Thesis for Master's Degree

Robust Imitation Learning with Attention Constraints and Risk Awareness for Motion Planning in Autonomous Driving

Jiyun Kim

Artificial Intelligence Graduate School

Gwangju Institute of Science and Technology

2026

Robust Imitation Learning with Attention Constraints and Risk Awareness for Motion Planning in Autonomous Driving

어텐션 제약과 위험 인지를 활용한 강건한 모방학습 기반 자율주행 모션 플래닝

Robust Imitation Learning with Attention Constraints and Risk Awareness for Motion Planning in Autonomous Driving

Advisor: Ue-hwan Kim

by

Jiyun Kim

Artificial Intelligence Graduate School

Gwangju Institute of Science and Technology

A thesis submitted to the faculty of the Gwangju Institute of Science and Technology in partial fulfillment of the requirements for the degree of Master of Science in the Artificial Intelligence Graduate School

Gwangju, Republic of Korea

December 09, 2025

Approved by

Professor Ue-hwan Kim

Committee Chair

Robust Imitation Learning with Attention Constraints and Risk Awareness for Motion Planning in Autonomous Driving

Jiyun Kim

Accepted in partial fulfillment of the requirements for

the degree of Master of Science

December 09, 2025

Committee Chair _____

Prof. Ue-hwan Kim

Committee Member _____

Prof. Kyung Joong Kim

Committee Member _____

Prof. Kyunghwan Choi

For my family, whose support made this possible

MS/AI 20241040 Jiyun Kim(김지윤). Robust Imitation Learning with Attention Constraints and Risk Awareness for Motion Planning in Autonomous Driving (어텐션 제약과 위험 인지를 활용한 강건한 모방학습 기반 자율주행 모션 플래닝). 2026. 56p. Advisor: Prof. Ue-hwan Kim.

Abstract

Imitation learning (IL) has become a common approach for autonomous-driving motion planning, but achieving robust and safe behavior remains difficult under distribution shift and rare hazard scenarios. A key limitation is that IL objectives typically prioritize mean accuracy (e.g., Average Distance Error), which can high-consequence failures in the long tail. Moreover, planners that use attention mechanisms, including Transformer-based planners, can exhibit attention collapse such as shortcut learning by over-relying on a small subset of state channels. This dissertation proposes a robust IL framework for motion planning that combines attention constraints and risk awareness. Mean-Deviation Constrained Attention (MDCA) promotes balanced state usage, and a Conditional Value-at-Risk (CVaR)-based tail-risk objective down-weights rare but potentially dangerous modes. Experiments on the nuPlan benchmark show improved robustness and safety-related metrics over State-Of-The-Arts planners.

MS/AI 20241040 Jiyun Kim(김지윤). Robust Imitation Learning with Attention Constraints and Risk Awareness for Motion Planning in Autonomous Driving (어텐션 제약과 위험 인지를 활용한 강건한 모방학습 기반 자율주행 모션 플래닝). AI대학원(학과). 2026. 56p. 지도교수: 김의환 교수님.

국 문 요 약

Imitation Learning(IL)은 자율주행에서 널리 사용되지만, 분포 변화나 위험 상황에서도 강건한 주행을 보장하기는 어렵다. 이는 학습 목표가 대게 평균 정확도(Average Distance Error)를 우선시하여, 발생 빈도는 낮지만 치명적인 실패를 충분히 반영하지 못하기 때문이다. 이렇게 최적화된 모델은 tail events에 대하여 강건하지 못해 tail risk를 남기는 한계로 이어지며, Transformer 기반 플래너와 같은 attention mechanism을 사용하는 플래너들은 ego-state의 일부에 과도하게 가중치가 집중되는 attention collapse 현상을 보일 수 있고, 이는 shortcut learning으로도 불린다. 본 논문은 이를 해결하고자 attention constraints와 risk awareness를 결합한 IL기반 강건 모션 플래닝 프레임워크를 제안한다. Mean-Deviation Constrained Attention(MDCA)으로 상태 정보를 균형있게 만들고, Conditional Value-at-Risk(CVaR) 기반의 tail-risk objective는 위험한 모드의 비중을 낮춘다. nuPlan 벤치마크 실험에서 제안 방법은 State-Of-The-Art 학습기반 플래너들 대비 강건성과 안전 관련 지표를 개선했으며, hazard 환경에서 큰 향상을 보였다.

Contents

List of Figures

Chapter 1

Introduction

1.1 Research background

Achieving motion planning that remains robust and safe across diverse, complex real-world scenarios is a central challenge in autonomous driving [1]. While early systems relied heavily on modular pipelines with hand-crafted rules and cost functions, recent progress has increasingly shifted toward data-driven approaches that learn driving policies from large-scale logs. Among them, Imitation Learning (IL) [2–5] has become a widely adopted paradigm, enabling planners to mimic expert demonstrations and capture rich interactions without exhaustive manual engineering. Despite its simplicity and strong empirical performance, IL does not inherently guarantee reliable behavior when the test-time conditions deviate from the training distribution.

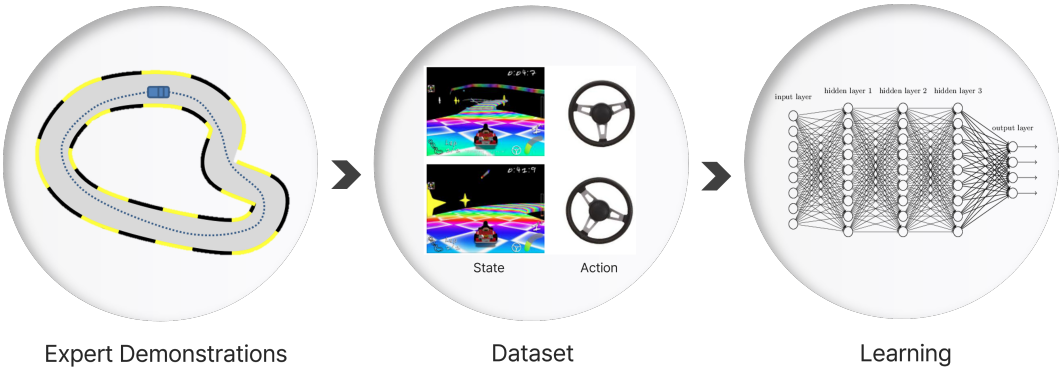


Figure 1.1: Imitation learning nutshell.

As illustrated in Figure 1.1, imitation learning learns a policy directly from expert

demonstrations, providing a practical route to autonomous decision making without explicitly specifying hand-crafted rules or cost functions. This paradigm is especially attractive in autonomous driving, where designing a complete reward or optimization objective that captures human-like driving and safety constraints is difficult, and where collecting large amounts of labeled outcomes for rare events can be expensive [6–10]. Instead, one can leverage driving logs recorded from human experts and treat them as supervision signals.

A standard formulation converts expert demonstrations into a dataset of state-action pairs. The state encodes what the agent observes about the environment at a given time step, and the action corresponds to the expert's control decision under that state. In the context of the illustrative example in Figure 1.1, a video-game setting can be interpreted similarly: the state may consist of rendered frames, HUD cues, and game context, while the action may consist of steering, acceleration, and braking inputs. Learning then amounts to training a policy that maps states to actions so that the agent reproduces the expert's behavior. In autonomous driving, the state typically includes ego kinematics, surrounding agents, and map context, and the action is represented as a planned future trajectory or low-level control; in both cases, the central idea is to learn the policy by matching expert decisions from data.

A key reason for the recent success of modern planning architectures is the ability to learn representations that integrate heterogeneous scene information at scale. Earlier sequence models such as RNNs [11, 12] process inputs step-by-step and can suffer from forgetting long-range information due to vanishing gradients, while CNN-style

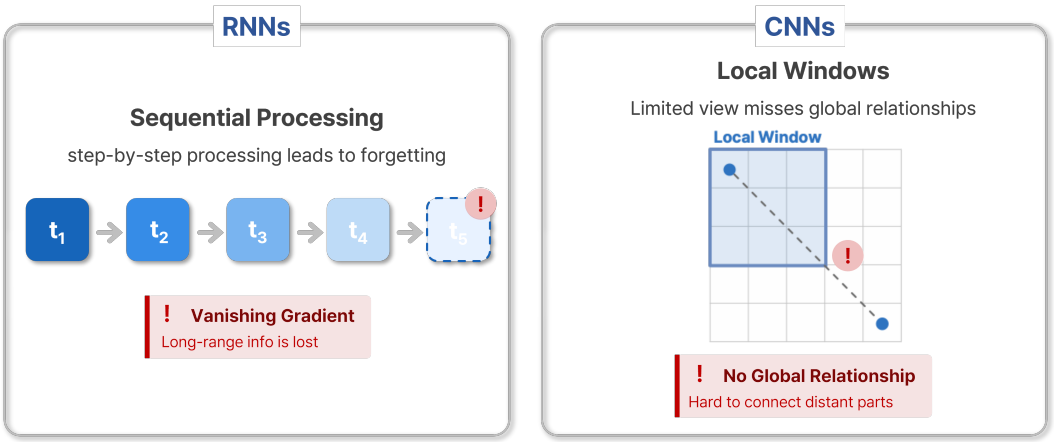


Figure 1.2: Limitations of sequential models and local-window models, motivating attention for global dependency modeling.

local processing [13] is limited by receptive fields and may miss global relationships (Figure 1.2). Attention mechanisms [14] address these limitations by directly modeling relationships between distant elements, enabling global context aggregation without relying on strictly sequential propagation or local windows. This property is particularly well matched to autonomous driving, where planning depends on long-range map geometry, multi-agent interactions, and subtle context cues that may be spatially and temporally distant. Consequently, Transformer-based planners [15–17] and attention-driven fusion have become a dominant design choice for learning-based motion planning, as they can dynamically emphasize the most relevant agents, lanes, and ego-state cues conditioned on the current situation. However, learning-based planners still face fundamental robustness challenges when deployed in interactive, closed-loop environments.

A well-known limitation of IL in sequential decision-making is compounding error [18]: small deviations from expert behavior can alter future observations and ac-

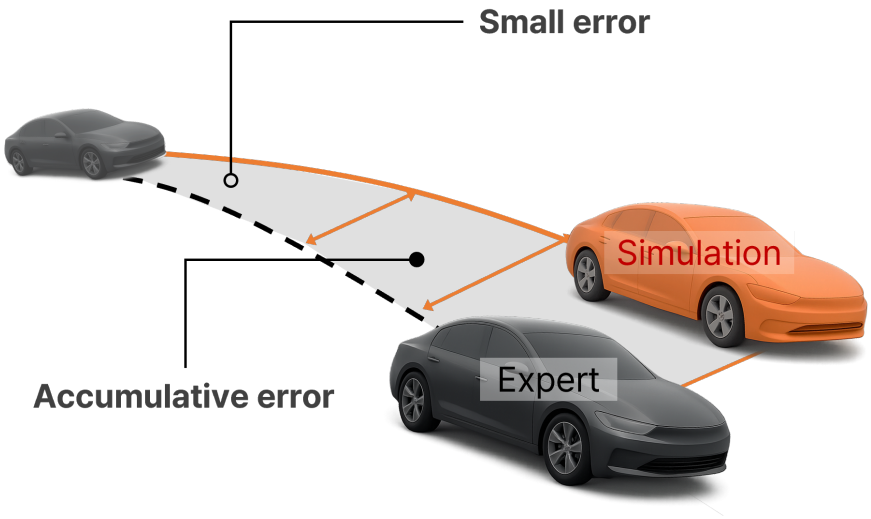


Figure 1.3: Compounding error in imitation learning.

cumulate over time, causing cascading failures during deployment. Prior work has explored strategies such as data augmentation [6, 19] and closed-loop data collection to alleviate this issue. However, even when compounding error is partially mitigated, IL-based planners can remain fragile due to a more subtle failure mode called shortcut learning [20–22].

Shortcut learning arises when the policy exploits spurious correlations in the training data instead of relying on causal cues that are necessary for safe driving [23, 24]. In autonomous driving, such shortcuts can be particularly harmful because they may appear successful under common conditions yet fail abruptly under occlusions, sensor noise, novel environments, or rare interactive situations, thereby degrading robustness and safety (Figure 1.4). This concern becomes more pronounced in attention-based planners (including Transformer planners), because shortcut-inducing correlations can be amplified when attention becomes overly concentrated on a narrow subset of tokens or ego-state channels, effectively ignoring complementary cues needed for stable

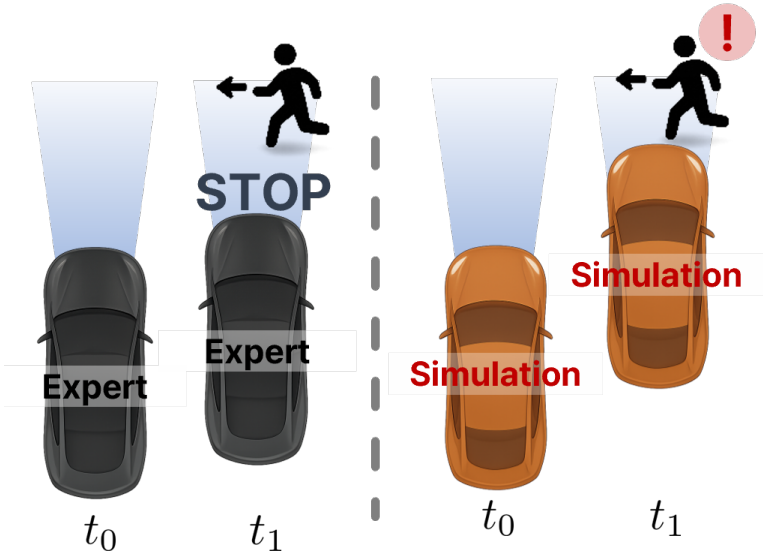


Figure 1.4: An example of shortcut learning under distribution shift: a policy that appears correct on common data can fail in simulation when a critical cue changes.

decision-making under distribution shifts. In this sense, attention is powerful for discovering structure, but excessively peaked attention can lead to brittle representations and make shortcut learning harder to avoid [25, 26].

In addition to representation-level brittleness, the learning objective itself can limit safety awareness. Many IL planners are trained with mean-focused losses that emphasize average trajectory accuracy (e.g., Average Distance Error). While such objectives improve nominal performance, safety-critical events are inherently rare and reside in the tail of the distribution. As a result, a planner optimized for average accuracy may still assign non-negligible probability mass to low-frequency but high-consequence trajectory modes, which can lead to hazardous behavior in long-tail scenarios. This motivates risk-aware formulations that explicitly account for tail events, such as Conditional Value-at-Risk (CVaR) [27–31], which quantifies adverse outcomes beyond average-case performance.

These observations suggest that robust and safe learning-based motion planning requires improvements at two complementary levels. First, the planner should reduce shortcut learning in feature representations by stabilizing attention allocation and discouraging over-reliance on a small subset of state channels. Second, the planner should incorporate risk awareness in multimodal decision making to suppress rare but potentially dangerous modes. This dissertation develops a robust imitation learning framework that addresses both aspects through attention constraints and tail-risk-aware trajectory scoring.

1.2 Research Motivation and Objectives

Motivated by the vulnerability of attention-based planners to shortcut learning and the limited safety supervision provided by mean-focused IL objectives, this dissertation develops a robust motion planning framework that promotes safety at both the feature-representation level and the trajectory-selection level.

The primary objective is to design an attention-based motion planner that is robust to distribution shift and sensitive to tail risks. To achieve this goal, the dissertation pursues the following research objectives:

1. **Mitigating shortcut learning via constrained attention.** To prevent the model from overfitting to spurious features, this dissertation introduces an Augmented Lagrangian Method (ALM)-based regularization strategy that constrains attention allocation across ego-state channels. The proposed constraint preserves relative importance among channels while encouraging a more balanced utiliza-

tion of the state representation, thereby improving robustness under distribution shifts.

2. **Improving safety via risk-aware learning and trajectory scoring.** To address the limitation of mean-based objectives, this dissertation incorporates a mode-wise tail-risk measure based on Conditional Value-at-Risk (CVaR). CVaR is used to construct risk-aware soft targets for multimodal outputs and to down-weight modes that are rare but potentially hazardous. This risk awareness further supports safer trajectory selection by reducing probability mass assigned to dangerous modes and discouraging high-likelihood yet unsafe behaviors in long-tail scenarios.

1.3 Contributions of the Dissertation

This dissertation presents a robust motion planning framework that addresses two central limitations of data-driven imitation learning for autonomous driving: (i) shortcut learning in feature representation and (ii) insufficient treatment of tail risks in trajectory selection. By integrating constrained optimization into attention mechanisms and incorporating risk-aware criteria into learning and decision making, the dissertation makes the following contributions:

1. **Constrained attention regularization for robust ego-state utilization.** This dissertation proposes a novel regularization method that formulates attention dispersion over ego-state channels as a constrained optimization problem, preventing attention collapse where the planner over-relies on a small subset of

state channels. In contrast to stochastic masking approaches (e.g., state dropout encoders) that discard input information, the proposed ALM-based constraint encourages balanced feature usage without information loss. The resulting regularizer is theoretically grounded, incurs negligible training overhead, and preserves the original inference architecture, improving robustness under distribution shifts on the nuPlan benchmark.

2. **Risk-aware trajectory scoring with a CVaR-based tail-risk measure.** To overcome the limitation of mean-focused IL objectives, this dissertation introduces a CVaR-based mechanism that explicitly quantifies tail risk at the mode level for multimodal planning. By down-weighting or filtering trajectory modes that exhibit high estimated risk despite high imitation likelihood, the approach reduces the probability mass assigned to rare but potentially dangerous behaviors, acting as a complementary safety layer in long-tail hazard scenarios.

3. **A unified framework combining attention constraints and risk awareness.** Finally, this dissertation integrates constrained attention regularization and CVaR-based risk awareness into a single planning framework. Comprehensive experiments on the nuPlan benchmark, including challenging hazard-focused settings, demonstrate that the proposed approach improves both robustness and safety-related performance compared to state-of-the-art learning-based planners. The results show consistent gains in closed-loop driving stability and collision avoidance, supporting the effectiveness of addressing robustness and safety jointly.

Chapter 2

Mean-Deviation Constrained Attention

2.1 Related Work: PlanTF

PlanTF [32] is a Transformer-based imitation planner developed and evaluated on the nuPlan benchmark [33, 34], with an emphasis on systematically revisiting which input features and which data augmentation strategies lead to reliable behavior. An important finding is that driving performance is not determined by model capacity alone. The planner's robustness and generalization depend strongly on whether the learned policy relies on causal cues from the environment or exploits spurious correlations present in the training data.

PlanTF encodes heterogeneous scene information such as ego kinematic states, surrounding agents, and map elements into token-like representations and uses attention modules to model their interactions. This attention-based fusion allows the planner to select relevant agents and map cues dynamically, rather than depending on pre-defined heuristics. Because urban driving often admits multiple plausible futures, PlanTF adopts a multi-modal prediction formulation that outputs multiple trajectory hypotheses and selects a suitable candidate under the learned scoring and selection mechanism. This design enables the model to represent uncertainty and capture diverse interactive outcomes in complex scenarios [35–37].

An important observation highlighted in PlanTF is that incorporating historical motion (e.g., past states or past trajectories) is not always beneficial. Although motion history can provide useful context, it can also encourage shortcut learning, where the model overfits to correlations in past trajectories instead of grounding its decisions in the surrounding scene. In such cases, the planner may appear strong under offline imitation metrics while exhibiting degraded behavior in closed-loop execution, particularly under occlusion, distribution shift, or rare hard scenarios [32, 38]. Conceptually, the policy can drift toward replaying patterns from the past rather than reasoning about why certain actions are required given the current environment and interactions.

To mitigate these false correlations, PlanTF introduces an attention-based State Dropout Encoder (SDE). The SDE stochastically drops [39] or masks parts of kinematic state information at the encoding stage, discouraging the network from consistently exploiting the same state cues as a shortcut. By perturbing state inputs during training, the model is encouraged to leverage a broader set of signals—especially surrounding agents and map geometry—thereby improving robustness and reducing reliance on misleading patterns. From this perspective, SDE serves as a practical regularization mechanism aimed at stabilizing learning in the presence of potentially shortcut-inducing inputs such as motion history.

Another notable aspect of PlanTF is that it provides a strong purely learning-based baseline that achieves competitive performance against state-of-the-art rule-based methods, without relying on hand-crafted driving heuristics. This makes PlanTF an appealing foundation for studying how additional principled constraints or safety

mechanisms can be integrated into Transformer-based imitation planning.

However, state dropout remains a heuristic that perturbs inputs and does not explicitly regulate how attention is allocated across tokens. Even with dropout, the model may still concentrate attention excessively on a small subset of remaining tokens, leading to attention collapse and brittle behavior [40] when distributed reasoning over multiple agents and constraints is necessary. Therefore, while PlanTF and SDE provide an effective starting point for reducing shortcut learning, there remains a need for a more principled approach that directly stabilizes attention allocation. Motivated by this limitation, the next sections introduce Mean-Deviation Constrained Attention (MDCA), which formulates attention regularization as a constrained optimization problem and enforces it via an Augmented Lagrangian Method (ALM)-based update rule.

2.2 Problem Formulation as a Constrained Optimization Problem

The proposed Mean-Deviation Constrained Attention (MDCA) is formulated as a constrained optimization problem [41]. The primary objective is to prevent attention collapse in ego-state encoding, which often leads to shortcut learning where the planner over-relies on a small subset of state channels.

To establish the mathematical basis for this constraint, the cross-attention mechanism used to aggregate ego-state features is first defined. In this architecture, the ego-state is summarized by a learnable query that attends to various input channels. Let $Q \in \mathbb{R}^{1\times d}$ be the learnable query vector representing the ego-vehicle, and let $K \in \mathbb{R}^{C\times d}$ and $V \in \mathbb{R}^{C\times d}$ be the keys and values projected from C ego-state channels.

The attention scores e_i are computed by the scaled dot-product between the query and each key:

$$e_i = \frac{QK_i^\top}{\sqrt{d}}, \quad \forall i \in \{1, \ldots, C\}, \tag{2.1}$$

where d is the feature dimension and K_i denotes the i-th channel's key vector. These scores are then normalized through a Softmax function to produce the attention weights a_θ:

$$a_{\theta,i} = \frac{\exp(e_i)}{\sum_{j=1}^{C} \exp(e_j)}, \quad a_\theta = \left[a_{\theta,1}, \ldots, a_{\theta,C}\right]^\top \in \Delta^{C-1}. \tag{2.2}$$

The resulting distribution a_θ exists on the probability simplex, where each weight $a_{\theta,i} \geq 0$ and $\sum_{i=1}^{C} a_{\theta,i} = 1$. These weights determine how much information is extracted from each ego-state channel to form the final representation.

MDCA regulates these weights to ensure that the model does not over-rely on a few specific features. For a minibatch of size B, let $a_\theta^{(n)}$ be the attention distribution for the n-th sample. To quantify the concentration level, define the mean-deviation statistic $D(\theta)$ as:

$$D(\theta) = \frac{1}{BC} \sum_{n=1}^{B} \sum_{i=1}^{C} \left| a_{\theta,i}^{(n)} - \frac{1}{C} \right|. \tag{2.3}$$

When attention collapses onto a small subset of channels, $D(\theta)$ increases. Conversely, if the weights are spread evenly, $D(\theta)$ decreases. MDCA constrains this deviation by introducing the following inequality constraint:

$$g(\theta) = D(\theta) - m \leq 0, \tag{2.4}$$

where $m > 0$ is a fixed margin that defines the maximum allowable deviation from a uniform distribution. This constraint prevents pathological attention concentration while maintaining the model's ability to focus on the most informative features.

Let $\mathcal{L}_{\text{task}}(\theta)$ denote the base imitation learning objective, such as trajectory regression and mode classification. The training objective with MDCA is formulated as a constrained optimization problem:

$$\min_{\theta} \ \mathcal{L}_{\text{task}}(\theta) \quad \text{s.t.} \quad g(\theta) \leq 0. \tag{2.5}$$

2.3 ALM-based Update Rule

The constrained optimization problem is solved using the Augmented Lagrangian Method (ALM) [42–45]. For the inequality constraint $g(\theta) \leq 0$, a hinge operator $[z]_+ = \max(z, 0)$ is employed, which ensures that the penalty is activated only when the attention dispersion exceeds the specified margin m.

Let $\lambda \geq 0$ be the Lagrange multiplier and $\rho > 0$ be the penalty parameter. The augmented Lagrangian objective for the primal parameters θ is defined as:

$$\mathcal{L}_{\text{aug}}(\theta, \lambda) = \mathcal{L}_{\text{task}}(\theta) + \lambda[g(\theta)]_+ + \frac{\rho}{2}\big([g(\theta)]_+\big)^2. \tag{2.6}$$

In each training iteration, the parameters θ are updated via backpropagation:

$$\theta \leftarrow \theta - \eta \, \nabla_\theta \mathcal{L}_{\text{aug}}(\theta, \lambda), \tag{2.7}$$

where η is the learning rate. After the primal update, the Lagrange multiplier λ is updated to adaptively adjust the constraint strength based on the degree of violation:

$$\lambda \leftarrow \max\Big(0,\ \lambda + \rho[g(\theta)]_+\Big). \tag{2.8}$$

In this framework, the penalty parameter ρ is kept constant throughout the training process. Keeping ρ constant simplifies hyperparameter tuning, while the multiplier update in Eq. (2.8) adaptively adjusts the effective constraint strength based on the magnitude of the current violation. This mechanism encourages the model to stay within the allowable deviation budget without necessitating a complex schedule for ρ, effectively balancing task performance and robust attention allocation.

Chapter 3

Risk-aware Motion Planning

3.1 Related Work : RAIL

Risk-aware imitation learning has been studied as a way to improve reliability in safety-critical domains by explicitly penalizing rare but high-consequence failures. A representative approach is Risk-Averse Imitation Learning (RAIL) [46], which extends Generative Adversarial Imitation Learning (GAIL) [47] by introducing a tail-risk objective on trajectory costs.

RAIL is motivated by the observation that policies learned with a risk-neutral objective can exhibit heavy-tailed distributions of trajectory costs, implying that catastrophic outcomes may occur more frequently than in expert behavior even when average performance is competitive. To quantify and reduce such tail events, RAIL adopts Conditional Value-at-Risk (CVaR) [27] as a conservative risk measure. CVaR focuses on the expected loss within the worst $(1-\alpha)$-fraction of outcomes, thereby directly targeting the upper tail of the cost distribution rather than its mean.

Concretely, RAIL defines a trajectory-cost random variable under the learned policy and augments the original GAIL objective with a CVaR term. Using an auxiliary variable ν and the standard hinge operator, the method constructs a differentiable surrogate $H_\alpha(\cdot, \nu)$ for CVaR and optimizes it jointly with the adversarial imitation

objective. A trade-off coefficient λ_{CVaR} controls the relative emphasis on tail-risk minimization versus imitation, and the formulation recovers the risk-neutral baseline when $\lambda_{\text{CVaR}} = 0$.

RAIL demonstrates that integrating a CVaR-based objective can reduce tail risk while maintaining comparable average performance on benchmark control tasks. This line of work is closely related to risk-aware motion planning, as it highlights the limitation of mean-focused objectives and provides a principled mechanism for suppressing rare but hazardous outcomes through tail-sensitive optimization.

3.2 Problem Formulation as a Clearance Tail Risk

Tail risk concerns rare but high-consequence outcomes that appear in the upper tail of a risk distribution. Let Z be a random variable representing risk values and let $\alpha \in (0, 1)$ denote a tail level. The Value-at-Risk (VaR) at level α is defined as the α-quantile of Z:

$$q_\alpha(Z) = \min \left\{ q \in \mathbb{R} \;\middle|\; \mathbb{P}(Z \leq q) \geq \alpha \right\}. \tag{3.1}$$

VaR provides a threshold that separates typical outcomes from tail events, but it does not characterize how severe the outcomes are beyond the threshold. Conditional Value-at-Risk (CVaR) complements VaR by measuring the expected risk conditioned on being in the tail:

$$\text{CVaR}_\alpha(Z) = \mathbb{E}[Z \mid Z \geq q_\alpha(Z)]. \tag{3.2}$$

As a result, CVaR is sensitive to both the frequency and the magnitude of tail events, making it suitable for penalizing rare but hazardous behaviors.

In motion planning, risk samples are naturally obtained along the prediction horizon. For each predicted mode $k \in \{1, \ldots, M\}$ over a horizon of length T, define the per-frame risk sequence

$$z^{(k)} = \left(z_1^{(k)}, \ldots, z_T^{(k)}\right), \tag{3.3}$$

and let Z be the random variable whose empirical distribution is induced by these T samples. Then $q_\alpha\left(z^{(k)}\right)$ corresponds to the empirical VaR of the sequence and $\mathrm{CVaR}_\alpha\left(z^{(k)}\right)$ summarizes the expected risk in the α-tail.

A clearance-based construction is used to instantiate $z_t^{(k)}$ at each time step. For mode k and step $t \in \{1, \ldots, T\}$, let $p_{\mathrm{ego}}^{(k)}(t) \in \mathbb{R}^2$ denote the ego position and let $p_{\mathrm{obs},i}(t)$ denote the position of obstacle i, where $i \in \{1, \ldots, N_{\mathrm{obs}}\}$. Using an effective radius

$$R_{\mathrm{eff}} = R_{\mathrm{ego}} + R_{\mathrm{obs}}, \tag{3.4}$$

the clearance to obstacle i is defined as

$$\mathrm{clr}_i^{(k)}(t) = \left\| p_{\mathrm{ego}}^{(k)}(t) - p_{\mathrm{obs},i}(t) \right\|_2 - R_{\mathrm{eff}}. \tag{3.5}$$

A soft minimum over obstacles provides a smooth approximation to the minimum

clearance at time t:

$$\mathrm{clr}_t^{(k)} = -\frac{1}{\beta_{\mathrm{clr}}} \log \sum_{i=1}^{N_{\mathrm{obs}}} \exp\Big(-\beta_{\mathrm{clr}}\, \mathrm{clr}_i^{(k)}(t)\Big), \qquad \beta_{\mathrm{clr}} > 0, \tag{3.6}$$

where β_{clr} controls the sharpness. Given a safety margin $m_{\mathrm{clr}} > 0$, the per-frame clearance risk is defined by

$$z_t^{(k)} = \mathrm{softplus}\Big(m_{\mathrm{clr}} - \mathrm{clr}_t^{(k)}\Big) \geq 0. \tag{3.7}$$

This construction yields near-zero penalties when the clearance is comfortable, while producing large penalties near collision configurations.

To obtain a time-wise tail-risk scalar for each mode, a CVaR-style mean excess term is computed from the empirical samples. Let $q_\alpha\big(z^{(k)}\big)$ denote the empirical α-quantile of the sequence $z^{(k)}$. The mean excess loss aggregates how strongly and how often per-frame risks exceed this VaR threshold:

$$r_k = \frac{1}{(1-\alpha)T} \sum_{t=1}^{T} \max\Big(z_t^{(k)} - q_\alpha\big(z^{(k)}\big)\,,\, 0\Big). \tag{3.8}$$

Modes whose risks remain below the VaR level satisfy $r_k \approx 0$, whereas larger r_k indicates that the trajectory spends more time in the high-risk tail region. For continuous distributions, one has

$$\mathrm{CVaR}_\alpha\big(z^{(k)}\big) = q_\alpha\big(z^{(k)}\big) + r_k, \tag{3.9}$$

so r_k can be interpreted as a tail severity term beyond the VaR threshold.

3.3 Risk-aware Mode Scoring

Given multimodal decoding, the planner outputs mode logits $\pi \in \mathbb{R}^M$ and the corresponding categorical distribution

$$p = \mathrm{softmax}(\pi) \in \Delta^{M-1}, \qquad p_k = \frac{\exp(\pi_k)}{\sum_{j=1}^{M} \exp(\pi_j)}. \tag{3.10}$$

In standard multimodal imitation learning, mode selection and supervision are largely driven by likelihood or regression errors, which does not explicitly suppress modes that contain rare but high-risk segments. Risk-aware mode scoring incorporates the clearance tail-risk scalar r_k from Eq. (3.8) into mode scoring.

An imitation-based score for mode k is defined as

$$s_k^{\mathrm{im}} = -\log p_k, \tag{3.11}$$

and the risk-aware score augments it with the clearance tail risk:

$$s_k = s_k^{\mathrm{im}} + \lambda_{\mathrm{cvar}}\, r_k, \qquad \lambda_{\mathrm{cvar}} \geq 0. \tag{3.12}$$

The coefficient λ_{cvar} controls the trade-off between imitation likelihood and tail-risk penalization. The label mode index used for subsequent regression and classification is selected by

$$m^* = \arg \min_{k \in \{1,\ldots,M\}} s_k. \tag{3.13}$$

This selection favors modes that are both likely under the decoder distribution and safe along the horizon in the sense of reduced tail risk.

Risk awareness is further used to shape the probability head via a tail-risk-aware soft target distribution that downweights high-risk modes even when their imitation likelihood is high. Let $\tilde{r}_k$ denote a normalized version of r_k computed within a minibatch (zero mean and unit variance) and clipped for stability. A risk-aware soft target $q \in \Delta^{M-1}$ is constructed by reweighting p:

$$q_k = \frac{p_k \exp(-\lambda_{\text{cvar}} \tilde{r}_k)}{\sum_{j=1}^{M} p_j \exp(-\lambda_{\text{cvar}} \tilde{r}_j)}. \tag{3.14}$$

The probability head is then trained by minimizing the KL divergence between the risk-aware targets q and the model distribution p:

$$\mathcal{L}_{\text{KL}} = \text{KL}(q \,\|\, p) = \sum_{k=1}^{M} q_k \log \frac{q_k}{p_k}. \tag{3.15}$$

This objective encourages the model to allocate less probability mass to modes that exhibit large clearance tail risk, thereby reducing the probability of selecting rare but potentially hazardous trajectories.

Together, Eq. (3.12)–(3.13) provide a risk-aware mode scoring rule for mode selection, while Eq. (3.14)–(3.15) provide a learning signal that shifts the multimodal distribution away from high-risk modes. This complements mean-focused imitation objectives by explicitly incorporating tail-risk sensitivity into multimodal decision making.

Chapter 4

Planner Architecture

4.1 Overall Planner Pipeline

Figure 4.1: Overview of proposed planner. Driving logs are converted into agent, map, and ego features, processed by a Transformer with ALM constrained ego attention and a CVaR based risk head, and the resulting plans are evaluated on nuPlan under OLS, NR-CLS, and R-CLS.

Figure 4.1 illustrates the overall pipeline of the proposed planner. Logged expert demonstrations are converted into structured scene features for the ego vehicle, surrounding agents, and map context. These heterogeneous inputs are encoded as token sequences and processed by a Transformer backbone to produce a multimodal distribution over future ego trajectories. During training, the planner is regularized by Mean-Deviation Constrained Attention (MDCA), which prevents attention collapse in ego-state encoding via an ALM-based constraint (Chapter 2). In addition, a CVaR-based tail-risk module evaluates clearance-derived risks along each candidate trajectory,

enabling risk-aware mode shaping and safer mode selection (Chapter 3). The resulting model is evaluated on the nuPlan benchmark under open-loop and closed-loop protocols.

4.1.1 Inputs and Tokenization

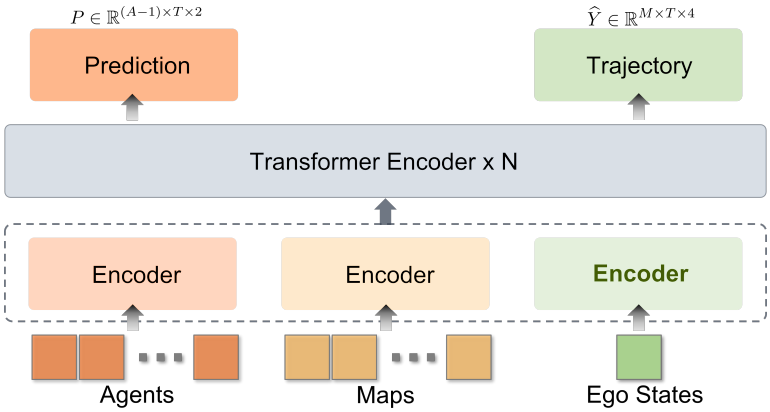


Figure 4.2: Planner architecture of Base (vanilla), PlanTF(with SDE), Ours: Car Planner (with constrained attention weights).

The planner operates on three heterogeneous inputs extracted from driving logs: surrounding-agent motion histories, the current ego state, and a local map snippet represented by lane polygons. Let A denote the number of agents in the local scene including ego, and let H be the history length. For each non-ego agent $i \in \{1, \ldots, A-1\}$, a history sequence $X_i^{\text{hist}} \in \mathbb{R}^{H\times d_{\text{kin}}}$ is formed from past kinematic states (e.g., position, heading, velocity). A dedicated agent encoder maps each history into a d-dimensional token $x_i^{\text{agt}} = f_{\text{agt}}(X_i^{\text{hist}}) \in \mathbb{R}^d$. In parallel, the local map is represented as a set of lane polygons. For each polygon $j \in \{1, \ldots, N_{\text{map}}\}$ with geometry and attributes, a polygon encoder produces a map token $x_j^{\text{map}} = f_{\text{map}}(\mathcal{P}_j) \in \mathbb{R}^d$.

The ego input is a six-dimensional state vector $s_{\text{ego}} = [x, y, \text{yaw}, v, a, s]^\top \in \mathbb{R}^C$,

where $C = 6$ and the channels correspond to position, heading, longitudinal speed, longitudinal acceleration, and steering angle. Unlike agent histories and map polygons, the ego state is not treated as a sequence. Each scalar channel is first embedded into a d-dimensional representation, and a single learnable query aggregates the channel embeddings through channel-wise attention, producing one ego token $x^{\text{ego}} \in \mathbb{R}^d$ and an associated attention distribution over channels. This design enables the planner to learn how strongly each ego-state channel contributes to the ego representation, which is later regularized by MDCA.

To preserve global geometry, an explicit positional encoding is added to agent and map tokens. For each agent token, the last observed pose is encoded as $[x, y, \cos\theta, \sin\theta]$, and for each map polygon the center pose is encoded in the same form. A small MLP projects this 4D pose into $\mathbb{R}^d$, and the resulting embedding is additively fused with the corresponding token feature. Finally, the ego token, agent tokens, and map tokens are concatenated into a single token sequence

$$X_0 = \left[x^{\text{ego}};\ x_1^{\text{agt}}, \ldots, x_{A-1}^{\text{agt}};\ x_1^{\text{map}}, \ldots, x_{N_{\text{map}}}^{\text{map}}\right] \in \mathbb{R}^{N \times d}, \tag{4.1}$$

where $N = 1 + (A-1) + N_{\text{map}}$. Key-padding masks are used to ignore invalid or padded tokens when the number of observed agents or map elements varies across scenes.

4.1.2 Transformer Backbone and Decoding

The concatenated token sequence is processed by a shared Transformer encoder with L layers and h attention heads per layer. Denoting the encoder by $f_{\text{enc}}(\cdot)$, the

encoded sequence is

$$X_L = f_{\text{enc}}(X_0) \in \mathbb{R}^{N \times d}. \tag{4.2}$$

Self-attention in the joint encoder models interactions between the ego vehicle, surrounding agents, and road geometry, allowing the planner to select and combine relevant context dynamically rather than relying on hand-crafted heuristics.

A multimodal trajectory decoder reads out from the encoded ego token (the first element of X_L) and predicts M candidate future trajectories over a horizon of length T. Each mode $k \in \{1, \ldots, M\}$ outputs a sequence

$$\hat{Y}^{(k)} \in \mathbb{R}^{T \times 4}, \tag{4.3}$$

where each step contains $(x, y, \cos\psi, \sin\psi)$ for continuity of heading representation. In addition, the decoder produces mode logits $\pi \in \mathbb{R}^M$, which define a categorical distribution over modes after softmax. At inference time, a single mode can be selected according to the learned scoring rule, and headings are recovered via $\text{atan2}(\sin\psi, \cos\psi)$.

To provide dense scene supervision and encourage interaction-aware representations, an auxiliary prediction head in parallel outputs future positions for non-ego agents:

$$\hat{P} \in \mathbb{R}^{(A-1) \times T \times 2}, \tag{4.4}$$

trained to match the corresponding logged future positions. This auxiliary objective stabilizes training and improves consistency between the planned ego motion and the predicted evolution of surrounding agents.

4.1.3 Mean-Deviation Constrained Attention (MDCA) for Ego-State Encoding

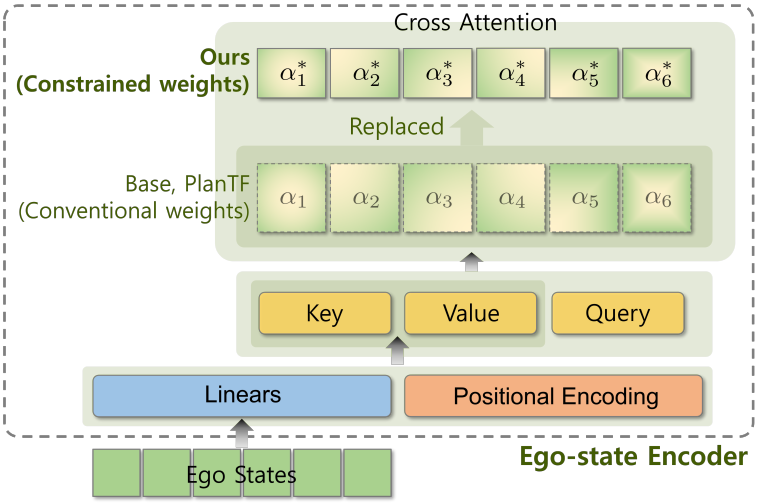


Figure 4.3: Illustration of MDCA applied to ego-state attention. The attention weights over ego-state channels are constrained to prevent collapse onto a small subset of channels, promoting balanced state usage.

The ego-state encoder constructs a compact ego token by aggregating C ego-state channels through channel-wise attention. Each scalar channel is first embedded into a d-dimensional representation and then combined by a query–key–value attention module, producing both an ego token and a channel attention distribution $a_\theta^{(n)} \in \Delta^{C-1}$ for each sample n. While this mechanism is effective for learning which state channels matter for planning, it can become brittle when the learned attention collapses onto a small subset of channels. Such attention collapse amplifies shortcut learning, where the planner over-relies on a few ego-state cues and ignores complementary information required for robust behavior under distribution shifts.

MDCA addresses this issue by explicitly constraining the channel-wise attention weights in the ego-state encoder. As illustrated in Figure 4.3, MDCA regularizes con-

ventional attention weights so that they do not become excessively peaked. The constraint is formulated using the mean-deviation statistic defined in Chapter 2, which measures the deviation of $a_\theta^{(n)}$ from the uniform allocation across channels. Rather than enforcing uniform attention, MDCA restricts the deviation to remain within a margin, thereby preserving the ability to emphasize informative channels while discouraging pathological concentration.

The constraint is enforced during training via an Augmented Lagrangian Method (ALM). The ALM penalty is added to the training objective only when the constraint is violated, and the corresponding Lagrange multiplier is updated online after each optimizer step. In the nuPlan experiments, the penalty parameter ρ is fixed, so the effective constraint strength adapts through the multiplier based on the observed violations. Importantly, MDCA only regulates the ego-state channel attention within the ego-state encoder; the Transformer backbone and decoding remain unchanged, and no additional computation is introduced at inference time.

4.1.4 CVaR based risk module for Mode Scoring

The planner incorporates a CVaR-based risk module to discourage rare but potentially dangerous trajectory modes in multimodal planning. Given the predicted ego trajectories $\{\hat{Y}^{(k)}\}_{k=1}^{M}$ and the predicted motions of surrounding agents, the module assigns a mode-wise tail-risk scalar r_k that summarizes the risk associated with insufficient clearance along the horizon. Figure 4.4 highlights that the risk computation is performed per mode and subsequently used for mode selection and distribution shap-

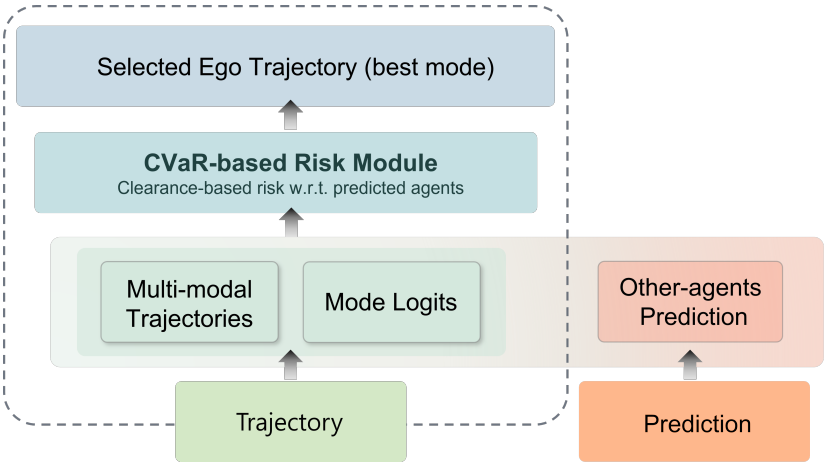


Figure 4.4: Illustration of the CVaR-based risk module. For each predicted trajectory mode, per-frame clearance risks are computed and aggregated into a tail-risk scalar using a CVaR-style mean excess formulation. This risk score informs risk-aware mode selection and distribution shaping.

ing.

For each mode k, the module computes a per-frame clearance sequence by measuring the distance margin between the predicted ego position and surrounding obstacles (or predicted agents), using a differentiable soft-minimum aggregation over obstacles at each time step (Figure 4.5). The resulting clearance is converted into a nonnegative per-frame risk sequence through a margin-based mapping, and a CVaR-style mean excess formulation aggregates these risks into the scalar r_k at tail level α, as defined in Chapter 3. This tail-risk design is sensitive to high-risk segments along the horizon, rather than average clearance, providing a principled signal for suppressing long-tail hazardous behaviors.

The tail-risk scalar r_k is used in two complementary ways. First, it is combined with the imitation-based mode likelihood to form a risk-aware score, so that the selected mode prefers trajectories that are both likely under the learned distribution and safe

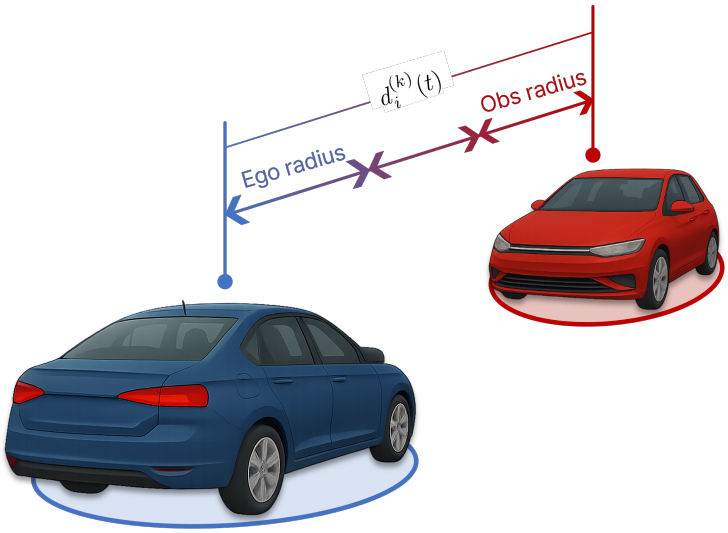


Figure 4.5: Clearance-based per-frame risk computation. For each predicted ego trajectory mode, the minimum clearance to surrounding obstacles is computed at each time step using a differentiable soft-minimum. The clearance is then transformed into a nonnegative risk value based on a safety margin.

along the horizon. Second, the same tail-risk values define a risk-aware soft target distribution that downweights risky modes and shapes the multimodal probability head through a KL divergence objective. This dual use allows the module to act not only as an inference-time safety filter, but also as a training-time signal that gradually shifts probability mass away from high-risk modes, improving safety-related behavior without altering the underlying Transformer architecture.

4.2 Training Objectives

The planner is trained under pure imitation learning with a multimodal trajectory output. For each training sample, the model predicts M candidate ego futures $\{\hat{Y}^{(k)}\}_{k=1}^{M}$ over a horizon of length T, mode logits $\pi \in \mathbb{R}^M$, and auxiliary future predictions for non-ego agents. The training objective combines imitation fidelity with two

additional mechanisms for robustness and safety: (i) an Augmented Lagrangian penalty that enforces the MDCA constraint on ego-state attention, and (ii) a tail-risk-aware shaping loss that discourages rare but potentially dangerous modes in the multimodal distribution. Accordingly, the overall objective consists of a base imitation loss for ego and non-ego trajectories, a KL divergence term derived from tail risk, and the MDCA Augmented Lagrangian term.

Let the expert ego trajectory be $Y \in \mathbb{R}^{T\times 4}$ represented by $(x, y, \cos\psi, \sin\psi)$, and let the k-th predicted mode be $\hat{Y}^{(k)} \in \mathbb{R}^{T\times 4}$. Using $(\cos\psi, \sin\psi)$ avoids discontinuities in angle regression around $\pm\pi$. For non-ego agents, let $P_{\text{gt}}, P \in \mathbb{R}^{(A-1)\times T\times 2}$ denote the ground-truth and predicted future positions, respectively. Let $\pi \in \mathbb{R}^M$ be the mode logits and $p = \text{softmax}(\pi) \in \Delta^{M-1}$ be the corresponding mode probabilities with components p_k.

The smooth ℓ_1 (Huber) loss with threshold $\delta = 1$ is applied element-wise over a set of valid indices $\mathcal{M}$:

$$L_{\text{smooth }1}(A, B) = \frac{1}{|\mathcal{M}|} \sum_{i\in\mathcal{M}} \begin{cases} \frac{1}{2}(A_i - B_i)^2, & |A_i - B_i| < \delta, \\ |A_i - B_i| - \frac{\delta}{2}, & \text{otherwise.} \end{cases} \tag{4.5}$$

A single supervised mode index m^* is selected for each sample (defined below). The base imitation objective supervises the selected ego mode, aligns the probability head with the selected mode, and trains auxiliary predictions of surrounding agents:

$$L_{\text{task}}(\theta) = L_{\text{smooth }1}\left(\hat{Y}^{(m^*)}, Y\right) + \text{CE}(p, m^*) + L_{\text{smooth }1}(P, P_{\text{gt}}), \tag{4.6}$$

where $\mathrm{CE}(p, m^*)$ denotes the cross-entropy between p and the one-hot label for m^*.

The ego-state attention constraint from Chapter 2 is enforced by an Augmented Lagrangian term. Using $[z]_+ := \max(0, z)$, the penalty is

$$L_{\mathrm{alm}}(\theta, \lambda) = \lambda[g(\theta)]_+ + \frac{\rho}{2}\big([g(\theta)]_+\big)^2, \tag{4.7}$$

with $\lambda \geq 0$ initialized at zero and $\rho > 0$ treated as a fixed hyperparameter in the nuPlan experiments. After each optimizer step, the multiplier is updated by

$$\lambda \leftarrow \lambda + \rho[g(\theta)]_+. \tag{4.8}$$

This update strengthens the constraint only when violations occur, preventing persistent attention collapse while preserving flexibility when the constraint is already satisfied.

For tail-risk-aware learning, a clearance-based per-mode tail-risk scalar r_k is computed along the horizon using the CVaR-style mean excess formulation from Chapter 3. Let $p_{\mathrm{ego}}^{(k)}(t) \in \mathbb{R}^2$ be the ego position of mode k at time step $t \in \{1, \ldots, T\}$, and let $p_{\mathrm{obs},i}(t) \in \mathbb{R}^2$ denote the position of obstacle (or predicted agent) $i \in \{1, \ldots, N_{\mathrm{obs}}\}$. Using an effective radius

$$R_{\mathrm{eff}} = R_{\mathrm{ego}} + R_{\mathrm{obs}}, \tag{4.9}$$

the clearance to obstacle i is defined as

$$\mathrm{clr}_i^{(k)}(t) = \big\|p_{\mathrm{ego}}^{(k)}(t) - p_{\mathrm{obs},i}(t)\big\|_2 - R_{\mathrm{eff}}. \tag{4.10}$$

A differentiable soft minimum over obstacles provides a smooth approximation to the minimum clearance at time t:

$$\mathrm{clr}_t^{(k)} = -\frac{1}{\beta_{\mathrm{clr}}} \log \sum_{i=1}^{N_{\mathrm{obs}}} \exp\Big(-\beta_{\mathrm{clr}}\, \mathrm{clr}_i^{(k)}(t)\Big), \qquad \beta_{\mathrm{clr}} > 0. \tag{4.11}$$

Given a safety margin $m_{\mathrm{clr}} > 0$, the per-frame clearance risk is defined by

$$z_t^{(k)} = \mathrm{softplus}\Big(m_{\mathrm{clr}} - \mathrm{clr}_t^{(k)}\Big) \geq 0, \tag{4.12}$$

and the corresponding risk sequence is $z^{(k)} = \big(z_1^{(k)}, \ldots, z_T^{(k)}\big)$. The Value-at-Risk (VaR) at tail level $\alpha \in (0, 1)$ is the empirical α-quantile

$$q_\alpha\big(z^{(k)}\big) = \min\Big\{q \in \mathbb{R} \;\Big|\; \mathbb{P}(z_t^{(k)} \leq q) \geq \alpha\Big\}, \tag{4.13}$$

where $\mathbb{P}$ denotes the empirical cumulative distribution function (CDF) over the T samples $z^{(k)}tt = 1^T$. The tail-risk scalar is then computed by the mean excess term beyond VaR:

$$r_k = \frac{1}{(1-\alpha)T} \sum_{t=1}^{T} \max\Big(z_t^{(k)} - q_\alpha\big(z^{(k)}\big)\,,\, 0\Big). \tag{4.14}$$

The tail risk is incorporated into mode selection during training. An imitation-based score for mode k is

$$s_k^{\mathrm{im}} = -\log p_k, \tag{4.15}$$

and the risk-aware score is

$$s_k = s_k^{\text{im}} + \lambda_{\text{cvar}} r_k, \tag{4.16}$$

where $\lambda_{\text{cvar}} \geq 0$ trades off imitation likelihood and tail risk. The supervised mode index is chosen by

$$m^* = \arg\min_k s_k. \tag{4.17}$$

This selection biases regression supervision toward modes that remain likely under the model while avoiding modes with elevated tail risk. Unlike best-of-M selection based solely on imitation error, the model probability and tail-risk score are used to directly encourage a safety-aware multimodal distribution.

Beyond selecting m^*, the multimodal probability head is shaped by a soft target that downweights risky modes. Let $\tilde{r}_k$ be a per-batch normalized version of r_k with zero mean and unit variance, optionally clipped for stability. A tail-risk-aware target distribution $q \in \Delta^{M-1}$ is defined as

$$q_k = \frac{p_k \exp(-\lambda_{\text{cvar}} \tilde{r}_k)}{\sum_{j=1}^{M} p_j \exp(-\lambda_{\text{cvar}} \tilde{r}_j)}. \tag{4.18}$$

The probability head is trained with a KL divergence

$$L_{\text{KL}} = \text{KL}(q \| p) = \sum_{k=1}^{M} q_k \log \frac{q_k}{p_k}. \tag{4.19}$$

This encourages the model to shift probability mass away from tail-risky modes while preserving the relative preference among safe candidates.

The final objective combines the imitation loss, the risk-aware KL shaping, and the MDCA Augmented Lagrangian penalty:

$$L(\theta) = L_{\text{task}}(\theta) + \lambda_{\text{KL}} L_{\text{KL}} + L_{\text{alm}}(\theta, \lambda), \tag{4.20}$$

where $\lambda_{\text{KL}} \geq 0$ controls the strength of tail-risk-aware distribution shaping.

Chapter 5

Experiments

5.1 Experimental Setup

To evaluate the performance of the proposed motion planning framework, experiments are conducted using nuPlan, the world's first large-scale machine learning-based planning benchmark provided by Motional [33, 34]. Unlike previous datasets focused on perception, nuPlan is specifically designed for the development and validation of autonomous driving planners in complex environments.

The dataset comprises approximately 1,300 hours of real-world driving data collected across four distinct cities: Boston, Pittsburgh, Las Vegas, and Singapore. These locations provide a diverse array of urban challenges, ranging from the narrow, congested streets of Boston to the left-hand traffic environment of Singapore. The inclusion of multiple cities supports robust generalization across different traffic laws and driving conditions. The nuPlan v1.1 dataset is utilized, which includes over 15,000 logs and is categorized into approximately 75 diverse driving scenarios (e.g., unprotected left turns, lane changes, and interaction with pedestrians).

All experiments are conducted on the official `test14-random` and `test14-hard` splits selected by [48]. The `test14-random` split consists of randomly sampled scenarios, while `test14-hard` is a curated set of challenging scenarios designed to stress-test

robustness under difficult interactions and distribution shift, each comprising about 280 scenarios across 14 scenario types. For comprehensive evaluation, both open-loop evaluation on logged expert trajectories and closed-loop simulation under interactive traffic are considered.

Three standard nuPlan metrics are reported: Open-Loop Score (OLS), Non-Reactive Closed-Loop Score (NR-CLS), and Reactive Closed-Loop Score (R-CLS), where higher is better. OLS measures imitation quality against logged expert trajectories in open-loop. NR-CLS evaluates closed-loop rollouts with non-reactive log-replay agents, while R-CLS evaluates rollouts with reactive agents, exposing safety-critical failures under interaction. Particular emphasis is placed on reactive closed-loop performance on `test14-hard` as a primary indicator of robustness. In addition, the standardized evaluation protocol reflects safety (e.g., collisions), comfort (e.g., jerk and lateral acceleration), and progress towards the goal.

All planner variants are trained for 20 epochs on NVIDIA RTX 4090 GPU with a total batch size of 32 using the Adam optimizer (learning rate 10^{-3}, weight decay 10^{-4}). To mitigate compounding errors in closed-loop rollouts, a state-perturbation augmentation is applied: bounded noise is added to the current ego state, and all coordinates are re-normalized with respect to the perturbed ego frame.

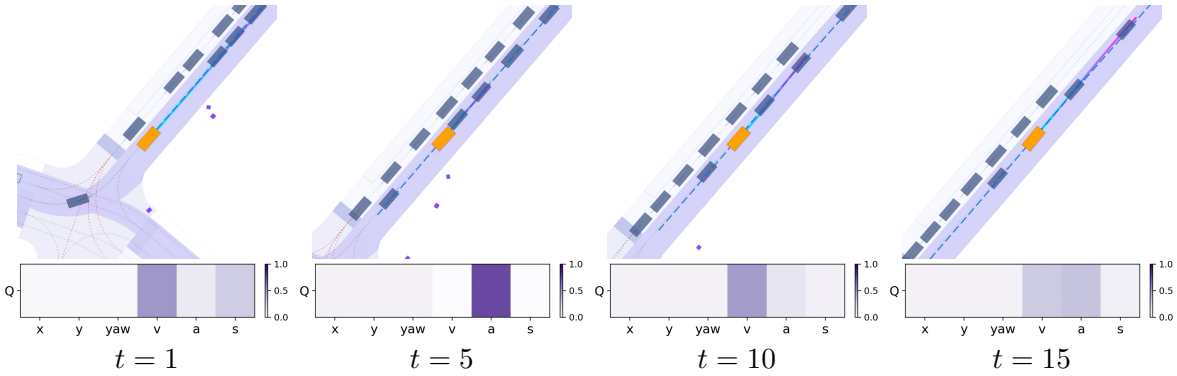
The following model variants are compared under a shared encoder–decoder backbone to ensure fair comparison. **Base** encodes the 6D ego state using a small MLP without ego-state attention regularization. **PlanTF (with SDE)** replaces the ego MLP with a single-query ego-state attention encoder and applies State Dropout En-

coder (SDE) during training. **CAR Planner** further regularizes the single-query ego-state attention with an augmented Lagrangian dispersion constraint (MDCA) on the ego-state attention weights. **Base+Risk** augments the Base model with a CVaR-based risk module that computes clearance-based tail risk per mode and uses it for risk-aware mode selection and soft targets, without MDCA. Finally, **CARE Planner** combines both components by extending CAR Planner with the CVaR-based risk module for tail-risk-aware supervision over multimodal outputs.

5.2 Experimental Results

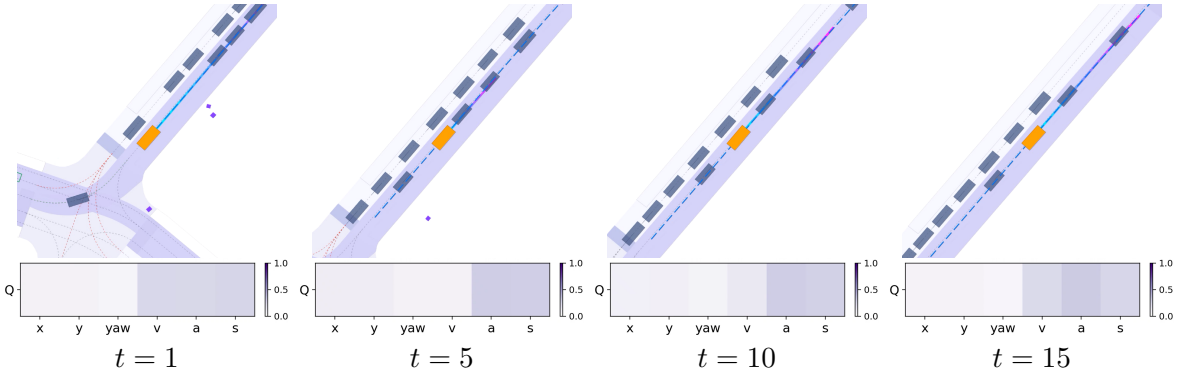
Figure 5.1 provides a qualitative comparison between PlanTF and the proposed planner using paired visualization of trajectory outputs and the corresponding ego-state attention. For each method, four snapshots at matched time steps ($t = \{1, 5, 10, 15\}$) are shown. In each snapshot, the top panel visualizes the predicted trajectory in the local map context, while the bottom panel shows the attention weights over ego-state channels. This paired layout is intended to interpret how ego-state utilization evolves as the scenario progresses, rather than treating attention maps as separate figures.

Across the time steps, PlanTF exhibits more concentrated attention patterns, where a small subset of ego-state channels tends to dominate for extended periods. In contrast, the proposed planner shows a more distributed and stable allocation of attention across channels, indicating reduced reliance on a single dominant cue. This observation motivates examining whether the attention constraint is actively enforced during training.



$t = 1$ $t = 5$ $t = 10$ $t = 15$

(a) PlanTF (top: trajectory, bottom: attention)

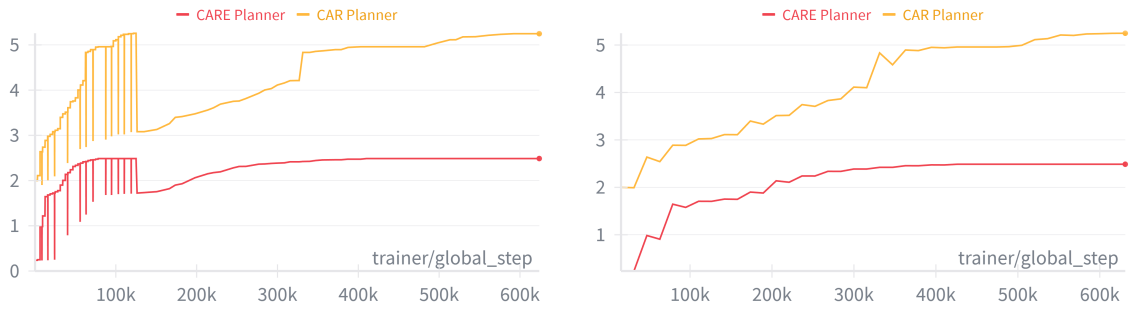


$t = 1$ $t = 5$ $t = 10$ $t = 15$

(b) CARE (top: trajectory, bottom: attention)

Figure 5.1: Qualitative comparison with paired visualization. For each time step, the trajectory visualization (top) and the corresponding attention map (bottom) are shown as a single unit.

The training dynamics of the Augmented Lagrangian optimization are illustrated in Figure 5.2, which plots the Lagrangian multiplier λ over training. Both curves share the global-step x-axis. The step-wise curve reports batch-level logs, whereas the epoch-aggregated curve summarizes the batch logs within each epoch. In the implementation, λ is updated at the end of each training batch after the optimizer step. Logging is performed before this update. As a result, the epoch-aggregated curve reflects averages of pre-update batch values and can appear lower in early training when λ increases rapidly within an epoch. The gradual increase and subsequent stabilization of λ indicate that constraint enforcement is adaptively strengthened when violations persist and later



(a) λ (step-wise) (b) λ (epoch-aggregated)

Figure 5.2: Lagrangian multiplier λ under ALM. Both curves share the global-step x-axis; the left shows batch-wise updates and the right shows the epoch-aggregated values.

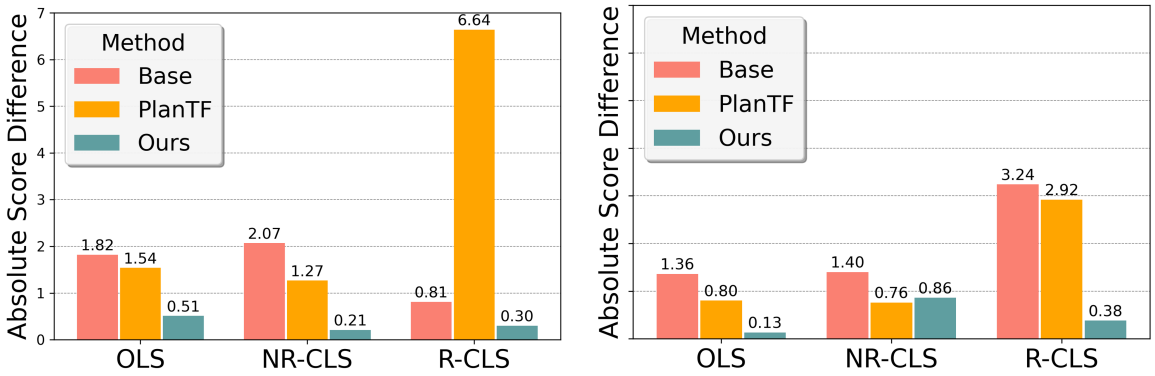
settles as the constraint becomes better satisfied. Notably, CARE exhibits consistently lower λ values than CAR, suggesting that the risk-aware distribution shaping objective can guide learning toward safer behaviors that require less corrective pressure to satisfy the attention constraint.

To verify whether MDCA mitigates shortcut learning, Table 5.1 reports ego-state ablation results comparing state6 and state5, and Figure 5.3 visualizes the corresponding absolute score gaps. Overall, CAR exhibits smaller absolute gaps between `state6` and `state5`, and the gap reduction is more pronounced on the hazard-focused `test14-hard` split. Overall, the reduced sensitivity to ego-state removal supports that the proposed regularization promotes more robust state utilization, consistent with the qualitative attention patterns in Figure 5.1.

Table 5.2 reports the overall scores on both nuPlan splits. On `test14-random`, CARE Planner achieves the best performance across all three metrics, reaching 89.13 (OLS), 87.48 (NR-CLS), and 82.52 (R-CLS). Compared to the Base model, this corresponds to improvements of +2.49 in OLS, +7.47 in NR-CLS, and +8.04 in R-CLS.

Table 5.1: Ego-state ablation results (state5 vs. state6).

Setting	Method	test14-random			test14-hard		
		OLS	NR-CLS	R-CLS	OLS	NR-CLS	R-CLS
state5	Base	84.82	77.94	73.67	81.12	63.90	49.87
	PlanTF	87.81	83.96	72.72	84.14	69.27	56.91
	CAR	87.16	85.12	78.01	86.18	68.62	64.26
state6	Base	86.64	80.01	74.48	82.48	65.30	53.11
	PlanTF	86.27	85.23	79.36	83.34	70.03	59.83
	CAR	87.67	84.91	78.31	86.31	69.48	64.64



(a) test14-random (b) test14-hard

Figure 5.3: Absolute score differences between state6 and state5. Larger gaps indicate higher sensitivity to removing an ego-state channel, reflecting stronger shortcut reliance.

Compared to PlanTF, CARE Planner improves OLS by +2.86, NR-CLS by +2.25, and R-CLS by +3.16, indicating consistent gains on the common scenario distribution.

More significant improvements are observed on `test14-hard`, which contains failure-prone and distribution-shifted conditions. CARE Planner achieves 87.76 (OLS), 72.65 (NR-CLS), and 67.17 (R-CLS), outperforming CAR Planner by +1.45 in OLS, +3.17 in NR-CLS, and +2.53 in R-CLS. The comparison between **Base+Risk** and **CARE Planner** further highlights the necessity of MDCA. Although Base+Risk improves over Base, it remains substantially below CARE Planner on `test14-hard` (R-CLS

Table 5.2: Overall score on nuPlan splits.

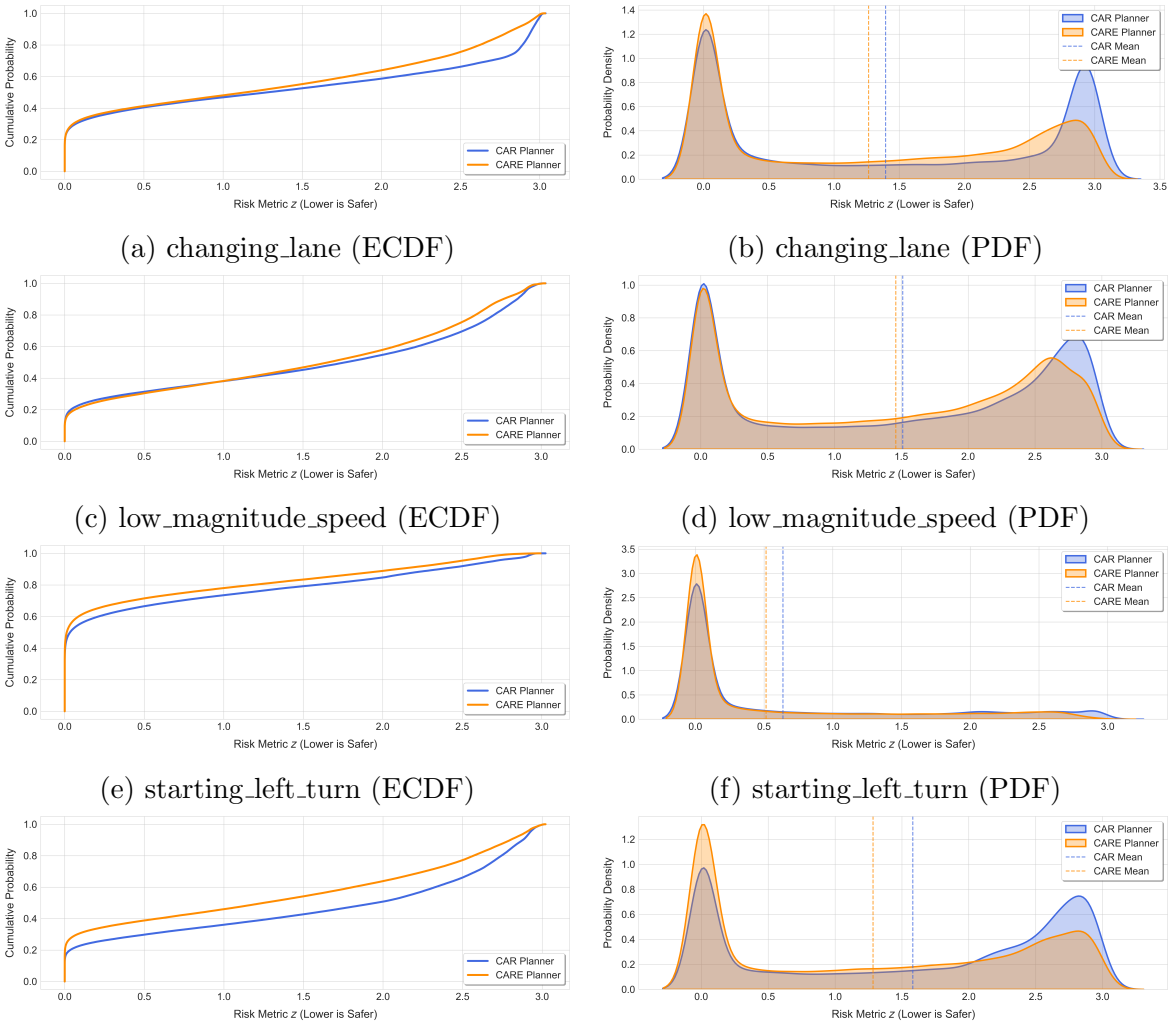
Method	test14-random			test14-hard		
	OLS	NR-CLS	R-CLS	OLS	NR-CLS	R-CLS
Base	86.64	80.01	74.48	82.48	65.30	53.11
PlanTF	86.27	85.23	79.36	83.34	70.03	59.83
CAR Planner	87.67	84.91	78.31	86.31	69.48	64.64
Base + Risk	88.19	86.05	80.36	83.92	69.65	59.73
CARE Planner	**89.13**	**87.48**	**82.52**	**87.76**	**72.65**	**67.17**

Table 5.3: Sub metrics on nuPlan test14-hard split under R-CLS.

Method	test14-hard (R-CLS)					
	Collisions	TTC	Drivable	Comfort	Progress	Speed
Base	88.11	81.50	92.64	88.23	72.79	98.02
planTF	85.84	80.88	92.64	93.01	84.55	97.01
CAR Planner	90.63	85.49	94.02	98.16	84.28	98.22
Base + Risk	89.52	83.82	92.27	90.80	81.61	97.17
CARE Planner	**92.03**	**87.55**	**95.22**	**99.26**	**84.55**	**98.67**

59.73 vs. 67.17). Since both variants incorporate risk information, the additional gain is attributed to MDCA, which encourages balanced utilization of ego-state channels and prevents attention collapse to a narrow shortcut feature. This property becomes critical in hazardous scenarios where robust state usage is required for reliable decision-making.

Table 5.3 presents sub-metrics on `test14-hard` under R-CLS. CARE Planner achieves the highest scores in Collisions (92.03), TTC (87.55), Drivable (95.22), and Comfort (99.26), while maintaining competitive Progress (84.55) and Speed (98.67). Relative to CAR Planner, the improvements are concentrated in safety-relevant factors (Collisions +1.40, TTC +2.06, Drivable +1.20, Comfort +1.10), indicating that the overall R-CLS gain is driven by safer interaction and better constraint satisfaction rather than



(a) changing_lane (ECDF)

(b) changing_lane (PDF)

(c) low_magnitude_speed (ECDF)

(d) low_magnitude_speed (PDF)

(e) starting_left_turn (ECDF)

(f) starting_left_turn (PDF)

(g) waiting_for_pedestrian_to_cross (ECDF)

(h) waiting_for_pedestrian_to_cross (PDF)

Figure 5.4: Risk distributions (ECDF and PDF) of the per-frame clearance risk z across representative scenario types. Lower z indicates safer behavior.

aggressive speed seeking.

The role of MDCA in mitigating shortcut learning is illustrated by the ego-state ablation analysis in Table 5.1 and Figure 5.3. Table 5.1 reports absolute scores when using `state6` versus `state5`, and Figure 5.3 visualizes the corresponding absolute score differences. Without explicit attention regularization, reducing the ego-state dimension can lead to large performance drops, implying dependence on specific channels.

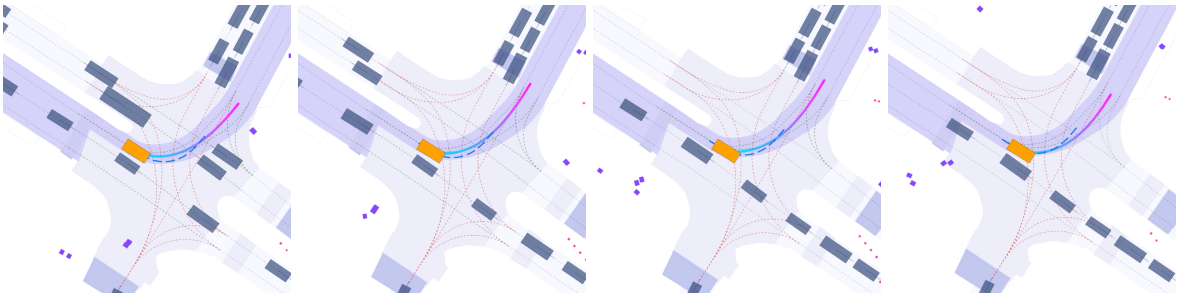
In contrast, CAR Planner exhibits consistently smaller gaps, particularly in R-CLS on `test14-hard` (0.38), compared to Base (3.24) and PlanTF (2.92). This reduced sensitivity indicates that MDCA discourages attention collapse and promotes distributed usage of ego-state inputs, thereby improving robustness under distribution shift. Combining this property with tail-risk-aware supervision explains the strong closed-loop performance of CARE Planner on `test14-hard`.

Figure 5.4 visualizes how tail-risk-aware learning reshapes the distribution of the risk metric z across representative scenario types. The top row of each scenario shows the empirical cumulative distribution function (ECDF), where the value at threshold z indicates the fraction of episodes with risk $\leq z$. The bottom row shows the probability density function (PDF), indicating how frequently each risk level occurs and how heavy the high-risk tail is. Across scenarios, CARE Planner shifts the ECDF upward and leftward compared to CAR Planner, implying that a larger fraction of episodes achieves lower risk. The PDF plots show reduced probability mass in the high-risk region and a slight shift of the distribution center toward smaller z, supporting that the risk module suppresses rare but severe high-risk episodes. Together with MDCA, this distributional effect is consistent with the improved safety-oriented closed-loop metrics and the strong performance on `test14-hard`.

Figure 5.5 presents qualitative comparisons at matched time steps ($t = \{1, 5, 10, 15\}$) for two representative scenarios. Each row corresponds to a single method–scenario pair, enabling direct visual inspection of how the predicted ego trajectory evolves under the same temporal alignment.

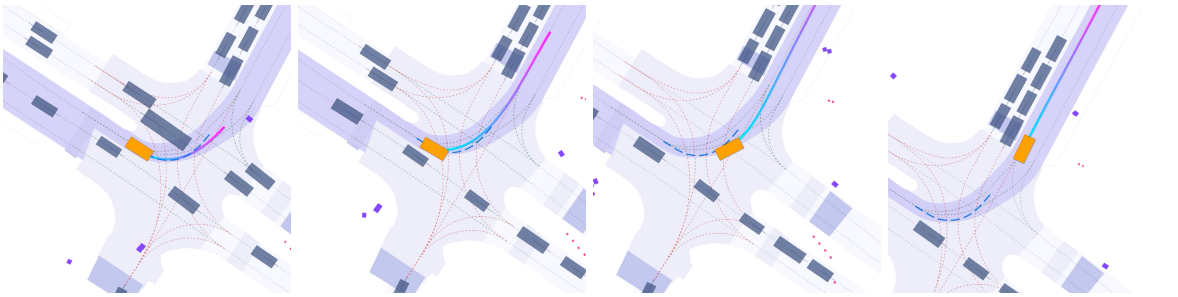
In the first scenario (rows (a)–(b)), PlanTF fails to complete the left-turn maneuver as the interaction unfolds, whereas CARE successfully executes the turn while maintaining a feasible trajectory within the lane geometry. In the second scenario (rows (c)–(d)), PlanTF exhibits an unsafe rollout that leads to a collision, while CARE preserves a safe margin to surrounding agents and continues with stable progress. These examples illustrate that the proposed planner yields more reliable closed-loop behavior in interactive or hazard-prone situations.

Overall, the qualitative results align with the quantitative improvements on the challenging `test14-hard` split. The combination of constrained ego-state attention (MDCA) and the risk-aware module contributes to a more robust planner compared to the SOTA baseline, reducing failure cases such as missed maneuvers and collisions under distribution-shifted conditions.



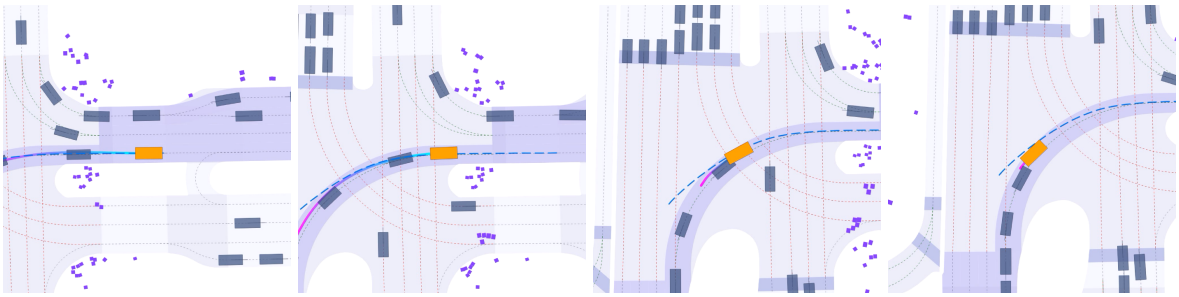
$t = 1$ $t = 5$ $t = 10$ $t = 15$

(a) PlanTF-1



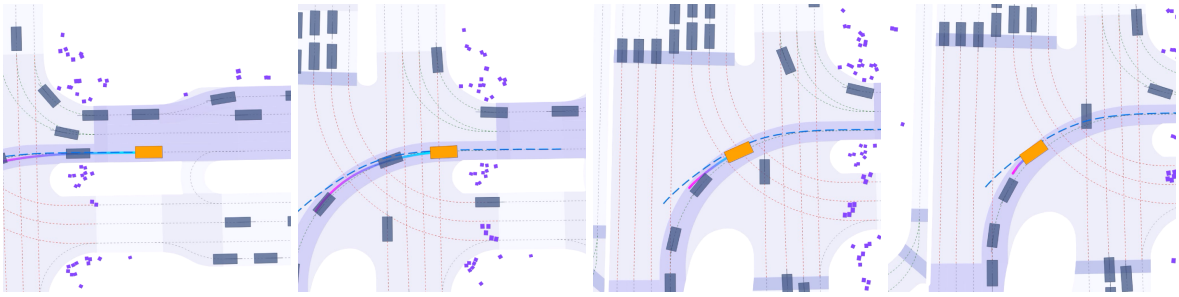
$t = 1$ $t = 5$ $t = 10$ $t = 15$

(b) CARE-1



$t = 1$ $t = 5$ $t = 10$ $t = 15$

(c) PlanTF-2



$t = 1$ $t = 5$ $t = 10$ $t = 15$

(d) CARE-2

Figure 5.5: Qualitative comparison at matched time steps ($t = \{1, 5, 10, 15\}$). Each row corresponds to a single method-scenario pair.

Chapter 6

Concluding Remark

6.1 Conclusion

This dissertation investigated robust closed-loop planning under distribution shift, where failures are often caused by two coupled factors: (i) shortcut learning in state-conditioned policies, and (ii) heavy-tailed risk induced by rare but severe events. The main contribution is a unified learning framework that combines constrained ego-state attention and tail-risk-aware supervision to improve reliability in challenging interactive scenarios.

First, Mean-Deviation Constrained Attention (MDCA) was introduced to prevent attention collapse and encourage balanced utilization of ego-state inputs. Rather than relying on a fixed penalty coefficient, the constraint was enforced using an Augmented Lagrangian Method (ALM), enabling adaptive regularization strength as training progresses. This design promotes stable optimization and supports robust state usage instead of over-reliance on a single dominant cue.

Second, a risk-aware learning component was incorporated to shape the trajectory distribution toward safer behaviors by suppressing high-risk tail events. While imitation learning can match average expert behavior, the proposed risk shaping explicitly targets rare failure cases that dominate closed-loop degradation. The resulting planner reduces

the frequency of high-risk rollouts and improves safety-oriented outcomes.

Experimental results on nuPlan demonstrated that the proposed planner achieves consistent gains on common scenarios and more pronounced improvements on hazard-focused evaluation. Sub-metric analyses indicated that improvements are driven primarily by safer interactions (e.g., reduced collision tendency and better TTC behavior) rather than aggressive progress seeking. Qualitative studies further supported these findings: paired trajectory–attention visualizations showed more stable and distributed attention patterns, ego-state ablation tests exhibited smaller performance gaps under state reduction, and risk distribution plots confirmed a reduced heavy-tail portion of the risk metric. Overall, the dissertation showed that combining MDCA-based attention regularization with tail-risk-aware supervision yields a more robust planner compared to strong baselines under distribution-shifted closed-loop conditions.

6.2 Future Work

Despite the observed improvements, several limitations remain and motivate future research directions.

Head-wise adaptive constraint control. The current MDCA formulation uses a single global Lagrangian multiplier λ to enforce the attention constraint. Although effective in many cases, a global λ can be overly coarse when multiple attention heads exhibit heterogeneous behaviors. As illustrated in Figure 6.1, a shared global margin may undesirably suppress attention in heads that already behave appropriately, while still failing to sufficiently correct a collapsed head. A natural extension is to introduce

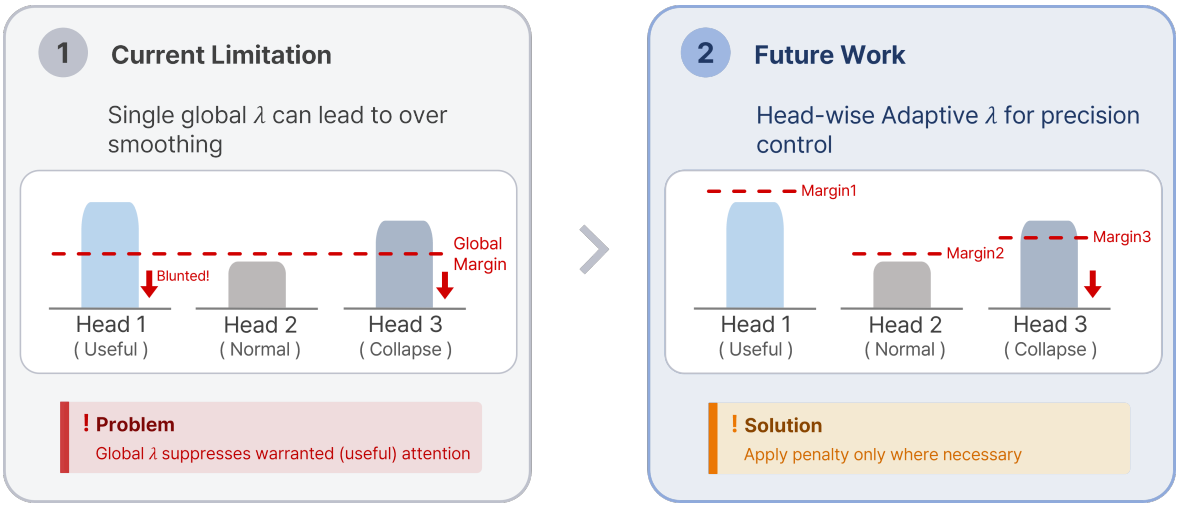


Figure 6.1: Limitations and future directions of Mean-Deviation Constrained Attention (MDCA).

head-wise (or channel-wise) adaptive multipliers, enabling more precise constraint enforcement by applying strong penalties only where necessary.

Broader evaluation beyond nuPlan. The experimental validation in this dissertation was conducted only on nuPlan. While nuPlan provides diverse urban scenarios and standardized closed-loop metrics, generalization to other large-scale benchmarks and real-world settings remains unverified. Future work should evaluate the proposed framework on additional autonomous-driving datasets and simulator stacks, and ultimately in more realistic environments with perception noise, map imperfections, and long-horizon interactions. Such studies would clarify robustness, transferability, and failure modes beyond the current evaluation domain.

Summary

Robust Imitation Learning with Attention Constraints and Risk Awareness for Motion Planning in Autonomous Driving

Robust closed-loop planning under distribution shift is challenging due to shortcut learning in state-conditioned policies and rare but severe high-risk events. This thesis proposes an imitation-learning framework that combines Mean-Deviation Constrained Attention (MDCA) with an Augmented Lagrangian Method (ALM) to prevent attention collapse and encourage balanced ego-state utilization, together with a tail-risk-aware objective that reshapes the multimodal trajectory distribution toward safer outcomes. The proposed planner is evaluated on the nuPlan benchmark and achieves consistent gains on the standard split and larger improvements on the hazard-focused split, with the gains concentrated in safety-relevant factors such as collision avoidance and TTC behavior rather than aggressive progress. Qualitative analyses further support the findings through paired trajectory–attention visualizations, reduced sensitivity under ego-state ablation, and reduced heavy-tail risk in distribution plots. Limitations include using a single global constraint coefficient across attention heads and evaluation restricted to nuPlan, motivating future work on head-wise adaptive constraints and broader validation on additional benchmarks and more realistic environments.

References

1. L. Chen, P. Wu, K. Chitta, B. Jaeger, A. Geiger, and H. Li, "End-to-end autonomous driving: Challenges and frontiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

2. D. A. Pomerleau, "Alvinn: An autonomous land vehicle in a neural network," *Advances in neural information processing systems*, vol. 1, 1988.

3. M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, *et al.*, "End to end learning for self-driving cars," *arXiv preprint arXiv:1604.07316*, 2016.

4. B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and autonomous systems*, vol. 57, no. 5, pp. 469–483, 2009.

5. S. Schaal, "Learning from demonstration," *Advances in neural information processing systems*, vol. 9, 1996.

6. M. Bansal, A. Krizhevsky, and A. Ogale, "Chauffeurnet: Learning to drive by imitating the best and synthesizing the worst," *arXiv preprint arXiv:1812.03079*, 2018.

7. M. Vitelli, Y. Chang, Y. Ye, A. Ferreira, M. Wołczyk, B. Osiński, M. Niendorf, H. Grimmett, Q. Huang, A. Jain, *et al.*, "Safetynet: Safe planning for real-world

self-driving vehicles using machine-learned policies," in *2022 International Conference on Robotics and Automation (ICRA)*, pp. 897–904, IEEE, 2022.

8. S. Pini, C. S. Perone, A. Ahuja, A. S. R. Ferreira, M. Niendorf, and S. Zagoruyko, "Safe real-world autonomous driving by learning to predict and plan with a mixture of experts," *arXiv preprint arXiv:2211.02131*, 2022.

9. Z. Huang, H. Liu, and C. Lv, "Gameformer: Game-theoretic modeling and learning of transformer-based interactive prediction and planning for autonomous driving," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3903–3913, 2023.

10. Z. Huang, H. Liu, J. Wu, and C. Lv, "Differentiable integrated motion prediction and planning with learnable cost function for autonomous driving," *IEEE transactions on neural networks and learning systems*, vol. 35, no. 11, pp. 15222–15236, 2023.

11. Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 157–166, 1994.

12. R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *International conference on machine learning*, pp. 1310–1318, Pmlr, 2013.

13. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied

to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 2002.

14. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," vol. 30, 2017.

15. K. Renz, K. Chitta, O.-B. Mercea, A. Koepke, Z. Akata, and A. Geiger, "Plant: Explainable planning transformers via object-level representations," *arXiv preprint arXiv:2210.14222*, 2022.

16. Y. Hu, J. Yang, L. Chen, K. Li, C. Sima, X. Zhu, S. Chai, S. Du, T. Lin, W. Wang, *et al.*, "Planning-oriented autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 17853–17862, 2023.

17. K. Chitta, A. Prakash, B. Jaeger, Z. Yu, K. Renz, and A. Geiger, "Transfuser: Imitation with transformer-based sensor fusion for autonomous driving," *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 11, pp. 12878–12895, 2022.

18. S. Ross, G. Gordon, and D. Bagnell, "A reduction of imitation learning and structured prediction to no-regret online learning," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pp. 627–635, JMLR Workshop and Conference Proceedings, 2011.

19. J. Zhou, R. Wang, X. Liu, Y. Jiang, S. Jiang, J. Tao, J. Miao, and S. Song, "Exploring imitation learning for autonomous driving with feedback synthesizer

and differentiable rasterization," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1450–1457, IEEE, 2021.

20. C.-C. Chuang, D. Yang, C. Wen, and Y. Gao, "Resolving copycat problems in visual imitation learning via residual action prediction," in *European Conference on Computer Vision*, pp. 392–409, Springer, 2022.

21. C. Wen, J. Lin, T. Darrell, D. Jayaraman, and Y. Gao, "Fighting copycat agents in behavioral cloning from observation histories," *Advances in Neural Information Processing Systems*, vol. 33, pp. 2564–2575, 2020.

22. R. Geirhos, J.-H. Jacobsen, C. Michaelis, R. Zemel, W. Brendel, M. Bethge, and F. A. Wichmann, "Shortcut learning in deep neural networks," *Nature Machine Intelligence*, vol. 2, no. 11, pp. 665–673, 2020.

23. A. Ilyas, S. Santurkar, D. Tsipras, L. Engstrom, B. Tran, and A. Madry, "Adversarial examples are not bugs, they are features," *Advances in neural information processing systems*, vol. 32, 2019.

24. A. D'Amour, K. Heller, D. Moldovan, B. Adlam, B. Alipanahi, A. Beutel, C. Chen, J. Deaton, J. Eisenstein, M. D. Hoffman, *et al.*, "Underspecification presents challenges for credibility in modern machine learning," *Journal of Machine Learning Research*, vol. 23, no. 226, pp. 1–61, 2022.

25. S. Jain and B. C. Wallace, "Attention is not explanation," *arXiv preprint arXiv:1902.10186*, 2019.

26. S. Serrano and N. A. Smith, "Is attention interpretable?," *arXiv preprint arXiv:1906.03731*, 2019.

27. R. T. Rockafellar, S. Uryasev, *et al.*, "Optimization of conditional value-at-risk," *Journal of risk*, vol. 2, pp. 21–42, 2000.

28. A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press, 2015.

29. A. Tamar, Y. Chow, M. Ghavamzadeh, and S. Mannor, "Policy gradient for coherent risk measures," *Advances in neural information processing systems*, vol. 28, 2015.

30. Y. Chow, A. Tamar, S. Mannor, and M. Pavone, "Risk-sensitive and robust decision-making: a cvar optimization approach," *Advances in neural information processing systems*, vol. 28, 2015.

31. A. Hakobyan, G. C. Kim, and I. Yang, "Risk-aware motion planning and control using cvar-constrained optimization," *IEEE Robotics and Automation letters*, vol. 4, no. 4, pp. 3924–3931, 2019.

32. J. Cheng, Y. Chen, X. Mei, B. Yang, B. Li, and M. Liu, "Rethinking imitation-based planners for autonomous driving," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 14123–14130, IEEE, 2024.

33. H. Caesar, J. Kabzan, K. S. Tan, W. K. Fong, E. Wolff, A. Lang, L. Fletcher,

O. Beijbom, and S. Omari, "nuplan: A closed-loop ml-based planning benchmark for autonomous vehicles," *arXiv preprint arXiv:2106.11810*, 2021.

34. N. Karnchanachari, D. Geromichalos, K. S. Tan, N. Li, C. Eriksen, S. Yaghoubi, N. Mehdipour, G. Bernasconi, W. K. Fong, Y. Guo, *et al.*, "Towards learning-based planning: The nuplan benchmark for real-world autonomous driving," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 629–636, IEEE, 2024.

35. H. Cui, V. Radosavljevic, F.-C. Chou, T.-H. Lin, T. Nguyen, T.-K. Huang, J. Schneider, and N. Djuric, "Multimodal trajectory predictions for autonomous driving using deep convolutional networks," in *2019 international conference on robotics and automation (icra)*, pp. 2090–2096, IEEE, 2019.

36. Y. Chai, B. Sapp, M. Bansal, and D. Anguelov, "Multipath: Multiple probabilistic anchor trajectory hypotheses for behavior prediction," *arXiv preprint arXiv:1910.05449*, 2019.

37. T. Phan-Minh, E. C. Grigore, F. A. Boulton, O. Beijbom, and E. M. Wolff, "Covernet: Multimodal behavior prediction using trajectory sets," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14074–14083, 2020.

38. J. Cheng, Y. Chen, and Q. Chen, "Pluto: Pushing the limit of imitation learning-based planning for autonomous driving," *arXiv preprint arXiv:2404.14327*, 2024.

39. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

40. M. S. A. Baig, S. A. Gillani, A. A. Khan, S. M. Shah, and M. O. Khan, "Attentiondrop: A novel regularization method for transformer models," *arXiv preprint arXiv:2504.12088*, 2025.

41. J. Kim and K. Choi, "Mitigating attention collapse via mean-deviation constrained optimization," Institute of Electrical and Electronics Engineers (IEEE), Nov. 2025.

42. D. P. Bertsekas, "Nonlinear programming," vol. 48, pp. 334–334, Taylor & Francis, 1997.

43. J. Nocedal and S. J. Wright, *Numerical optimization*. Springer, 2006.

44. F. Fioretto, P. Van Hentenryck, T. W. Mak, C. Tran, F. Baldo, and M. Lombardi, "Lagrangian duality for constrained deep learning," in *Joint European conference on machine learning and knowledge discovery in databases*, pp. 118–135, Springer, 2020.

45. S. Park and P. Van Hentenryck, "Self-supervised primal-dual learning for constrained optimization," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, pp. 4052–4060, 2023.

46. A. Santara, A. Naik, B. Ravindran, D. Das, D. Mudigere, S. Avancha, and B. Kaul, "Rail: Risk-averse imitation learning," *arXiv preprint arXiv:1707.06658*, 2017.

47. J. Ho and S. Ermon, "Generative adversarial imitation learning," *Advances in neural information processing systems*, vol. 29, 2016.

48. D. Dauner, M. Hallgarten, A. Geiger, and K. Chitta, "Parting with misconceptions about learning-based vehicle motion planning," in *Conference on Robot Learning*, pp. 1268–1281, PMLR, 2023.

Acknowledgements

2년간의 석사 과정은 제게 참 다사다난한 시간이었습니다. 때로는 막막하고 조급함에 힘들었던 순간도 많았지만, 그럴 때마다 곁에서 건네주신 따뜻한 도움과 조언 덕분에 끝까지 걸어올 수 있었습니다. 각자의 자리에서 묵묵히 최선을 다하는 연구실 사람들을 보며, 부족한 저 또한 마음을 다잡고 이 과정을 잘 마무리할 수 있었던 것 같습니다.

먼저, 앞장서서 연구실을 이끌어 주시고 언제나 아낌없이 생각과 조언을 나눠주신 승훈님, 지난 2년간 옆자리에서 수많은 재밌는 대화를 나누고, 서로 장난도 치며 힘든 순간을 가볍게 웃어넘길 수 있게 해준 옆자리 메이트 동화님, 고민을 기꺼이 들어주고 응원해주며 늘 겸손한 태도로 제 마음가짐을 돌아보게 해준 민석님, 산책도 하고 커피도 마시며 함께 환기할 수 있었던 명석님, 연구도 생활도 끊임없는 자세로 늘 중심을 잡아주었던 효찬님, 그리고 누구보다 제 일에 진심으로 기뻐해주시고 따뜻하게 응원해주셨던 mouri 박사님께도 깊이 감사드립니다. 늘 성실하고 멋진 모습으로 자극이 되어준 창은씨와 naol, 그리고 angus까지, 여러분 모두가 있었기에 저는 더 힘을 내고 더 열심히 할 수 있었습니다.

무엇보다도 가장 큰 감사의 마음은, 제게 아낌없는 기회를 주시고 끝까지 믿어주신 최경환 교수님께 전하고 싶습니다. 연구의 방향뿐 아니라 늘 열정적이신 교수님을 바라보며 태도와 기준까지 배울 수 있었던 소중한 시간이었습니다. 진심으로 감사드립니다.

마지막으로, 어쩌면 심통도 많이 부리고 칭얼대기도 했던 막내를 언제나 변함없이 지지해준 가족에게 감사하다는 말을 전합니다. 엄마, 아빠, 그리고 언니 늘 제 편이 되어줘서 감사하고, 받은 사랑을 꼭 보답하겠습니다. 세상에서 제일 사랑합니다.

이 글에 다 담지 못한 AI대학원 동기들께 전하는 감사의 마음까지 포함해, 이 여정을 함께해주신 모든 분들께 진심으로 감사드리며 늘 응원합니다.