

ParkPredict: 駐車場における車両の動きと意図の予測

Xu Shen^{1,*}, Ivo Batkovic^{2,3,*}, Vijay Govindarajan^{1,*}, Paolo Falcone², Trevor Darrell¹, and Francesco Borrelli¹

概要 駐車場におけるドライバーの行動を予測する問題を調査する。駐車場は、典型的な道路網よりも構造化されておらず、コンパクトな空間における複雑でインタラクティブな操縦を特徴とする環境である。CARLAシミュレータを用いて駐車場環境を開発し、人間の駐車操作のデータセットを収集する。次に、マルチモーダル長短期記憶(LSTM)予測モデルと畳み込みニューラルネットワーク(LSTM-CNN)を物理ベースの拡張カルマンフィルタ(EKF)ベースラインと比較することで、モデルの複雑さと特徴情報の影響を研究する。我々の結果は、1)意図はうまく推定できること(LSTMとCNN-LSTMモデルでトップ1精度はおおよそ85%、トップ3精度はほぼ100%)、2)人間のドライバーの意図する駐車場に関する知識は駐車軌跡の予測に大きな影響を与える、3)環境の意味表現は長期予測を改善することを示している。

I. INTRODUCTION

自律走行技術は飛躍的に進歩したが、自律走行車(AV)はまだ大きな課題に直面している。ロバストな知覚[1]、予測[2]、他の参加者との実際の交通シナリオにおけるインタラクション[3]、[4]、特に人間主導の車両は難しい。他の推定行動によっては、AVの行動は保守的すぎたり、攻撃的すぎたり、最悪の場合安全でなかったりする。

この問題は、駐車場のようなコンパクトで構造化されていない領域では特に困難であり、近接した人間エージェントとの多数の相互作用 [5] を特徴とし、最適でないポリシー [6] との混雑につながる可能性がある。このような環境において、センシングと通信インフラを装備する可能性は、コネクテッドAVのアルゴリズムによるソリューションを可能にするのに役立つかもしれない[7]。

車両運動予測の問題を調査するために広範な研究が行われ、様々なレベルのアクセス可能な情報に基づいてアルゴリズムが開発されている。

物理ベースの手法は、直感的な短期予測を提供するためにポーズ履歴を使用し[8]、カルマンフィルタ[9]は、車両ダイナミクスを時間的に前方に伝播し、軌道または到達可能なセットを予測するために使用される[10]。一方、リカレントニューラルネットワーク(RNN)、特にLSTM(Long Short-Term Memory)ネットワークのようなデータ駆動型手法は、車両パラメータや入力プロファイルを明示的に知ることなく、運動モデルを学習することができる。このエンコーダーである。

This work is partly supported by Berkeley DeepDrive (BDD) and the Wallenberg Artificial Intelligence, Autonomous Systems and Software Program (WASP) funded by Knut and Alice Wallenberg Foundation.

* Indicates equal contribution.

¹ University of California, Berkeley, CA, USA ({xu.shen, govindarajan, fborelli, trevor}@berkeley.edu).

² Chalmers University of Technology, Gothenburg, Sweden ({ivo.batkovic, falcone}@chalmers.se).

³ Zenuity AB, Gothenburg, Sweden (ivo.batkovic@zenuity.com).

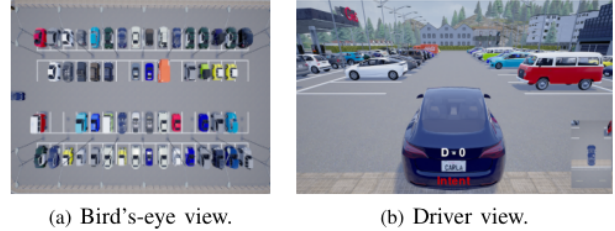


Fig. 1. Parking lot scenario.

LSTMはポーズ履歴を処理し、このシーケンスの要約を生成し、予測のためにデコーダに渡される[11], [12]。

車両の運動は、旋回、車線維持、車線変更などの意図に直接影響されるため、意図情報を運動情報に埋め込むことで、予測性能を向上させることができる[13]。LSTMエンコーダは、意図を予測するためにモーションデータの表現を学習し、解釈可能なマルチモーダル予測を生成する[14]。高速道路や都市走行では、通常、車線グラフに基づいて意図が分類される[15]、[16]、[17]。

軌跡や環境構造を明示的に処理する必要がないため、畳み込みニューラルネットワーク(CNN)は、LSTMと組み合わせて、画像シーケンスから情報を合成し予測を行うために使用されてきた[18]。最近の研究では、これらのアーキテクチャは、車両形状、ダイナミクス[20]、マルチモーダル運動予測のための衝突回避制約[21]など、鳥瞰図(BEV)意味画像[19]に適用されている。

上述した既存の研究のほとんどは、構造化された環境に焦点を当てたものである。例えば、交差点、車線、信号機からなる、よく定義された道路網[22]が存在する。しかし、駐車場などの環境では、次のような課題が生じる。

- 1) 駐車場内の人間運転車両に関する公開データセットは限られている[23];
- 2) 駐車場の操縦は、高速道路走行よりも一般的に複雑で[24]、困難である;
- 3) コンパクトな空間と周囲の物体の近接性により、衝突や混雑のリスクが高まる。

本研究では、特徴量とモデル抽象度が与えられたとき、人間のドライバーの意図する駐車場と将来の軌跡を予測する問題に焦点を当てる。本論文は以下の貢献を提供する:

- 1) CARLAシミュレータとレーシングホイールインタフェースを用いた駐車場シミュレーション環境を開発する;

- 2) 前方駐車と後方駐車両方の操作を含む人間の駐車行動の注釈付きデータセットと、様々な駐車スポットの選択を収集する；
- 3) 我々は、意図的な軌道予測とマルチモーダルな軌道予測の両方を提供できる階層的なLSTMモデル構造を提案する；
- 4) 駐車場の形状を捉えたセマンティックBEV画像に視覚エンコーダを適用したネストされたCNN-LSTMモデル構造を提案する；
- 5) 物理学に基づくカルマンフィルタのベースラインと、より複雑度の高いLSTMやCNN-LSTMモデルを比較し、モデルの複雑度、特徴の複雑度、アクセス可能な情報量が予測性能に与える影響を調査する。

本稿の構成は以下の通りである。セクションIIでは、実験デザインとデータセット生成について述べる。セクションIIIでは、意図の分類と動き予測のために設計されたアルゴリズムについて詳しく説明する。セクションIVでは、予測アルゴリズムの結果について述べる。最後に、セクションVでは、我々の主要な発見と今後の研究のためのアイデアを要約する。

II. 実験デザインとデータセット

本節では、駐車場シミュレーション環境と実験セットアップの概要を説明する。我々は、軌跡、最終的な駐車スポットの位置、および信号の意図からなる、人間の運転デモからデータセットを生成した。このデータセットは、インテントとモーションの予測に使用された。

A. 駐車場シナリオ

駐車場のデモを制御して収集するために、Town04から改良したカスタム駐車場マップを備えたCARLAシミュレータとCARLA ROSブリッジ[25]を使用した。駐車場は4列からなり、各16スポットである。各試験において、静的車両は、中央の2列に位置する8つのフリースポットオプションのみが利用できるように、駐車場に産卵された。これらのフリースポットの具体的な位置は、被験者から多様な駐車場のデモを収集するために、試行ごとに变化させた。フリースポット構成の例を図1(a)に示す。

このような設定のもと、被験者は、指定された順方向または逆方向の駐車操作の後、選択したフリースポットに駐車するよう指示された。被験者が駐車場を選択するとき、ボタンを押して意図を判断させるように指示された。被験者は、ロジクールG27レーシングホイールを使用して、自車両のブレーキ、スロットル、ステアリングを制御した。この実験では、エゴ車両のみが移動し、シーン内の他の車両はすべて静止したまま駐車した。図1(b)にドライバビューを可視化する。

合計10人の被験者が30回の順方向駐車と30回の逆方向駐車の実験を行い、合計600回のデモンストレーションを行った。各デモンストレーションでは、エゴ・ビークルの運動状態履歴と、駐車場がいつ選択されたかを知るための意図信号が記録された。さらに、各駐車場の位置を収集した。衝突を含む、または意図的なシグナルを含まないデモはすべてフィルタリングされた。

さらに、駐車場の構成は、エゴ車両と他のすべての駐車車両のバウンディングボックスとともに記録された。

B. Dataset Generation

まず、各デモに含まれるデータに対して使用されるいくつかの表記法を紹介する。

- 時刻 t における車両の姿勢は、 $\sim z(t) = x(t) \ y(t) \ \theta(t)^T$ で表される。車両は xy 平面上にあると仮定し、高度、ピッチ、ロールの変動を無視する。時刻 t における駐車場占
- 有率を次のように表す。

$$\mathcal{O}(t) = \begin{bmatrix} x_{s,1}(t) & y_{s,1}(t) & f_{s,1}(t) \\ \vdots & \vdots & \vdots \\ x_{s,G}(t) & y_{s,G}(t) & f_{s,G}(t) \end{bmatrix},$$

であり、サイズ $G \times 3$ の行列である。ここで、 G は駐車場の数である。各行 j において、最初の2つのエントリ $x_{s,j}(t)$, $y_{s,j}(t)$ はスポットの位置であり、最後のエントリ $f_{s,j}(t)$ はスポットが自由であれば1に設定されるバイナリ変数である。

- ドライバーの意図のグラントゥールス分布は、サイズ $G+1$ のワンホットベクトルである $g(t)$ で示される。 $(G+1)$ 番目の要素は未確定の意図に対応する。意図が示された後、 $g(t)$ は被験者が最終的に駐車する場所に対応する。
- 駐車場鳥瞰図 $I(t)$ は、図 1(a)の駐車場環境を表す $W \times 3$ の形状 H の意味画像である。3つのチャンネルは、それぞれ駐車場マーク、静止車両バウンディングボックス、エゴ車両バウンディングボックスに対応する。

各デモンストレーションにおいて、図2に示すように、被験者が運転を開始する前と駐車した後の時間間隔を削除した。タイムステップ $\Delta t = 0.1s$ が与えられると、残りの部分のデモは履歴ホライズン $N_{hist} = 5$ 、予測ホライズン $N_{pred} = 20^1$ の短いスニペットに処理された。各スニペットはさらに処理され、以下の特徴が生成された：

- 1) motion history of $\bar{z}(t)$: $\mathcal{Z}_{hist}(t) \in \mathbb{R}^{N_{hist} \times 3}$;
- 2) parking spot and occupancy $\mathcal{O}(t) \in \mathbb{R}^{G \times 3}$;
- 3) image history of $I(t)$: $\mathcal{I}_{hist}(t) \in \mathbb{R}^{N_{hist} \times H \times W \times 3}$;

and labels:

- 1) $z(t)$ の未来運動: $\mathcal{Z}_{future}(t) \in \mathbb{R}_R^{N_{pred} \times 3}$ グラ
- 2) ウンドトゥールスドライバの意図: $g(t)$.

すべての履歴特徴は、水平線 N_{pred} 、ステップサイズ Δt で、 t から時間的に逆サンプリングされることに注意。同様に、予測ラベルは水平線 N_{future} 、ステップサイズ Δt で t から時間的に前方にサンプリングされる。

この処理の結果、データセット $D = \{(\mathcal{Z}_{hist}^{(i)}, 0^{(i)}, \mathcal{I}_{hist}^{(i)}), (\mathcal{Z}_{future}^{(i)}, g^{(i)})\}_{i=1}^M$ 。ここで、上付き文字 (i) は i 番目のデータセットインスタンスに対応し、インスタンスの総数は $M = 20850$ である。最初のタプルは特徴量に対応し、2番目のタプルはラベルに対応し、各エントリの時間依存性は今後削除する。

処理されたデータセットだけでなく、詳細は <https://bit.ly/parkpredict> にある。

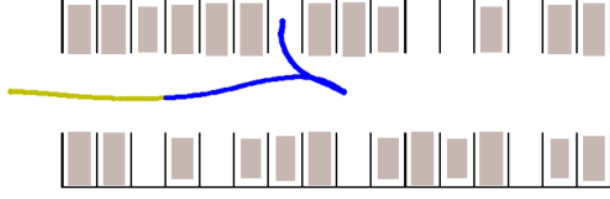


図2. 図2. 中央の2行に焦点を当てたデモの例。網掛けのバウンディングボックスは駐車車両を表す。当初、人間のドライバーは未確定な意図を持っています(黄色の部分)。そして、ドライバーはスポットインテントを決定し、軌跡は青で示される。

III. METHODOLOGY

データセットDが与えられたとき、予測問題を意図と軌跡推定のサブ問題に分割する。入力履歴 $X_{hist}^{(i)}$ と占有率 $0^{(i)}$ を用いて、予測されたドライバーの意図分布 $g^{\{i\}}$ と予測された将来の N_{pred} -ステップ軌跡 $Z^{\{i\}}$ を生成する。モデルと特徴の複雑さの両方が異なる3つのモデルを評価する: EKFベースライン、LSTMネットワーク、CNN-LSTMネットワーク。以下の節では、各モデルに提供されるモデル構造と入力(i)履歴特徴 X_{hist} について説明する。

A. 速度一定での拡張カルマンフィルタ(EKF)

ベースラインとして、 $X_{hist}^{(i)} = (Z_{hist}^{(i)})$ とする等速を仮定したEKFを用いる。以下の状態ダイナミクスと測定モデルを使用する。

1) State Dynamics:

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \\ \theta_{k+1} \\ v_{k+1} \\ \omega_{k+1} \end{bmatrix} = \begin{bmatrix} x_k + v_k \cos(\theta_k) \Delta t \\ y_k + v_k \sin(\theta_k) \Delta t \\ \theta_k + \omega_k \Delta t \\ v_k \\ \omega_k \end{bmatrix} + q_k, \quad q_k \sim \mathcal{N}(0, Q)$$

2) Measurement Model:

$$\hat{z}_k = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_k \\ y_k \\ \theta_k \\ v_k \\ \omega_k \end{bmatrix} + r_k, \quad r_k \sim \mathcal{N}(0, R)$$

外乱共分散Qは、各デモンストレーションで提供されたポーズと速度の情報を用いて推定された。ノイズの共分散は、グランドトゥールースのポーズ測定を反映するために、 $R = \text{diag}(1e-3, 1e-3, 1e-3)$ として選択された。まず、姿勢履歴 Z_{hist} に対して時間更新と計測更新(i)を行い、時間更新によって Z_{future} を外挿する。次に、 $g^{\{i\}}$ は、ユークリッド距離の逆数に基づいて、フリースポットに確率を割り当てる(すなわち、 $0^{(i)}$ を用いる)ことで、最終的な予測位置を推定する。あるスポットが距離閾値(20m)を超えた場合、対応する確率が未確定カテゴリである $g^{\{i\}}$ の(G + 1)番目の要素に追加される。

B. 長短期記憶ネットワーク(LSTM)

我々の提案するLSTMモデルを図3に示す。EKFと同様に、 $X_{hist}^{(i)} = (Z_{hist}^{(i)})$ であるが、意図と軌跡の推定は逆の順序で行われる。特に、EKFとは異なり、このモデルは以下の階層構造を持つマルチモーダルな意図条件付き軌道を予測する。

1) 意図予測モジュール: 意図予測モジュールは $X_{hist}^{(i)}$ と $0^{(i)}$ を入力とし、意図確率分布を推定する:

$$\hat{g}^{(i)} = \mathcal{F}_{\text{intent}}(X_{hist}^{(i)}, O^{(i)})$$

ポーズ履歴と駐車場占有率は、まずエンコーダLSTMスタックに渡され、次に完全接続層によって処理される。ソフトマックス出力層は予測される意図分布 $g^{\{i\}}$ を生成する。

ここで、 $j \in \{1, \dots, G+1\}$ はインテントインデックスを表す。

- 予測値 $g^{\{i\}}$ とグランドトゥールース $g^{(i)}$ の間のクロスエントロピーは、予測分布をグランドトゥールースラベルに近づける:

$$J_1^{\text{intent}} = - \sum_{j=1}^{G+1} g_j^{(i)} \log(\hat{g}_j^{(i)}).$$

- ドライバーの確率性を考慮するために、予測分布 $g^{\{i\}}$ の負のエントロピー:

$$J_2^{\text{intent}} = -\mathcal{H}(\hat{g}^{(i)}) = - \sum_{j=1}^{G+1} \hat{g}_j^{(i)} \log(\hat{g}_j^{(i)}).$$

- すでに占有されている駐車場を予測するペナルティ:

$$J_3^{\text{intent}} = \sum_{j=1}^G \max\{(\hat{g}_j^{(i)} - f_{s,j}^{(i)}), 0\}.$$

したがって、最終的な目的関数は

$$J^{\text{intent}} = J_1^{\text{intent}} + J_2^{\text{intent}} + J_3^{\text{intent}}.$$

2) 軌道予測モジュール: 軌道予測モジュールは $X_{hist}^{(i)}$ と意図インデックス $j \in \{1, \dots, G+1\}$ を入力とし、 N_{pred} タイムステップの将来の軌道を推定する:

$$\hat{Z}_{\text{future},j}^{(i)} = \mathcal{F}_{\text{traj}}(X_{hist}^{(i)}, j).$$

このモジュールは、マルチモーダル予測を生成するために、意図予測モジュールと順次動作する。エンコーダはまず(i) X_{hist} を処理する。エンコーダの最終的な隠れ状態とセルの状態は、デコーダの初期化に使用される。次に、各意図インデックスjに対して、デコーダとそれに続く完全接続(i)層は、確率 $g_j^{(i)}$ に関連する予測された未来の軌跡 $Z_{future,j}^{(i)}$ を返す。

しかし、学習時には、このモジュールを意図予測と切り離し、グランドトゥールースラベル $g^{(i)}$ のみを使用する:

$$\hat{Z}_{\text{future},gt}^{(i)} = \mathcal{F}_{\text{traj}}(X_{hist}^{(i)}, \underset{j}{\text{argmax}} g_j^{(i)}).$$

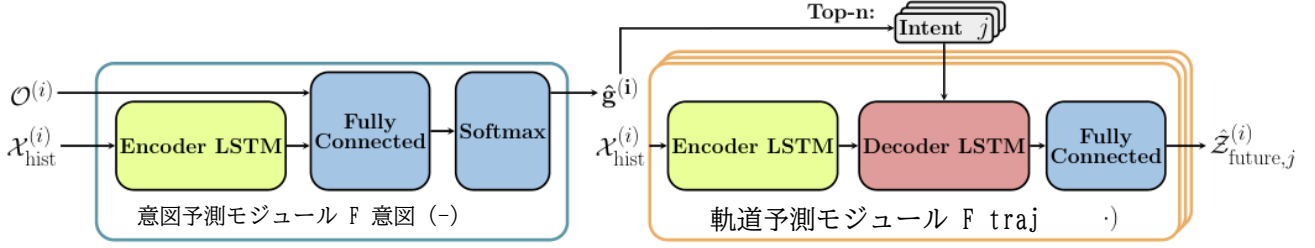
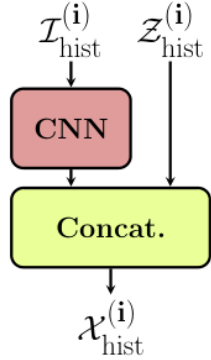


図3. マルチモーダルLSTM予測モデル



Type / Stride	Parameters
Conv / s1	$8 \times (3 \times 3)$
Max Pool / s2	Pool 2×2
Batch Norm	—
Conv / s2	$16 \times (3 \times 3)$
Max Pool / s2	Pool 2×2
Batch Norm	—
Flatten	—
Dropout	$p = 0.5$
Fully connect	128
Dropout	$p = 0.5$
Fully connect	16

図4. CNNのアーキテクチャ。右の表は左図のCNN前処理ブロックを記述している。

目的関数 J^{traj} は位置の平均二乗誤差(MSE)を用いて定義される。具体的には、 $Z_{\text{est}}^{(i)}$ と $Z_{\text{est}}^{(i)}$ のk行目をそれぞれ $\tilde{z}_k^{(i)}$ と $\hat{z}_k^{(i)}$ とすると、次のようになる：

$$J^{traj} = \frac{1}{N_{pred}} \sum_{k=1}^{N_{pred}} \sqrt{(\tilde{z}_k^{(i)} - \hat{z}_k^{(i)})^T \text{diag}(1, 1, 0) (\tilde{z}_k^{(i)} - \hat{z}_k^{(i)})}$$

C. 畳み込みニューラルネットワークLSTM(CNN-LSTM)

CNN-LSTMモデルはLSTMモデルと同様に図3と同じ構造を用いる。しかし、(i)図4に示すように、 \mathcal{I}_{hist} の各画像を前処理するCNNは1つである。生成された画像特徴量は、動き特徴量と連結される。したがって、CNN-LSTMモデルでは、 $\mathcal{X}_{hist}^{(i)} = (F_{CNN}(\mathcal{I}_{hist}^{(i)}), \mathcal{Z}_{hist}^{(i)})$ となり、 $F_{CNN}(-)$ はCNNが行う視覚符号化演算を表す。これは、動きと視覚の特徴をLSTM時間エンコーダで融合する[21]、[26]で使われたアプローチから着想を得ている。

D. 予測評価

各予測アルゴリズムの性能を比較するために、5重クロスバリデーションを使用し、LSTMとCNN-LSTMはバッチサイズ32で200エポック学習される。ここで変数 M_{302} は、評価される対応するホールドアウト集合のカーディナリティに対応する。

1) トップn精度：集合 $G_n(g^{(i)})$ が、予測される意図分布 $g^{(i)}$ に最も可能性の高いn個の意図カテゴリを含むとする。そして、トップnの精度は

$$A_n = \frac{1}{M} \sum_{i=1}^{\tilde{M}} \sum_{j=1}^{G+1} g_j^{(i)} \mathbb{I}(g_j^{(i)} \in G_n(\hat{g}^{(i)}))$$

ここで $I_{(-)}$ は指標関数である。

2) 平均距離誤差：予測された \hat{z}_{future} と実際の軌跡 z_{future} が与えられたとき、予測タイムステップの関数として位置誤差を見ることができる。タイムステップkにおける平均距離誤差 d_k は次の通りである：

$$d_k = \frac{1}{M} \sum_{i=1}^{\tilde{M}} \sqrt{(\tilde{z}_k^{(i)} - \hat{z}_k^{(i)})^T \text{diag}(1, 1, 0) (\tilde{z}_k^{(i)} - \hat{z}_k^{(i)})}$$

IV. RESULTS

本節では、選択した評価指標を用いて、各アルゴリズムの意図予測能力と軌跡予測誤差を比較する。情報量とマルチモーダリティが予測性能に与える影響について調査する。簡潔にするため、以降の図ではCNN-LSTMをCNNに短縮している。

A. 意図の分類

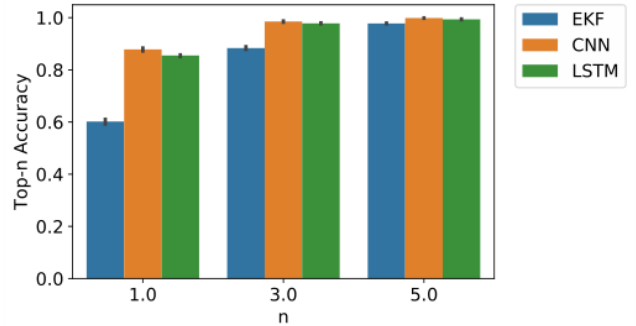


図5. トップnの意図分類精度、 A_n .

図5は、評価した全モデルにおけるトップn精度 A_n を示す。予想通り、LSTMとCNN-LSTMはすべてのnでEKFベースラインを上回り、およそ85%のトップ1精度とほぼ100%のトップ3またはトップ5精度を達成した。これは、LSTMとCNN-LSTMが意図予測タスクに特化して学習されるのに対し、EKFの軌跡予測ステップの後には発見的逆距離アプローチが使用されるという事実から導かれる。

B. 軌道予測

このセクションでは、以下のモデルと情報レベルを見ていく。 $\in \{\text{LSTM}, \text{CNN}\}$ ：

- ? no intent: 軌跡モジュール $F_{traj}(-)$ を再学習し、ゼロ
intent入力のみを適用するintent不可知モデル。
- gt intent: $F_{traj}(-)$ をintentモジュール $F_{intent}(-)$ から切り
離し、グラントゥルースのintent $g^{(i)}$ のみを用いて $\mathcal{L}_{\{future, gt\}^{(i)}}$ を予測するintent条件付きモデル。
- マルチモーダル: $F_{traj}(-)$ を $F_{intent}(-)$ から $g^{(i)}$
} の上位 n エントリに適用する、意図条件付きマルチモ
ーダルモデル。ここでは $n = 3$ を選ぶ。

図6は、各タイムステップ k における平均距離誤差 d_k に対する意図知識と情報レベルの影響を捉えたものである。CNN-LSTMとLSTMの両モデルにおいて、より微妙な動きモデルを学習するため、長期予測においてEKFを上回ることが観察された。LSTMモデルでは、意図なしモデルはグラントゥルースモデルと同等であるため、意図知識の利点は最小である。しかし、CNN-LSTMの場合、時間ステップ12の後に見られるように、意図の追加知識が予測を助ける。このことは、意味的なBEV画像が、モデルが予測のための意図ラベルをより良く理解するのに役立つことを示唆している。

図7は、各タイムステップ k における平均距離誤差 d_k に対するマルチモーダル予測の利点を示している。マルチモーダル予測では、上位 $n = 3$ 個のintentラベルを取り、3個の別々のロールアウトを生成することに注意。ここで報告する結果は、グラントゥルースの軌跡に最も近いロールアウトを見つけ(すなわち、位置の平均二乗誤差 J^{traj} を使用)、この単一のロールアウトを使用して d_k を評価することによって計算される。すべての意図条件付きモデルは、予測地平において、意図知識なしのモデルと同等か、それ以上の性能を発揮することに注意する。このことは、たとえ意図が正確に予測され、知られていなかったとしても、この追加情報は、意図にとらわれない予測に比べ、軌道予測誤差を減少させることができることを示している。これは、軌跡予測のマルチモーダルな性質が、ユニモーダルな軌跡予測に比べて、より多くの進化の可能性とドライバーモデルの確率性を捉えることができるためと考えられる。また、CNN-LSTMモデルは長期予測に適しており、駐車場の形状がより有益である可能性が高いことが観察された。

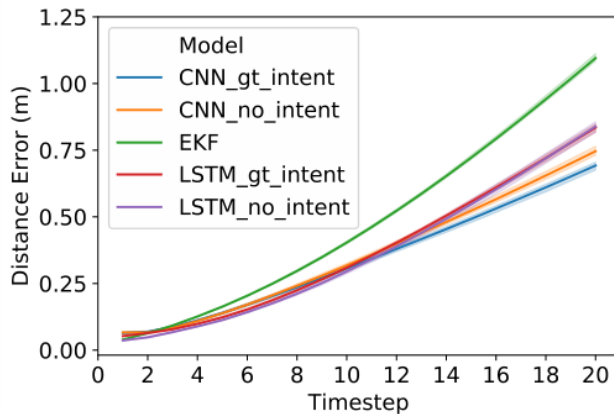


図6. 図6. モデルと意図知識のレベルを変化させた場合の平均距離誤差 d_k 対タイムステップ k 。

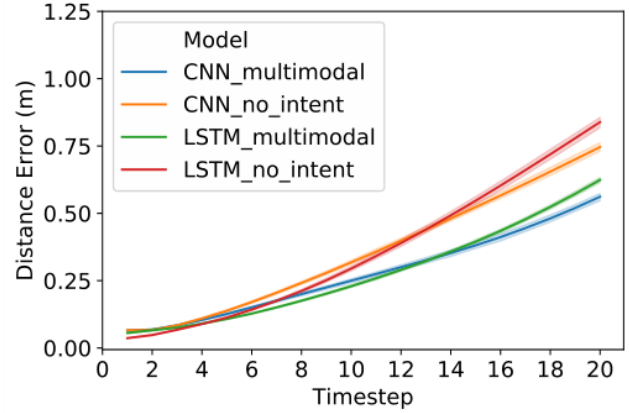


図7. 平均距離誤差 d_k 対タイムステップ k は、意図にとらわれないユニモーダル予測と、マルチモーダルな意図条件付き予測の中で最良のロールアウトを比較する。

C. シナリオ例

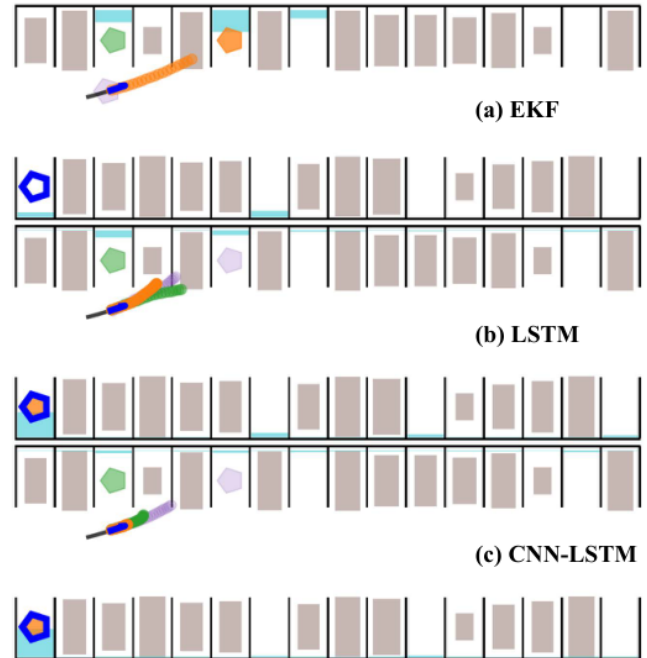


図8. モデル間の予測例。黒の曲線はポーズ履歴 $Z_{hist}^{(i)}$ を表し、青の五角形と曲線はそれぞれグラントゥルースの将来の意図 $g^{(i)}$ と動き $Z_{future}^{(i)}$ を表し、オレンジ、緑、紫の五角形と曲線は上位3つの意図と軌跡予測に対応する。これらの次数と透明度は確率 $g^{(i)}$ を反映しており、これはスポットのシアンバーで可視化されている。軌跡上に五角形がマークされている場合、それは未確定の意図に対応する。

図8は、予測アルゴリズムが予測インスタンスのサンプルに対してどのように比較されるかを示している。3つのサブ図は駐車場の2次元レイアウトを示しており、塗りつぶされた斜線ボックスは占有駐車場である。意図予測では、EKFの近接ベースのヒューリスティックは、車両の前方にある最も近いスポットを優先するが、

車両が後方にあるグラントゥールスのスポットに反転する可能性がある場合を見逃している。代わりに、LSTMとCNN-LSTMは上位3つの候補のグラントゥールスを捉えており、ドライバーが操作でバックアップすることを選択するかもしれないデータからも学習していることを意味する。

軌道予測では、EKFはダイナミクスを外挿するため、大きな遅延を伴う単一の軌道予測しか提供しない。LSTMとCNN-LSTMは、よりグラントゥールスに適合し、他の近傍スポットに対してマルチモーダルな振る舞いを提供する。LSTMとCNN-LSTMの上位3つの候補は、LSTMが占有情報をより良く活用し、CNN-LSTMが意味的な画像入力を通してより詳細な障害物表現を学習するため、障害物をより認識する。したがって、どちらのモデルも逆マヌーバをより重視する。

V. DISCUSSION

本研究では、特徴量とモデルの複雑さを変化させながら、人間の運転者の駐車意図と操縦を予測する問題を検討した。カスタムCARLAシミュレータ駐車場環境を構築し、人間の駐車場操作のデータセットを生成するために使用した。我々は、EKFベースラインに対して、意図条件付きLSTMとCNN-LSTM予測モデルを比較し、推定されたとしても、軌跡予測に意図情報を提供することの利点を指摘した。さらに、障害物や駐車場の形状を符号化することで、セマンティックBEV画像は長期的な予測性能の向上に役立つ。階層型フレームワークは、駐車場のような比較的非構造的な環境において、マルチモーダルなドライバー行動予測を提供することが可能である。

今後の課題として、[20]のように運動学的制約に応じたペナルティを組み込んで、CNN-LSTMモデルに用いられるコスト関数を調査したい。さらに、実世界の行動をより広く変化させるデータセットを拡張することで、マルチエージェント設定における自律駐車のための確率制御フレームワークにおけるマルチモーダル予測の適用を期待する。

REFERENCES

- [1] J. Janai, F. Güney, A. Behl, and A. Geiger, "Computer vision for autonomous vehicles: Problems, datasets and state-of-the-art," *arXiv preprint arXiv:1704.05519*, 2017.
- [2] S. Lefèvre, D. Vasquez, and C. Laugier, "A survey on motion prediction and risk assessment for intelligent vehicles," *ROBOMECH Journal*, vol. 1, no. 1, p. 1, dec 2014. [Online]. Available: <http://www.robomechjournal.com/content/1/1/1>
- [3] I. Batkovic, M. Zanon, M. Ali, and P. Falcone, "Real-time constrained trajectory planning and vehicle control for proactive autonomous driving with road users," in *2019 18th European Control Conference (ECC)*, June 2019, pp. 256–262.
- [4] K. Driggs-Campbell, V. Govindarajan, and R. Bajcsy, "Integrating intuitive driver models in autonomous planning for interactive maneuvers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 12, pp. 3461–3472, 2017.
- [5] A. Marshall, "Why Are Parking Lots So Tricky for Self-Driving Cars?" [Online]. Available: <https://www.wired.com/story/why-are-parking-lots-so-tricky-for-self-driving-cars/>
- [6] X. Shen, X. Zhang, and F. Borrelli, "Autonomous Parking of Vehicle Fleet in Tight Environments," 2019. [Online]. Available: <http://arxiv.org/abs/1910.02349>
- [7] J. Guanetti, Y. Kim, and F. Borrelli, "Control of connected and automated vehicles: State of the art and future challenges," *Annual Reviews in Control*, vol. 45, pp. 18–40, 2018.
- [8] A. Houenou, P. Bonnifait, V. Cherfaoui, and W. Yao, "Vehicle trajectory prediction based on motion model and maneuver recognition," in *2013 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2013, pp. 4363–4369.
- [9] R. Schubert, E. Richter, and G. Wanielik, "Comparison and evaluation of advanced motion models for vehicle tracking," in *2008 11th international conference on information fusion*. IEEE, 2008, pp. 1–6.
- [10] M. Althoff, O. Stursberg, and M. Buss, "Model-based probabilistic collision detection in autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 2, pp. 299–310, 2009.
- [11] S. H. Park, B. Kim, C. M. Kang, C. C. Chung, and J. W. Choi, "Sequence-to-Sequence Prediction of Vehicle Trajectory via LSTM Encoder-Decoder Architecture," *IEEE Intelligent Vehicles Symposium, Proceedings*, vol. 2018-June, no. Iv, pp. 1672–1678, 2018.
- [12] B. D. Kim, C. M. Kang, J. Kim, S. H. Lee, C. C. Chung, and J. W. Choi, "Probabilistic vehicle trajectory prediction over occupancy grid map via recurrent neural network," *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, vol. 2018-March, pp. 399–404, 2018.
- [13] Y. Hu, W. Zhan, and M. Tomizuka, "Probabilistic prediction of vehicle semantic intention and motion," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, June 2018, pp. 307–313.
- [14] N. Deo and M. M. Trivedi, "Multi-Modal Trajectory Prediction of Surrounding Vehicles with Maneuver based LSTMs," *IEEE Intelligent Vehicles Symposium, Proceedings*, vol. 2018-June, no. Iv, pp. 1179–1184, 2018.
- [15] T. Streubel and K. H. Hoffmann, "Prediction of driver intended path at intersections," *IEEE Intelligent Vehicles Symposium, Proceedings*, no. Iv, pp. 134–139, 2014.
- [16] J. Schulz, C. Hubmann, J. Lochner, and D. Burschka, "Interaction-Aware Probabilistic Behavior Prediction in Urban Environments," *IEEE International Conference on Intelligent Robots and Systems*, pp. 3999–4006, 2018.
- [17] T. Gindele, S. Brechtel, and R. Dillmann, "Learning driver behavior models from traffic observations for decision making and planning," *IEEE Intelligent Transportation Systems Magazine*, vol. 7, no. 1, pp. 69–79, 2015.
- [18] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2625–2634.
- [19] S. Casas, W. Luo, and R. Urtasun, "IntentNet: Learning to Predict Intention from Raw Sensor Data," *CoRL*, vol. 87, no. CoRL, pp. 947–956, 2018.
- [20] H. Cui, T. Nguyen, F.-C. Chou, T.-H. Lin, J. Schneider, D. Bradley, and N. Djuric, "Deep kinematic models for physically realistic prediction of vehicle trajectories," *arXiv preprint arXiv:1908.00219*, 2019.
- [21] 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," pp. 2090–2096, 2019.
- [22] C. Vallon, Z. Ercan, A. Carvalho, and F. Borrelli, "A machine learning approach for personalized autonomous lane change initiation and control," in *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2017, pp. 1590–1595.
- [23] D. A. Vasquez Govea, T. Fraichard, and C. Laugier, "Growing Hidden Markov Models: An Incremental Tool for Learning and Predicting Human and Vehicle Motion," *The International Journal of Robotics Research*, vol. 28, no. 11-12, pp. 1486–1506, Nov. 2009.
- [24] X. Zhang, A. Liniger, A. Sakai, and F. Borrelli, "Autonomous Parking Using Optimization-Based Collision Avoidance," in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, dec 2019, pp. 4327–4332.
- [25] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, 2017, pp. 1–16.
- [26] H. Xu, Y. Gao, F. Yu, and T. Darrell, "End-to-end learning of driving models from large-scale video datasets," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2174–2182.