

## 自律走行のための模倣に基づくプランナー再考

Jie Cheng<sup>1</sup>, Yingbing Chen<sup>1</sup>, Xiaodong Mei<sup>1</sup>, Bowen Yang<sup>1</sup>, Bo Li<sup>2</sup> and Ming Liu<sup>1,3</sup>

概要- 近年、模倣型ドライビングプランナーが大きな成功を収めている。しかし、標準化されたベンチマークがないため、様々なデザインの有効性は不明なままである。新たにリリースされたnuPlanは、大規模な実世界データセットと、公平な比較のための標準化されたクローズドループベンチマークを提供することで、この問題に対処している。このプラットフォームを利用して、我々は模倣ベースのプランナーの2つの基本的かつ未開拓の側面、すなわち、エゴプランニングに不可欠な特徴と、複合エラーを減らすための効果的なデータ補強技術について包括的な研究を行う。さらに、現在の学習システムでは見過ごされてきた模倣のギャップを強調する。最後に、我々の知見を統合し、強力なベースラインモデルであるPlanTFを提案する。我々の結果は、よく設計された純粋に模倣に基づくプランナが、手作業で作成されたルールを含む最先端の手法と比較して、非常に競争力のある性能を達成でき、ロングテールのケースで優れた汎化能力を示すことを示している。我々のモデルとベンチマークは公開されている。プロジェクトサイト <https://jchengai.github.io/planTF>.

## 1. INTRODUCTION

学習ベースのプランナは、従来のルールベースのプランナに取って代わる、自律走行のためのスケーラブルなソリューションであると考えられている[1]-[3]。このことは、近年、大きな研究関心を引き起こしている。特に、模倣ベースのプランナ[4]-[12]は、シミュレーションや実世界のシナリオで顕著な成功を収めることが報告されている。とはいえ、これらのプランナーは、標準化されたベンチマークがないため、多様なカスタム条件(例えば、様々なデータセット、メトリクス、シミュレーションセットアップ)で主に訓練され、評価される。その結果、実用的な学習ベースのシステムを構築するための効果的な設計選択肢を比較し、要約することは困難になる。

最近、大規模なnuPlan [13]データセットが、標準化されたシミュレーションベンチマークとともに公開され、学習されたモーションプランナーを進歩させる新たな機会が提供された。この新しいベンチマークによって、我々は、学習ベースのプランナのいくつかの一般的で重要でありながら十分に研究されていない設計上の選択について詳細な調査を行い、将来の研究のための建設的な提案を提供することを目的とする。本論文では、模倣ベースのプランナの2つの包括的かつ基本的な側面、すなわち、プランニングに必要なエゴの特徴と、データ増強の効果的なテクニックに焦点を当てる。

模倣に基づく計画モデル[5]~[11]の大半は、予測モデルの成功を踏襲しており、入力特徴として自律走行車(AV)の過去の軌跡を本質的に組み込んでいるが、模倣学習(IL)は過去の観測から近道を取得する傾向があることが頻繁に指摘されている[4]、[14]~[16]。

Jie Cheng, Yingbing Chen, Xiaodong Mei, Bowen Yang は香港科学技術大学(中国、香港特別行政区)に所属している。{jchengai,ychengz,xmeiab,byangar}@connect.ust.hk <sup>2</sup> Bo Li はロータス・テクノロジー社に所属している。libo@lotuscar.com.cn <sup>3</sup>Ming Liu is also with the Hong Kong University of Science and Technology (Guangzhou), Nansha, Guangzhou, China. eelium@ust.hk

我々の研究は、AVの過去の動作が、クローズドループの性能の大幅な劣化につながることを再確認するものである。プランナーはAVの現在の状態のみを利用することで、性能向上を実現する。意外なことに、AVの現在のポーズ(位置と方位)を純粋に使用することで、より優れた閉ループ性能を達成する。このことは、速度、加速度、ステアリングなど、一般的に計画に重要とみなされる追加的な運動学的属性が、性能低下につながることを示唆している。この現象をより深く理解するために、AVの状態が結果の軌道に与える影響を評価する感度分析を行う。我々の実験から、プランナは過去の運動データがない場合でも、その運動学的状態からショートカットを利用することを学習できることが明らかになった。この課題を軽減するために、我々は、運動学的状態を利用するプランナが最適な全体性能を達成することを可能にする、単純でありながら非常に効果的な注意ベースの状態ドロップアウトエンコーダを導入した。

模倣学習には複合誤差があることも知られている[17]。摂動に基づく補強[4]-[6]は、逸脱からの回復をプランナに指示するためによく採用される戦略である。我々は、履歴摂動、状態摂動、将来の補正を含む、様々なオーグメンテーション技術を探求する包括的な実験を行う。さらに、オーグメンテーションの有効性のために、適切な正規化が不可欠であることを実証する。さらに、現在の学習フレームワークの中で無視されている模倣のギャップを特定し、その潜在的な影響を説明する。

最後に、我々の発見を組み合わせることで、我々の標準化されたnuPlanベンチマークにおいて、最先端の競合に対して強力な性能を示す、純粋な学習ベースのベースラインモデルを提供する。我々の貢献は以下のように要約される:

- 1) 自我計画に必要な特徴について詳細な調査を行い、主流の慣行に反する直感に反する結果を得た。さらに、全体として最も高い性能を達成する、効果的な注意に基づく状態脱落エンコーダを導入した。
- 2) 我々は、様々なオーグメンテーション技術を含む包括的な実験を行い、それによって複合エラーを軽減する効果的な戦略を解明した。さらに、現在の学習フレームワークでは、見過ごされている模倣のギャップが確認された。
- 3) By combining our findings, we provide an open baseline model with strong performance. All our code, benchmarks, and models will be publicly released, as a reference for future research.

## II. RELATED WORK

模倣ベースのプランナは、その収束の容易さとデータに対する典型的なスケラビリティにより、学習ベースのプランナの中で非常に好まれている。入力タイプによって2つのグループに分類することができます:

1) エンドツーエンド。エンドツーエンド(E2E)手法[18]~[29]は、生のセンサー入力を使用して将来の軌道を直接生成する。クローズドループCARLAベンチマーク[30]とオープンソースコミュニティの協調的な取り組みを活用することで、E2E手法は短期間で目覚ましい進歩を遂げた: 最初の基本的なCNNベースのアプローチ(LBC [23], CILRS [22])から、マルチモーダル融合(Transfuser [26], NEAT [25], MMFN [27], Interfuser [26], ThinkTwice [29]), および統合された知覚と計画戦略(LAV [24], STP3 [19], VAD [21])を包含する。しかし、シミュレーション環境の制約から、これらの手法は一般的に低速で機能し、シミュレーションされた交通エージェントの挙動はリアルさと多様性に欠ける。データ駆動型交通シミュレーション[31], [32]や現実的なセンサーエミュレーション[33], [34]などの新しく興味深い研究は、これらの問題を軽減する可能性を秘めている。

2) 中位から中位。これらのアプローチ[4], [6]-[12], [35], [36]は、知覚後の結果を入力として利用し、記録された実世界のデータから直接学習することができる。Chaffernet [4]は、共変量シフトを緩和するために摂動軌道の合成を導入しており、この実践はその後の研究で一般的になっている。[5]は、学習データをさらにオンポリシーロールアウトで補強する。いくつかの研究により、実車両の運用能力が実証されている(SafetyNet [6], UrbanDriver [7], SafetyPathNet [9])。多くの場合、プランナーのロバスト性を高めるために、ポストオブティマイザ(DIPP [11], GameFormer [10], hotplan [35], pegasus [37])を搭載している。ホットプランを除く上記の方法は、すべてAVの履歴モーションを使用している。本研究では、このカテゴリに焦点を当て、標準化されたデータとベンチマークに基づくいくつかの重要な設計上の選択について、詳細な調査を提供する。

模倣を超えて。環境損失の活用[4], [5], ILと強化学習の統合[38]~[40], クローズドループ学習[41]~[43]としても知られる敵対的学習の組み込みなど、純粋な模倣学習(IL)の本質的な限界[14]~[16]を克服することを目的とした研究もある。我々の研究は、純粋なILベースのプランナが限界に達しておらず、適切な設計によって大幅に改善できることを示している。

### III. RETHINK I MITAION ベースのランナー

データセットから専門家の軌跡を模倣して学習したプランナを用いた都市ナビゲーションのタスクを考える。各計画反復において、プランナは、2秒間の履歴ウィンドウまでの周囲物体の追跡データ、自車両の現在と過去の運動状態、信号機に関する情報、高精細(HD)マップ、速度制限、指定ルートなど、様々な入力を受け取る。プランナーは、その後の8秒間の軌道を生成するタスクを課される。

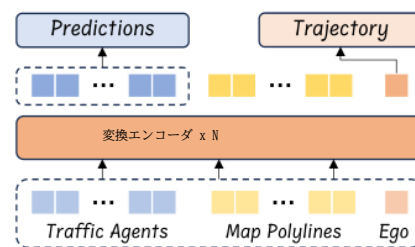


図1. ベースラインモデルの概要エージェント、マップ、エゴの特徴は別々にエンコードされ、その後連結され、その後、変換エンコーダ層のスタックによって処理される。ベースラインモデルは、シーンレベルでの自車両の交通エージェントと計画を共同で予測する。

特に断りのない限り、本論文ではプランナーから出力される変更されていない軌道を採用していることに注意する必要がある。プランナー固有のパフォーマンスを評価するために、ルールベースの緊急停止や最適化後などのパフォーマンス向上技術を意図的に取り入れることは避けている。

nuPlan [13]は、自律走行車のための大規模な閉ループMLベースの計画ベンチマークである。このデータセットには、4つの都市中心部から収集された1300時間の記録された運転データが含まれ、自動ラベリングツールを用いて75のシナリオタイプにセグメンテーションされている。

シミュレーションを行う。シミュレーション環境として、nuPlanの閉ループシミュレータを使用する。各シミュレーションは、10Hzのレートで15秒間のロールアウトを伴う。軌道追従のためにLQRコントローラを採用し、制御コマンドは内部運動学自転車モデルを通じて自律走行車の状態を更新するために利用される。背景交通の挙動はシミュレーションモードによって異なり、非反応的(ログリプレイ)または反応的である。

メトリクス。オープンループスコア(OLS)、非反応性クローズドループスコア(NR-CLS)、反応性クローズドループスコア(R-CLS)を含む、nuPlanが提供する公式評価指標を採用する。R-CLSとNR-CLSは、シミュレーション中にインテリジェント・ドライバ・モデル(IDM)[44]を介してバックグラウンド・トラフィック制御を組み込むという点で異なるだけで、同じ計算方法論を共有している。クローズドループスコアは包括的な複合スコアであり、交通ルールの遵守、人間の運転の類似性、車両のダイナミクス、目標達成、およびシナリオに特有の他のメトリクスなどの要因の重み付けされた組み合わせによって達成される。スコアは0から100までの範囲で評価される。メトリクスの詳細な説明と計算については、[45]を参照してください。

ベースライン。ベースラインとして、我々は先行研究[46]のモーションフォアキャスティングバックボーンモデルをプランニングタスクに適応させた。図1は、ベースラインモデルの概要を簡潔に示している。その単純さにもかかわらず、アーキテクチャは主に複数のTransformerエンコーダ[47]で構成され、大きなモデリング能力を示している。詳細はコードベースを参照されたい。

ベンチマーク。全ての実験において、訓練と評価のためにデータ分割を標準化する。学習フェーズでは、nuPlan学習セットの75のシナリオタイプすべてを利用し、シナリオの総数を1Mフレームに制限する。

| Models        |                    | Test14-random |              |              | Test14-hard  |              |              |
|---------------|--------------------|---------------|--------------|--------------|--------------|--------------|--------------|
| Input feature | Variants           | OLS           | NR-CLS       | R-CLS        | OLS          | NR-CLS       | R-CLS        |
| w/ history    | shared encoder     | 90.20         | 56.50        | 56.28        | <b>88.25</b> | 48.60        | 51.32        |
|               | seperate encoder   | <b>90.28</b>  | 61.02        | 59.85        | 86.77        | 51.98        | 49.34        |
| w/o history   | state3 (x, y, yaw) | 81.13         | <b>85.99</b> | <b>79.38</b> | 71.43        | 68.44        | <b>63.14</b> |
|               | state4 (+vel.)     | 86.42         | 81.32        | 75.75        | 82.30        | 68.15        | 62.51        |
|               | state5 (+acc.)     | 87.71         | 81.76        | 74.51        | 84.54        | <b>68.67</b> | 54.91        |
|               | state6 (+steer)    | 88.45         | 83.32        | 77.52        | 85.93        | 65.15        | 55.99        |

表 1. 異なる入力特徴量の結果履歴を入力とするモデルにおいて、「共有」エンコーダと「分離」エンコーダは、エージェントとエゴの両方が共有履歴エンコーダを利用するか、それとも別個のエンコーダを利用するかを示す。履歴入力のないモデルの場合、+は表中の前のモデルと比較して、追加の状態変数を含めることを意味する。値が高いほど、すべてのメトリクスでより良いパフォーマンスを示し、最良のメトリクスは太字で強調表示されている。

評価フェーズでは、nuPlan Planning Challengeで指定された14種類のシナリオを採用し、それぞれ20のシナリオから構成される。(1)Test14random:シナリオを各タイプからランダムにサンプリングし、選択後に固定する。(2)Test14-hard:ロングテールシナリオにおけるプランナーの性能を調べるため、最先端のルールベースプランナー(PDM-Closed [48])を用いて各タイプのシナリオを100回実行し、その後、各タイプの最も性能の低い20個のシナリオを選択する。シナリオの例はプロジェクトページで見ることができる。オンラインリーダーボードの提出が終了したため、すべての評価は nuPlan 公開テストセットで実施した。

#### A. 入力機能が違いを生む

本節では、以下の問いを解決することを目的とする:(1)過去のモーションデータはプランニングに不可欠か?(2)そうでない場合、自律走行車の現在の状態はすべてプランナーの性能向上に寄与するか?これらの疑問を解決するために、我々はベースラインモデルから得られた2つのバリエーションセットを含む調査を実施した。Test14-randomとTest14-hardベンチマークの結果を表1に示す。2つの履歴変種のうち、1つは他の交通エージェントと履歴エンコーダを共有し、もう1つは自車両の過去の動作に別個の履歴エンコーダを採用している。状態のみのモデルの場合、車両の姿勢、速度、加速度、操舵角など、従来のプランナーに不可欠ないくつかの重要な状態変数を精査した。実験結果に基づき、以下の知見を得た:

履歴は必要ない。過去のモーションデータを取り入れたモデルは、優れたオフポリシー評価性能(OLS)を示すが、クロードループメトリクスでは、状態ベースのモデルと比較して著しく劣る性能を示す。この現象は、よく知られた「コピーキャット」問題[16]や学習ショートカット[49]に起因している可能性があり、プランナーは、根本的な原因要因を包括的に把握することなく、過去のデータからの外挿に頼っている。さらに、履歴モデルのオープンループ性能における優位性は、状態のみのモデルでは状態数が増加するにつれて急速に減少する。したがって、履歴運動は計画モデルには必要ないと結論づけられる。

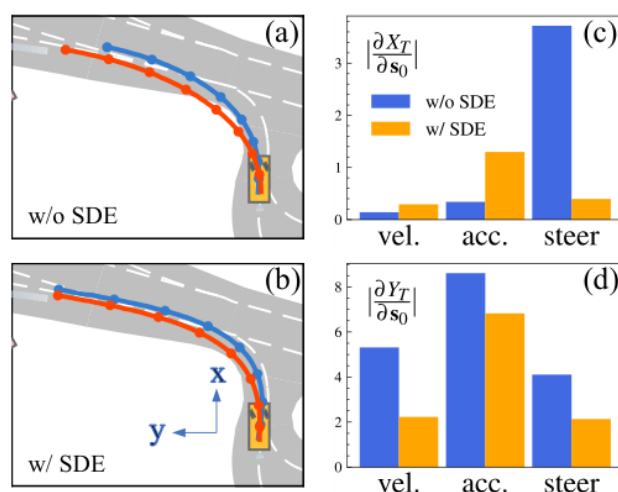


図2. 左側は、AVのステアリング角度を0.15~0.5radに調整した状態6モデルの計画軌跡を示す。右側は、AVの運動状態に対する軌道終点の位置に関する勾配の大きさを示している。

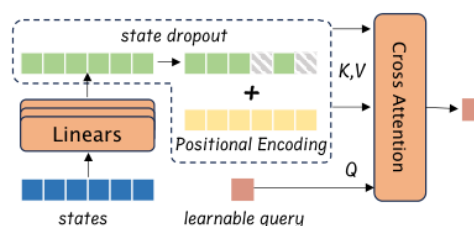


図3. 注意に基づく状態脱落エンコーダの説明図。

運動学的状態におけるショートカット学習速度や加速度などの運動状態は、軌道計画における安全性と快適性を確保するための重要な初期境界条件として機能する。とはいえ、自律走行車(AV)のポーズ(位置と方位からなる)のみに依存するstate3モデルが、CLSメトリクスの点で運動学的状態を組み込んだ他のモデルを大幅に上回ることを発見したのは驚きである。この現象をより深く理解するために、state6モデルの左折ケースを研究する。

|        |     | Test14-random |               |               | Test14-hard   |               |               |
|--------|-----|---------------|---------------|---------------|---------------|---------------|---------------|
|        | SDE | OLS           | NR-CLS        | R-CLS         | OLS           | NR-CLS        | R-CLS         |
| staet3 | -   | 81.13         | 85.99         | 79.38         | 71.43         | 68.44         | 63.14         |
| state5 | ✗   | 87.71         | 81.76         | 74.51         | 84.54         | 68.67         | 54.91         |
|        | ✓   | 88.80 (+1.09) | 86.73 (+4.97) | 75.75 (+1.24) | 84.29 (-0.25) | 71.28 (+2.61) | 61.88 (+6.97) |
| state6 | ✗   | 88.55         | 83.19         | 74.79         | 85.89         | 67.57         | 58.99         |
|        | ✓   | 87.07 (-1.48) | 86.48 (+3.29) | 80.59 (+5.80) | 83.32 (-2.57) | 72.68 (+5.11) | 61.70 (+2.71) |

表 II. Test14-randomとTest14-hardベンチマークにおける状態ドロップアウトエンコーダ(SDE)の実験結果。SDEを用いたモデルは、OLSで高い性能を維持しながら、CLSで大幅な改善を得る。

図2aに示すように、ステアリング角度を0.15(青)から0.5(赤)radに変更した場合、モデルは望ましくないオフロード軌道を生成する。我々は、このモデルは過去の観測結果がない場合でも、運動学から誤った相関を学習するという仮説を立てた。

状態ドロップアウトエンコーダ。我々の仮定に立ち向かうために、図3に示すような注意に基づく状態ドロップアウトエンコーダ(SDE)を提案する。各状態変数は、位置エンコーディングと組み合わせられる前に、線形レイヤーを通して個々の埋め込みを受ける。学習可能なクエリは、クロスアテンションモジュールを通して状態埋め込みを集約する。学習中、埋め込まれた状態(位置と方位を除く)トークンは、ある確率で削除される。エンコーダは、補助情報へのアクセスに部分的な制約を課すことで、モデルに行動の根本原因を明らかにさせる。一方、運動学的属性にアクセスできる場合、モデルは計画能力を向上させることができる。state5とstate6モデルで状態ドロップアウトエンコーダを実装し、その結果を表IIに示す。この結果から、SDEの利用はモデルの閉ループ性能を大幅に向上させることが示された。重要なことは、state3と比較した場合、state5とstate6にSDEを追加したモデルは、クロズドループスコアが改善されただけでなく、オープンループスコアが大幅に向上したことであり、SDEの有効性を示す説得力のある証拠となる。state3モデルは、すべての運動学的情報を失うため、基本的に曖昧であることを指摘する(OLSの性能の低さによって裏付けられている)。図2(a)(b)はstate6モデルの比較計画結果、図2(c)(d)は初期運動状態 $s_0$ に対する終点位置の勾配( $X_T$ ,  $Y_T$ )の大きさの比較結果である。この結果は、SDEを採用したモデルは運動状態の変動に対する感度が低く、その結果、より弾力的な計画結果が得られることを示している。

#### B. データの増強と正規化

データ補強は、ILベースのモデルが偏差から回復する方法を学習するための一般的な手法である。本節では、複合エラーを軽減するための効果的な戦略を探ることを目的として、様々なデータ補強技術に関する包括的な実験を行う。図4には、さまざまなオーグメンテーション戦略が表示されている。図4(a)では、すべての座標を自律走行車の中心に対して正規化した運転シナリオの例が描かれている。図4(b)では、ランダムにサンプリングされたノイズ(摂動)

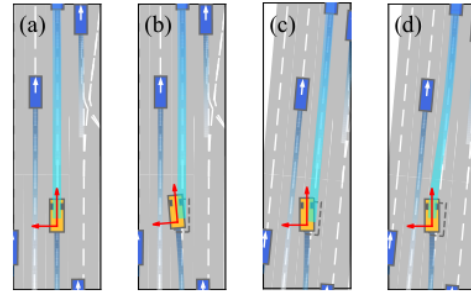


図4 (a) 元のシナリオ。(b) AVの現在の状態にランダムノイズを加え、履歴の動きを平滑化する。(c) シナリオの座標は、AVの摂動位置に基づいて再正規化される。(d) 制約付き非線形最適化を用いて、補正された将来の軌道を生成する。

|                | P | RN | FC | OLS   | NR-CLS | R-CLS |
|----------------|---|----|----|-------|--------|-------|
| history        | ✗ | ✗  | ✗  | 88.99 | 65.84  | 65.58 |
|                | ✓ | ✓  | ✓  | 89.94 | 65.14  | 66.03 |
| state3         | ✗ | ✗  | ✗  | 78.92 | 71.86  | 70.87 |
|                | ✓ | ✗  | ✗  | 80.85 | 74.28  | 71.69 |
|                | ✓ | ✓  | ✗  | 81.13 | 85.99  | 79.38 |
|                | ✓ | ✓  | ✓  | 79.28 | 81.35  | 76.60 |
| state5         | ✗ | ✗  | ✗  | 88.44 | 80.67  | 74.50 |
|                | ✓ | ✗  | ✗  | 89.20 | 79.85  | 72.43 |
|                | ✓ | ✓  | ✗  | 87.71 | 81.76  | 74.51 |
|                | ✓ | ✓  | ✓  | 86.43 | 82.10  | 74.71 |
| state6<br>+SDE | ✗ | ✗  | ✗  | 88.33 | 77.28  | 74.10 |
|                | ✓ | ✗  | ✗  | 87.71 | 77.70  | 75.18 |
|                | ✓ | ✓  | ✗  | 87.07 | 86.48  | 80.59 |
|                | ✓ | ✓  | ✓  | 85.50 | 82.95  | 76.09 |

表III. Test14-randomベンチマークにおける、異なるオーグメンテーションと正規化の組み合わせの結果。P:摂動、RN:再正規化、FC:未来補正。

はAVの現在の状態に加算され、それに応じて履歴の状態が平滑化される。図4(c)では、シナリオの座標が摂動後の自律走行車の中心に対して再正規化されていることが示されている。図4(d)は、非線形最適化による整流された未来軌道の生成を示す。基本的に、図4(b)と図4(d)に描かれた両戦略は、共通の目的に合致している。



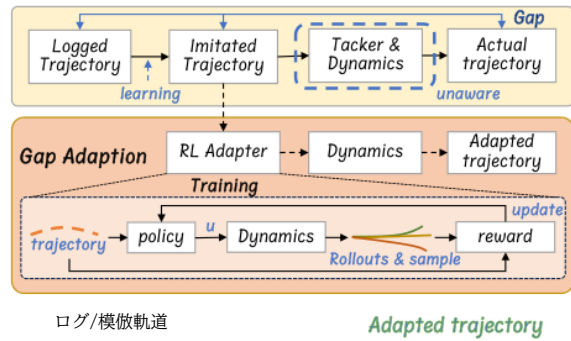


図5. 模倣ギャップと提案するRLアダプタの説明図。

|                  | Test14-random |       | Test14-hard |       |
|------------------|---------------|-------|-------------|-------|
| Log-replay +     | NR-CLS        | R-CLS | NR-CLS      | R-CLS |
| Perfect tracking | 96.63         | 77.38 | 91.61       | 71.34 |
| LQR              | 94.03         | 75.86 | 85.96       | 68.80 |
| RL Adapter       | 96.3          | 77.13 | 91.65       | 71.62 |

表 IV. Test14-randomとTest14-hardベンチマークにおける、異なるトラックを用いたログリプレイプランナー(完全模倣)の実験結果。LQRはnuPlanベンチマークで使用するデフォルトのトラックであり、RLアダプターは模倣ギャップに対処するために提案された方法である。

車両を専門家の軌道に戻すように誘導する。

表IIIは、4つのモデルバリエーションについて、様々な増強戦略で行った実験の結果を示している。これらの結果から、以下の知見が得られた：(1)履歴モデルと状態モデル5の場合、どのデータ補強も実質的な強化を示さない。我々は、これら2つのモデルが直面する主な課題は、因果関係の混乱、すなわち、歴史的または運動学的状態からの外挿の問題であると仮定する。(2) state3とstate6+SDEモデルでは、摂動は非常に重要であるが、適切な正規化でのみ機能する。例えば、state3モデルのNR-CLSスコアは摂動と再正規化で71.86から85.99に向上し、摂動のみを使うよりはるかに高い(74.28)。このことは、学習とテストの間、データ分布を近づけることが重要であることを示唆している。(3)修正されたガイドとなる将来の軌道を提供することは、プラスの効果をもたらさない。考えられる理由の1つは、手動で生成された軌跡が専門家の軌跡分布と一致しないことである。エキスパート軌道を直接監視として使用することは、元の分布を保持し、小さな偏差をトラックで簡単に修正できるため、より効果的な選択である。

### C. 隠された模倣のギャップ

模倣のギャップ。最も一般的な模倣学習フレームワークの中で、モデルはデータセットから記録された専門家の足跡を模倣する。

| Reward Term       | Expression                                     | Weight |
|-------------------|--|--------|
| Position Tracking | $e^{-15}  \mathbf{p}_{xy}-\mathbf{p}_{xy}^*  $ | 1.0    |
| Action            | $  \mathbf{u}  ^2$                             | -0.01  |
| Action Rate       | $  \dot{\mathbf{u}}  ^2$                       | -0.1   |
| Lon. Acc. limit   | $\mathbb{1}(\dot{v} > 2.4)$                    | -1     |
| Jerk limit        | $\mathbb{1}( \ddot{v}  > 4.0)$                 | -1     |
| Yaw rate          | $\mathbb{1}( \dot{\theta}  > 0.95)$            | -0.5   |

表 V. RLアダプタの報酬項と表現。アクションuは加速度とステアリングレートを含む。vとθはAVの縦方向と方位角を意味する。

我々は、この学習フレームワークが模倣における隠されたギャップを生じさせ、顕著な性能劣化につながる可能性がある主張する。図5に示すように、一般にエキスパート軌道として知られる記録された軌道は、模倣ベースのプランナーの学習中にグランドトゥールースとして機能する。生成された模倣された軌道は、その後、下流のトラックと基礎となるシステムダイナミクスによって処理され、AVの最終的な軌道が得られる。とはいえ、訓練中のトラックとダイナミクスに関する知識不足のため、模倣が完璧であっても、実際の軌跡は記録された軌跡から大きくずれる可能性がある。この主張は、表IVに示した実験結果を支持するものである。注目すべきは、Test14hardにおけるLog-replay + LQR法のNR-CLSは、完全追跡と比較して5.65の大幅な減少を示していることである。

RLアダプタ。一つの可能な解決策は、軌跡点ではなく、制御コマンドを直接模倣することである。とはいえ、このアプローチは特定の車両モデルに大きく依存するため、軌跡ベースの手法よりも一般化しやすく、解釈しにくい。別のアプローチとして、微分可能な運動モデルを軌跡デコーダに組み込む方法がある[5], [51]。しかし、微分可能性を確保するために、運動学的モデルが単純化されすぎていることが多い。この課題に取り組むために、このギャップを埋めるために設計された強化学習ベースの軌道アダプタ(RL Adapter)を導入する。RLアダプタは、模倣された軌道を、基礎となるダイナミクスを考慮しながら、関連する制御コマンドに変換する。その利点は2つある。第一に、プランナーを再トレーニングすることなく、様々な車両モデルに適応できることである。第二に、車両モデルに制約を課さず、微分不可能な車両モデル(例えば、高忠実度の実車両ダイナミクスモデル)と互換性を保つ。アダプタの学習過程を図5に、報酬を表Vに示す。ポリシーの最適化にはPP0[52]を使用し、学習率は $1e-3$ で80Kステップで学習を終了する。表IVに示すように、RLアダプタは完全追跡と同様の性能を示し、模倣ギャップを埋める能力を強調している。プランナーの学習プロセスに組み込むことができることに気づき、今後の課題として残す。

### IV. C IV. 最先端技術との比較

実装の詳細。我々の発見を統合して、我々は、以下のような完全学習ベースのベースライン計画モデルを提案する。

| Planners            |                  | Test14-random |              |              | Test14-hard  |              |              |          |
|---------------------|------------------|---------------|--------------|--------------|--------------|--------------|--------------|----------|
| Type                | Method           | OLS           | NR-CLS       | R-CLS        | OLS          | NR-CLS       | R-CLS        | Time(ms) |
| Expert              | Log-replay       | 100.0         | 94.03        | 75.86        | 100.0        | 85.96        | 68.80        | -        |
| Rule-based          | IDM [44]         | 34.15         | 70.39        | 72.42        | 20.07        | 56.16        | 62.26        | 32       |
|                     | PDM-Closed [48]  | 46.32         | 90.05        | <b>91.64</b> | 26.43        | 65.07        | 75.18        | 140      |
| Hybrid <sup>†</sup> | GameFormer [10]  | 79.35         | 80.80        | 79.31        | 75.27        | 66.59        | 68.83        | 443      |
|                     | PDM-Hybrid [48]  | 82.21         | <b>90.20</b> | 91.56        | 73.81        | 65.95        | <b>75.79</b> | 152      |
| Learning-based      | RasterModel [13] | 62.93         | 69.66        | 67.54        | 52.4         | 49.47        | 52.16        | 82       |
|                     | UrbanDriver [7]  | 82.44         | 63.27        | 61.02        | 76.9         | 51.54        | 49.07        | 124      |
|                     | GC-PGP [50]      | 77.33         | 55.99        | 51.39        | 73.78        | 43.22        | 39.63        | 160      |
|                     | PDM-Open [48]    | 84.14         | 52.80        | 57.23        | 79.06        | 33.51        | 35.83        | 101      |
|                     | PlanTF (Ours)    | <b>87.07</b>  | 86.48        | 80.59        | <b>83.32</b> | <b>72.68</b> | 61.7         | 155      |

表 VI. 最先端技術との比較。実行時間には、Pythonコードに基づく特徴抽出とモデル推論が含まれる。

<sup>†</sup> indicates these methods' final output trajectory relies on rule-based strategies or post-optimization.

プランニングトランスフォーマー(PlanTF)。具体的には、状態注意ドロップアウトエンコーダを組み込んだ状態6モデルを採用し、ドロップアウト率は0.75とする。学習中、0.5の確率で状態摂動を適用する。パッチサイズ128、重み減衰 $1e-4$ 、25エポックでモデルを学習する。初期学習率は $1e-3$ に設定され、コサイン的にゼロに減衰する。

方法 PlanTFの性能をいくつかの最先端のプランナーと比較する。(1) RasterModelは[13]で提供されたCNNベースのプランナである。(2) UrbanDriver[7]は、PointNetベースのポリラインエンコーダとTransformerに基づくベクトル化プランナである。ここでは、オープンループの再実装を使用し、履歴摂動は学習時に採用する。(3) GameFormer [10]は、レベルkゲームに基づくDETRのような対話型予測・計画フレームワークであり、最終的な軌道を生成するためのポストオブティマイザを組み込んでいる。(4) PDM\* [48]は2023年nuPlan計画チャレンジの優勝ソリューションである。PDM-Closedは、IDM[44]を異なるハイパーパラメータでアンサンブルする、純粋にルールベースのアプローチである。PDM-Hybridは、オープンループ予測性能を向上させるためにオフセット予測器を追加したPDM-closedの変形である。PDM-OpenはIDMベースのプランナを用いない純粋な学習コンポーネントである。結果は、公開されているコードを使用し、標準的な1Mデータ分割で学習させたものである。

結果 Test14-randomとTest14-hardベンチマークの比較結果を表VIに示す。まず、提案するPlanTFは、全てのメトリクスにおいて、他の全ての純粋な模倣ベースの手法を、特に閉ループ性能の点で大幅に上回る。また、広く認知されているIDMを凌駕する唯一の学習ベースの手法であり、ILにおける適切な設計の重要性を強調している。第二に、ルールベースやハイブリッド手法と比較した場合、PlanTFは、トリッキーな手作りのルールや戦略を必要とせず、高い競争力を持つCLSを維持しながら、優れたOLSを実現する。

注目すべきは、我々のアプローチがTest14-hardベンチマークで最高のNR-CLSを達成したことで、ルールベースの手法は通常のシナリオ(Test14-random)では良好な性能を発揮するが、ロングテールの状況(Test14-hard)では汎化に苦勞していることがわかる。一方、PlanTFはより強力な汎化能力を示す。

## V. CONCLUSION

本研究では、標準化されたnuPlanベンチマークを利用することで、模倣ベースのプランナのいくつかの重要な設計側面を系統的に検証する。我々の発見は、破局的ショートカット学習は、一般的に、過去の運動や単一フレームの運動状態のような入力特徴に対して起こることを明らかにした。これは、AVの現在位置のみに基づいて計画を立てることで、クローズドループの性能が向上するという予想外の結果につながる。この問題を軽減するために、ショートカット学習問題に効果的に対処する、素直な注意ベースの状態脱落エンコーダ(SDE)を導入する。SDEの実装により、state6モデルは総合的に最高の性能を達成した。データ補強は模倣ベースのプランナーにおけるもう一つの重要な要素である。我々の結果は、摂動は複合誤差を減らすために不可欠であるが、適切な特徴正規化によってのみ有効であることを示している。さらに、摂動を受けた場合でも、元のエキスパート軌道は信頼できるトレーニングのグランドトゥルースのままであることが観察される。これらの知見に加え、我々は、プランナーの性能にかなりの影響を与える、基礎となるシステムダイナミクスに対するモデルの認識不足によって引き起こされる、無視された模倣ギャップを特定した。この問題を解決するために、我々は強化学習ベースのアダプタを提案する。我々の発見を取り入れることで、提案する純粋に学習ベースのベースラインモデルPlanTFは、最先端のアプローチと比較して印象的な性能を示し、複雑なルールベースの戦略やポスト最適化を採用する手法と同等である。このことは、模倣学習ベースのプランナーにとって、適切な設計選択が重要であることを浮き彫りにしている。

限界と今後の課題。純粋な模倣ベースのプランナの境界を押し広げるにもかかわらず、我々の方法は、オープンループ学習とクローズドループテストの間の基

本的なミスマッチによって制約を受ける。閉ループ情報とシステムダイナミクスをトレーニングプロセスに組み込むことが、今後の研究の方向性を構成する。

## APPENDIX

Val14ベンチマークでの追加結果。Val14[48]ベンチマークでの比較結果(表VII)を示す。Val14には、14種類のシナリオからなる1180のシナリオが含まれている。

| Method          | OLS   | NR-CLS | R-CLS |
|-----------------|-------|--------|-------|
| Log-replay      | 100   | 94     | 80    |
| IDM [44]        | 38    | 77     | 76    |
| GC-PGP [50]     | 82    | 57     | 54    |
| PlanCNN [13]    | 64    | 73     | 72    |
| PDM-Hybrid [48] | 84    | 93     | 92    |
| PlanTF (Ours)   | 89.18 | 84.83  | 76.78 |

表 VII. Val14ベンチマークにおけるSOTAとの比較。他の手法の結果は[48]から引用した。

状態ドロップアウト率に関するアブレーション。表VIIIは、state6+SD Eモデルのドロップアウト率を変えた場合のアブレーション研究である。

| Model  | dropout | OLS   | NR-CLS | R-CLS |
|--------|---------|-------|--------|-------|
| state6 | -       | 88.33 | 77.28  | 74.10 |
|        | 0.25    | 89.11 | 81.70  | 78.44 |
|        | 0.50    | 89.12 | 83.71  | 77.52 |
|        | 0.75    | 87.07 | 86.48  | 80.59 |

表 VIII. SDEの状態脱落率に関するアブレーション研究。

## REFERENCES

- [1] L. Chen, P. Wu, K. Chitta, B. Jaeger, A. Geiger, and H. Li, "End-to-end autonomous driving: Challenges and frontiers," *arXiv preprint arXiv:2306.16927*, 2023.
- [2] S. Hagedorn, M. Hallgarten, M. Stoll, and A. Condurache, "Rethinking integration of prediction and planning in deep learning-based automated driving systems: A review," *arXiv preprint arXiv:2308.05731*, 2023.
- [3] J. Cheng, Y. Chen, Q. Zhang, L. Gan, C. Liu, and M. Liu, "Real-time trajectory planning for autonomous driving with gaussian process and incremental refinement," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 8999–9005.
- [4] 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.
- [5] 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)*. IEEE, 2021, pp. 1450–1457.
- [6] M. Vitelli, Y. Chang, Y. Ye, A. Ferreira, M. Wolczyk, B. Osinski, M. Niendorf, H. Grimmer, 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)*. IEEE, 2022, pp. 897–904.
- [7] O. Scheel, L. Bergamini, M. Wolczyk, B. Osinski, and P. Ondruska, "Urban driver: Learning to drive from real-world demonstrations using policy gradients," in *Conference on Robot Learning*. PMLR, 2022, pp. 718–728.
- [8] J. Cheng, R. Xin, S. Wang, and M. Liu, "Mnp: Multi-policy neural planner for urban driving," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 10 549–10 554.
- [9] 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," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 10 069–10 075.
- [10] Z. Huang, H. Liu, and C. Lv, "Gameformer: Game-theoretic modeling and learning of transformer-based interactive prediction and planning for autonomous driving," *arXiv preprint arXiv:2303.05760*, 2023.
- [11] 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*, 2023.
- [12] K. Guo, W. Jing, J. Chen, and J. Pan, "Ccil: Context-conditioned imitation learning for urban driving," *arXiv preprint arXiv:2305.02649*, 2023.
- [13] 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," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2021.
- [14] U. Muller, J. Ben, E. Cosatto, B. Flepp, and Y. Cun, "Off-road obstacle avoidance through end-to-end learning," *Advances in neural information processing systems*, vol. 18, 2005.
- [15] D. Wang, C. Devin, Q.-Z. Cai, P. Krähenbühl, and T. Darrell, "Monocular plan view networks for autonomous driving," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 2876–2883.
- [16] 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.
- [17] 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*. JMLR Workshop and Conference Proceedings, 2011, pp. 627–635.
- [18] W. Zeng, W. Luo, S. Suo, A. Sadat, B. Yang, S. Casas, and R. Urtaun, "End-to-end interpretable neural motion planner," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8660–8669.
- [19] S. Hu, L. Chen, P. Wu, H. Li, J. Yan, and D. Tao, "St-p3: End-to-end vision-based autonomous driving via spatial-temporal feature learning," in *European Conference on Computer Vision*. Springer, 2022, pp. 533–549.
- [20] 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*, 2023, pp. 17 853–17 862.
- [21] B. Jiang, S. Chen, Q. Xu, B. Liao, J. Chen, H. Zhou, Q. Zhang, W. Liu, C. Huang, and X. Wang, "Vad: Vectorized scene representation for efficient autonomous driving," *arXiv preprint arXiv:2303.12077*, 2023.
- [22] F. Codevilla, E. Santana, A. M. López, and A. Gaidon, "Exploring the limitations of behavior cloning for autonomous driving," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9329–9338.
- [23] D. Chen, B. Zhou, V. Koltun, and P. Krähenbühl, "Learning by cheating," in *Conference on Robot Learning*. PMLR, 2020, pp. 66–75.
- [24] D. Chen and P. Krähenbühl, "Learning from all vehicles," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 222–17 231.
- [25] K. Chitta, A. Prakash, and A. Geiger, "Neat: Neural attention fields for end-to-end autonomous driving," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15 793–15 803.
- [26] K. Chitta, A. Prakash, B. Jaeger, Z. Yu, K. Renz, and A. Geiger, "Transfuser: Imitation with transformer-based sensor fusion for au-



- tonomous driving,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.
- [27] Q. Zhang, M. Tang, R. Geng, F. Chen, R. Xin, and L. Wang, “Mmfn: Multi-modal-fusion-net for end-to-end driving,” in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 8638–8643.
  - [28] H. Shao, L. Wang, R. Chen, H. Li, and Y. Liu, “Safety-enhanced autonomous driving using interpretable sensor fusion transformer,” in *Conference on Robot Learning*. PMLR, 2023, pp. 726–737.
  - [29] X. Jia, P. Wu, L. Chen, J. Xie, C. He, J. Yan, and H. Li, “Think twice before driving: Towards scalable decoders for end-to-end autonomous driving,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 21 983–21 994.
  - [30] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, “Carla: An open urban driving simulator,” in *Conference on robot learning*. PMLR, 2017, pp. 1–16.
  - [31] M. Igl, D. Kim, A. Kuefler, P. Mougin, P. Shah, K. Shiarlis, D. Anguelov, M. Palatucci, B. White, and S. Whiteson, “Symphony: Learning realistic and diverse agents for autonomous driving simulation,” in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 2445–2451.
  - [32] Q. Li, Z. Peng, L. Feng, C. Duan, W. Mo, B. Zhou, *et al.*, “Scenarionet: Open-source platform for large-scale traffic scenario simulation and modeling,” *arXiv preprint arXiv:2306.12241*, 2023.
  - [33] A. Amini, T.-H. Wang, I. Gilitschenski, W. Schwarting, Z. Liu, S. Han, S. Karaman, and D. Rus, “Vista 2.0: An open, data-driven simulator for multimodal sensing and policy learning for autonomous vehicles,” in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022.
  - [34] Z. Wu, T. Liu, L. Luo, Z. Zhong, J. Chen, H. Xiao, C. Hou, H. Lou, Y. Chen, R. Yang, Y. Huang, X. Ye, Z. Yan, Y. Shi, Y. Liao, and H. Zhao, “Mars: An instance-aware, modular and realistic simulator for autonomous driving,” *CICAI*, 2023.
  - [35] Y. Hu, K. Li, P. Liang, J. Qian, Z. Yang, H. Zhang, W. Shao, Z. Ding, W. Xu, and Q. Liu, “Imitation with spatial-temporal heatmap: 2nd place solution for nuplan challenge,” *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2023.
  - [36] K. Renz, K. Chitta, O.-B. Mercea, A. S. Koepke, Z. Akata, and A. Geiger, “Plant: Explainable planning transformers via object-level representations,” in *Conference on Robot Learning*. PMLR, 2023, pp. 459–470.
  - [37] W. Xi, L. Shi, and G. Cao, “An imitation learning method with data augmentation and post processing for planning in autonomous driving,” *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2023.
  - [38] Y. Lu, J. Fu, G. Tucker, X. Pan, E. Bronstein, B. Roelofs, B. Sapp, B. White, A. Faust, S. Whiteson, *et al.*, “Imitation is not enough: Robustifying imitation with reinforcement learning for challenging driving scenarios,” *NeurIPS 2022 Workshop on Machine Learning for Autonomous Driving*, 2022.
  - [39] H. Liu, Z. Huang, J. Wu, and C. Lv, “Improved deep reinforcement learning with expert demonstrations for urban autonomous driving,” in *2022 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2022, pp. 921–928.
  - [40] Z. Huang, J. Wu, and C. Lv, “Efficient deep reinforcement learning with imitative expert priors for autonomous driving,” *IEEE Transactions on Neural Networks and Learning Systems*, 2022.
  - [41] N. Baram, O. Anschel, I. Caspi, and S. Mannor, “End-to-end differentiable adversarial imitation learning,” in *International Conference on Machine Learning*. PMLR, 2017, pp. 390–399.
  - [42] E. Bronstein, M. Palatucci, D. Notz, B. White, A. Kuefler, Y. Lu, S. Paul, P. Nikdel, P. Mougin, H. Chen, *et al.*, “Hierarchical model-based imitation learning for planning in autonomous driving,” in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 8652–8659.
  - [43] G. C. K. Couto and E. A. Antonelo, “Hierarchical generative adversarial imitation learning with mid-level input generation for autonomous driving on urban environments,” *arXiv preprint arXiv:2302.04823*, 2023.
  - [44] M. Treiber, A. Hennecke, and D. Helbing, “Congested traffic states in empirical observations and microscopic simulations,” *Physical review E*, vol. 62, no. 2, p. 1805, 2000.
  - [45] Motional. nuplan metrics. [Online]. Available: <https://nuplan-devkit.readthedocs.io/en/latest/metrics.description.html>
  - [46] J. Cheng, X. Mei, and M. Liu, “Forecast-MAE: Self-supervised pre-training for motion forecasting with masked autoencoders,” *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023.
  - [47] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.
  - [48] D. Dauner, M. Hallgarten, A. Geiger, and K. Chitta, “Parting with misconceptions about learning-based vehicle motion planning,” in *Conference on Robot Learning (CoRL)*, 2023.
  - [49] 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.
  - [50] M. Hallgarten, M. Stoll, and A. Zell, “From prediction to planning with goal conditioned lane graph traversals,” *arXiv preprint arXiv:2302.07753*, 2023.
  - [51] H. Cui, T. Nguyen, F.-C. Chou, T.-H. Lin, J. Schneider, D. Bradley, and N. Djuric, “Deep kinematic models for kinematically feasible vehicle trajectory predictions,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 10 563–10 569.
  - [52] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” *arXiv preprint arXiv:1707.06347*, 2017.