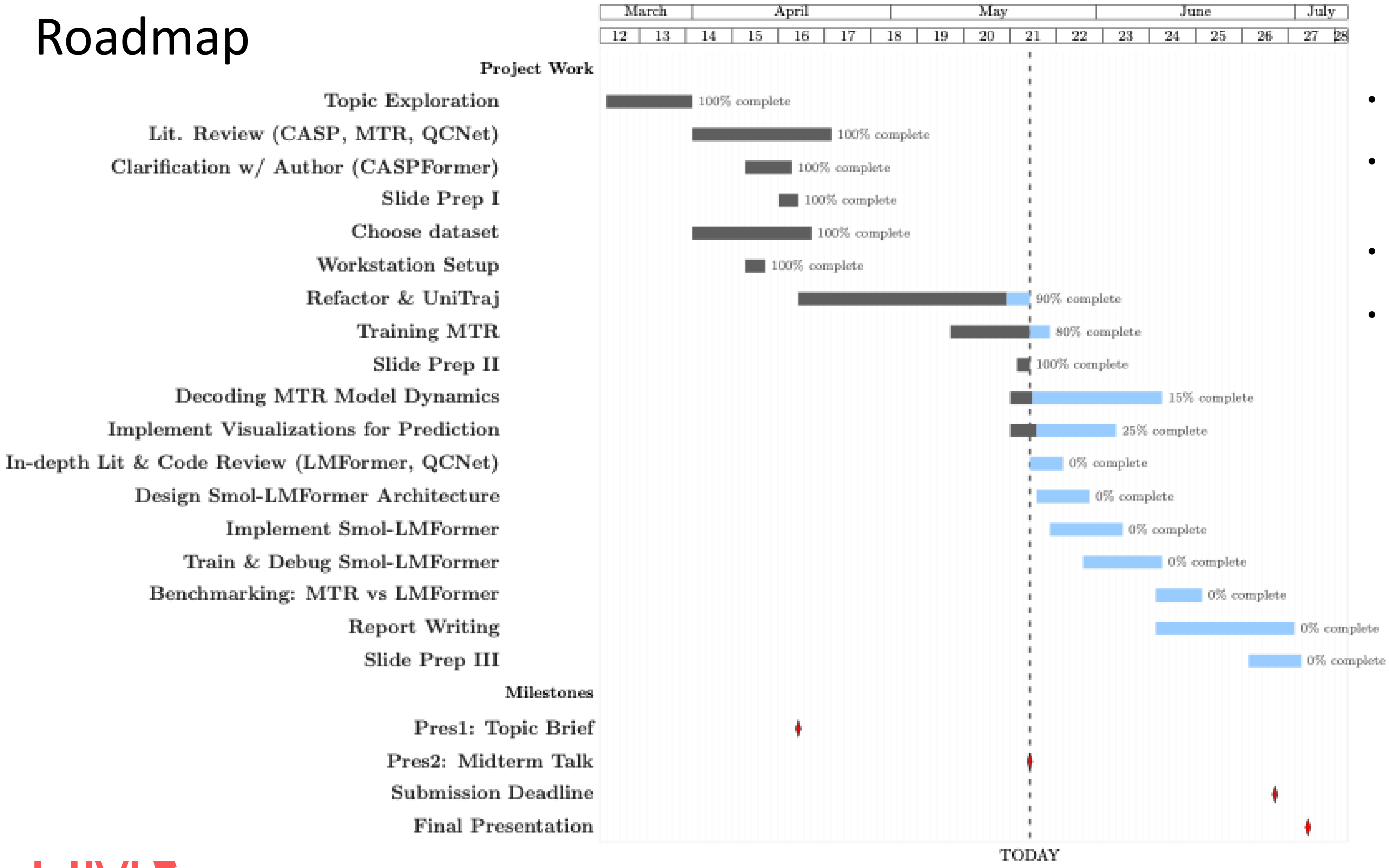




Current Challenges & Issues

- Reduced scope of project – *Very-Smol-Single-Agent-Motion-Forecasting*
- No BBOX's from AV2, only agents of type vehicle.
- Poor coding standards, visualization and documentation within UniTraj
  - *No typing hints, doc-strings, cluttered and opaque code data-processing and config handling, incorrect usage of framework like PyTorch Lightning and WandB, disregarded original train:test:val splits, bad path and file-handling, no logging, no good exception handling, no usage of good design pattern…*

Roadmap



- ML Infrastructure ✅
- Model Components & Metrics ⏳
- Train Reference model (⏳ .. ✅)
- Implement Smol Model ⏳

Discussion