Hochschule München
University of Applied Sciences

Fakultät für Informatik und Mathematik
Department of Computer Sciences and Mathematics

Multimodal Trajectory Prediction in Multi-Agent Scenarios

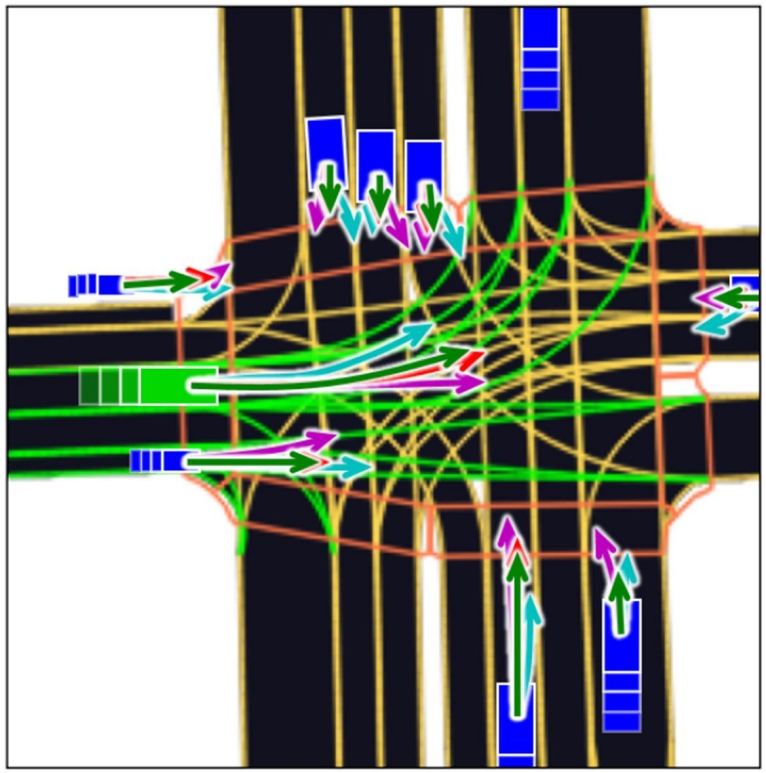
Seminar: Video Analysis & Object Tracking

16 April 2025

Lukas Röß, Jan Duchscherer

Introduction & Motivation

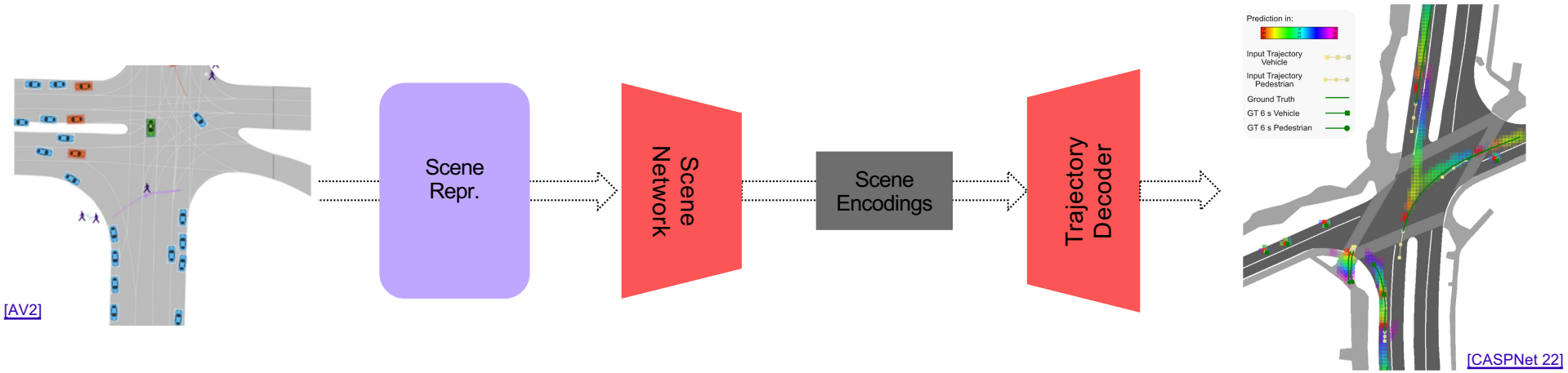
- What is Multimodal Trajectory Prediction?
- Why Multi-Agent Scenarios Matter?
- Challenges



Explainable multimodal trajectory prediction using attention models 24

From Scene to Prediction

The General Motion Forecasting Pipeline



$$\mathbf{X}_d \in \mathbb{R}^{N \times T_{\text{in}} \times \|F_d\|}$$

$$\mathbf{X}_s \in \mathbb{R}^{K \times \|F_s\|}$$

$$\boldsymbol{\pi} \in \mathbb{R}^{N \times M}$$

$$\boldsymbol{\mu}, \mathbf{b} \in \mathbb{R}^{N \times M \times T_{\text{out}} \times 2}$$

$F_d = [x, y, \psi, \dot{x}, \dot{y}, \ddot{x}, \ddot{y}, w, l, c]$

N = # agents

T_{in} = past timesteps

M = # modes

F_s = polyline encodings

K = # map objects

T_{out} = pred horizon

Selected Approaches

CASPFormer (2024):
(SOTA on nuScenes)

- HD Map-Free
- CNN + ConvLSTM Encoder
- Deformable Attention Decoder

MTR (2022):
(MTR v3 SOTA on Waymo)

- Map-Aware
- Intention-Based Queries
- Modular Transformer Design

Project Objective

Main Goal:

- Implement CASPFormer in the UniTraj framework
- Train MTR on dataset within UniTraj if no pre-trained model will be shared

Comparative Analysis:

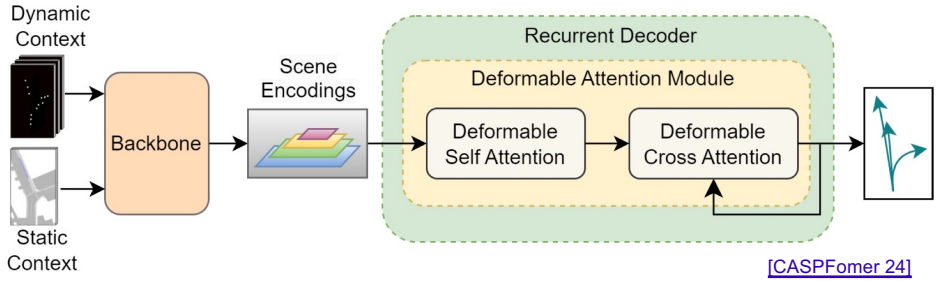
- Compare models based on algorithmic structure
- Evaluate both models on a shared dataset using key metrics:
  - Average Displacement Error (ADE)
  - Final Displacement Error (FDE)
  - Miss Rate (MR)

Challenges:

- CASPFormer source code not available
  - Use fallback
  - Or implement SmolCASPFormer
- Adapt data processing for the dataset of selected approaches
- Train/Fine-tune MTR Model
- Ensure fair and comparable evaluation, especially if different dataset are used datasets

CASPFormer

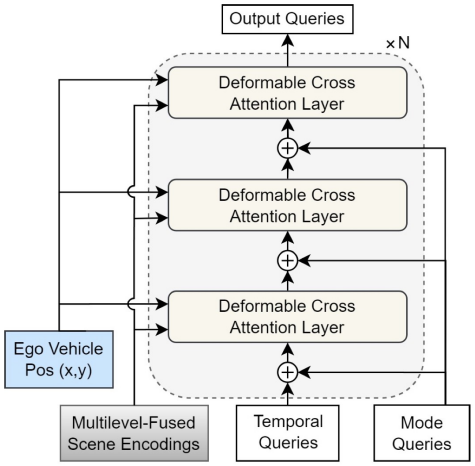
BEV (Birds Eye View)
⟹ CNN + ConvLSTM
⟹ Multi-Scale Scene Encodings
⟹ Deformable (recurrent) Transformer
(Mode Queries and Temporal Queries)



[CASPFomer 24]

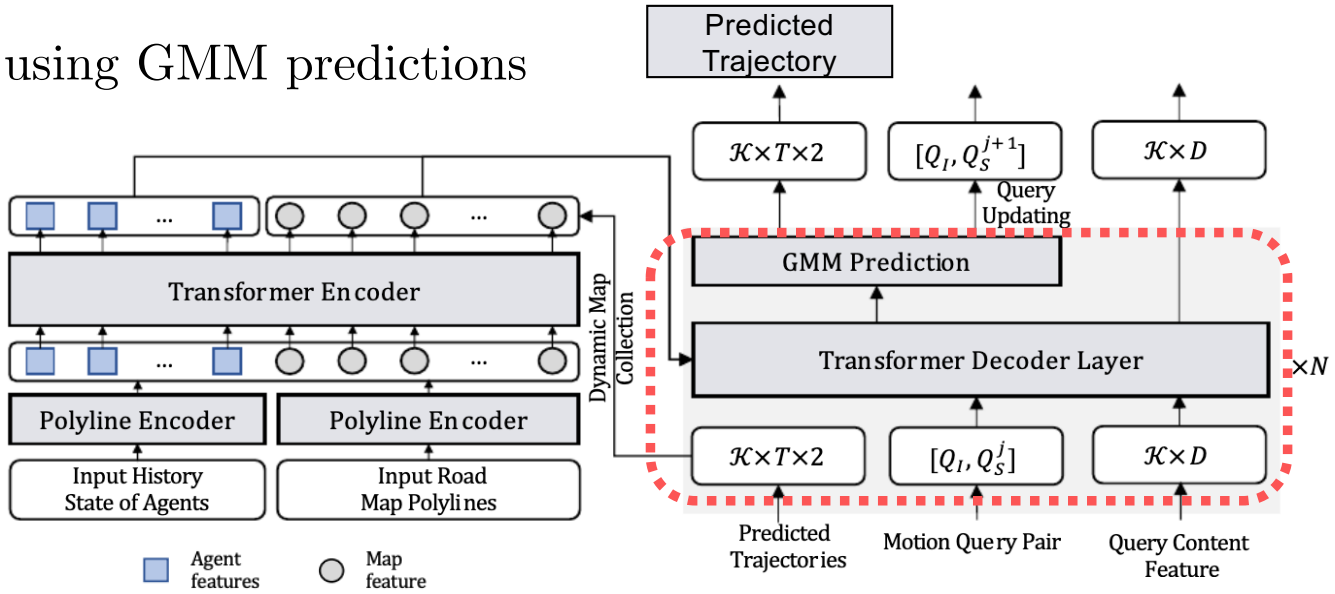
⚡ No Code available!
⟹ Try implementing smol version from scratch? 🥺
⟹ QCNet as source code blueprint
(loss functions, train pipeline, encodings...)

⟹ Most likely will not yield a working prototype!



Motion TRansformer (MTR)

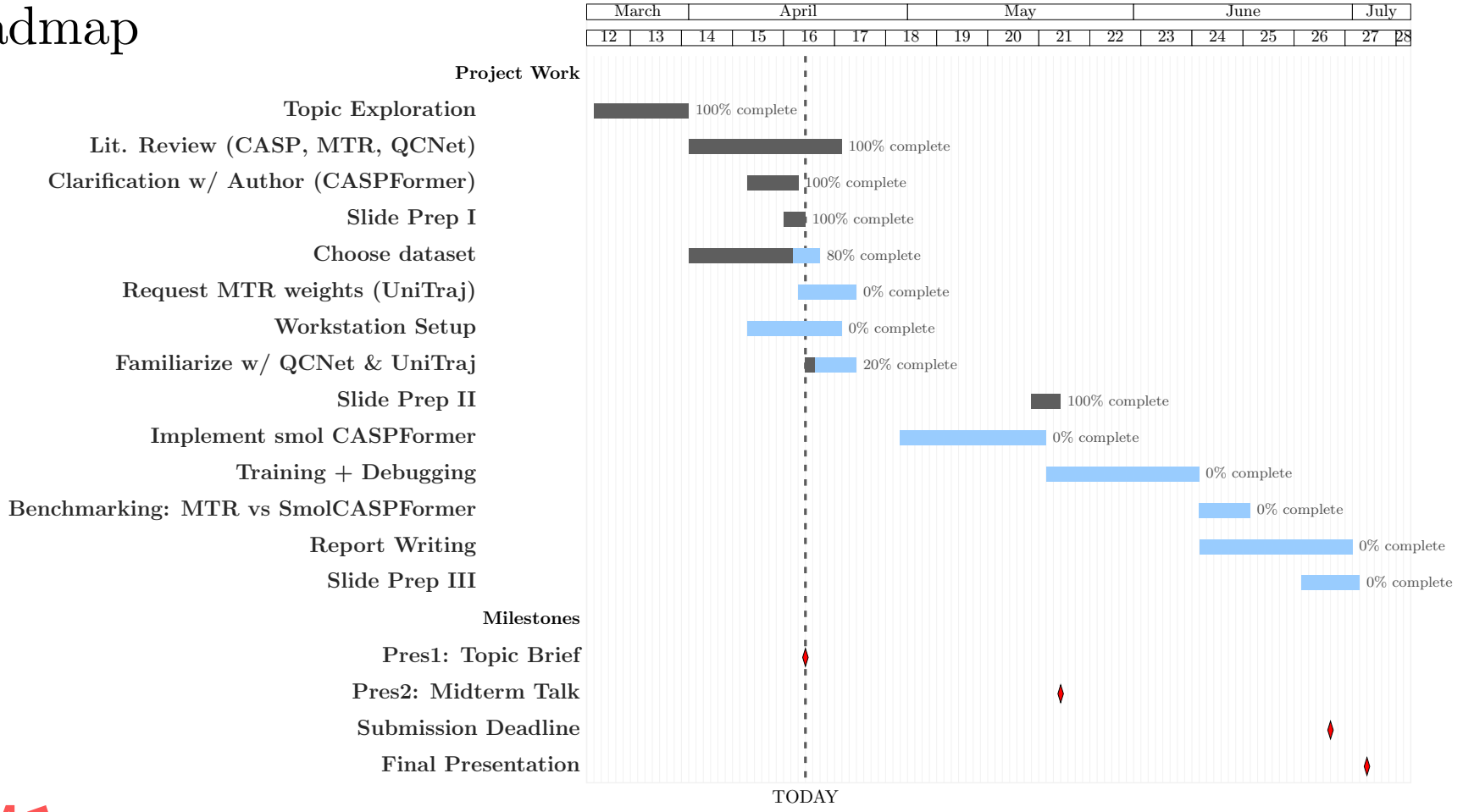
- Modular Design
- Goal-conditioned decoding
- Multimodal output using GMM predictions



MTR 22

--- Motion Decoder

Roadmap


March
April
May
June
July
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
Project Work
Topic Exploration
100% complete
Lit. Review (CASP, MTR, QCNet)
100% complete
Clarification w/ Author (CASPFormer)
100% complete
Slide Prep I
100% complete
Choose dataset
80% complete
Request MTR weights (UniTraj)
0% complete
Workstation Setup
0% complete
Familiarize w/ QCNet & UniTraj
20% complete
Slide Prep II
100% complete
Implement smol CASPFormer
0% complete
Training + Debugging
0% complete
Benchmarking: MTR vs SmolCASPFormer
0% complete
Report Writing
0% complete
Slide Prep III
0% complete
Milestones
Pres1: Topic Brief
Pres2: Midterm Talk
Submission Deadline
Final Presentation
TODAY

Discussion