自律走行車のための視覚と推測航法に基づくエンドツーエンドの駐車

Conference Paper · June 2018

DOI: 10.1109/IVS.2018.8500558

2 authors:

Vijay John
RIKEN Keihanna
77 PUBLICATIONS **1,514** CITATIONS
SEE PROFILE

Swarn singh Rathour
Toyota Technological Institute
14 PUBLICATIONS **52** CITATIONS
SEE PROFILE

Vision and Dead Reckoning-based End-to-End Parking for Autonomous Vehicles

Conference Paper · June 2018

DOI: 10.1109/IVS.2018.8500558

2 authors:

Vijay John
RIKEN Keihanna
77 PUBLICATIONS **1,514** CITATIONS
SEE PROFILE

Swarn singh Rathour
Toyota Technological Institute
14 PUBLICATIONS **52** CITATIONS
SEE PROFILE

自律走行車のための視覚と推測航法に基づくエンドツーエンドの駐車

概要- 本論文では、エンドツーエンドの運転のための視覚と推測航法を組み合わせた駐車システムを提案する。標準的な自律駐車フレームワークは複数のモジュールを含み、各モジュールにはそれぞれ制限がある。一方、提案する駐車フレームワークは、単一のエンドツーエンドモジュールで構成されており、これらの固有の制限を軽減している。提案するディープラーニングベースの駐車システムでは、前方および後方に取り付けた単眼カメラを用いて、ステアリング角度とギアの状態を予測するために、新しい反復2段階学習フレームワークを利用する。提案するフレームワークの第一段階では、エンコーダ・デコーダアーキテクチャを用いて、前方または背面の単眼カメラの複数のフレームから、操舵角の軌跡の初期推定値を予測する。ギアの状態推定値を用いて、ステアリング推定に使用するカメラを選択する。ギアの状態は、初期化中にあらかじめ定義され、その後、提案するフレームワークの第2段階で推定される。提案フレームワークの第2段階では、最適なステアリング角度とギアの状態を推定するために、車両のヘディング角度と絶対位置とともに、ステアリング角度の軌跡の初期推定値が長期短期記憶ネットワークへの入力として与えられる。提案するフレームワークは、取得したデータセットで検証される。ベースラインアルゴリズムとの比較分析、詳細なパラメトリック分析を行う。実験結果は、提案フレームワークがベースラインのエンドツーエンドアルゴリズムよりも優れていることを示している。

I. Introduction

自律走行と先進運転支援システム(ADAS)の分野では、大きな成果を上げている。主要な自動車メーカーの多くは、インテリジェント・パーキング・アシスト・システム(IPAS)を採用しており、パーキング・アシストと半自律型パーキングが市場における主流の自動車技術となっている。初期の駐車支援技術は、1台の屋台(前後クリアランス±70cm)を使って、空の平行駐車スペースに逆転させる機能を備えている。また、これらのシステムは、0.5～1.5mの距離から駐車空間を検出する能力を有している。前バージョンのアップグレードである現在のIPAS技術は、複数人(前後クリアランス±40cm)の駐車が可能である。市販されているにもかかわらず、駐車支援システムには固有の限界があり、駐車問題を完全に解決していない。さらに、構造化された駐車場だけでなく、あるレベルでの人的介入 [2], [3] が必要である。したがって、本稿では、完全自律型セルフパーキングシステムを開発するために、視覚と推測航法を組み合わせたパーキングシステムを紹介する。本論文の主な目的は、現在の自律駐車システムの限界(構造化された駐車場、隣接車両などへの依存)を克服するための学習ベースのフレームワークを開発することである。

本論文の主な貢献は以下の通りである：

- 新しい2段階の深層学習ベースのエンドツーエンド駐車システム(図2)。構造化された駐車場と構造化されていない駐車場で駐車できる。
- エンドツーエンドの運転におけるエンコーダ・デコーダのアーキテクチャの使用。ベースラインのエンドツーエンドドライビングフレームワーク[セクションIV]では、画像から抽出された特徴が回帰ネットワークへの入力として与えられる。一方、エンコーダ・デコーダのアーキテクチャでは、デコーダ出力マップの顕著な特徴が回帰ネットワークへの入力として与えられる。これは推定精度を向上させることが示されている(セクションV)。
- 提案する深層学習フレームワークの予測精度を向上させるために、DRから得られる車両の方位と移動距離を利用する。(第III章)

本稿の残りの部分は以下のように分割される。セクションIIでは、レビュー作業の概要と現在の駐車技術の限界について述べる。セクションIIIでは、エンドツーエンドの駐車のために開発された提案アルゴリズムを定義する。セクションIVでは、提案するディープラーニングベースのエンドツーエンド駐車アルゴリズムの訓練と検証のためのデータセットの準備について説明する。セクションVでは、提案する学習フレームワークと他のベースラインであるエンドツーエンド学習フレームワークの比較検討と、提案するフレームワークのパラメトリックなバリエーションを示す。最後に、セクションVIにおいて、本論文の主な貢献を列挙することで、本論文の結論を述べる。

II. 文献調査

ほとんどの主要自動車メーカー(トヨタ、BMW、フォード、フォルクスワーゲン、メルセデス・ベンツなど)は、半自律型駐車やIPASシステムを搭載した自動車をロールアウトしている。様々な半自律型またはIPAS技術の基本は、依然として類似している。まず、システムは適切な駐車スペースの検出から始まる。次に、システムは、車両周囲の障害物の安全な距離をドライバーに知らせるために、最適なアプローチを定式化する。上記の駐車手順を実行するために、半自律駐車またはIPASシステムは、複数の距離ベースのセンサーとカメラで構成され、関係する変数の大部分を検出するために、車両の前部と後部のバンパーに取り付けられている。したがって、駐車システムは、環境認識、経路生成、制御、衝突回避など、複数のサブモジュールに分けることができる。これらのモジュールはそれぞれ、それ自体が困難な研究課題であり、研究コミュニティによる個別の注意が必要である。半自律型または駐車支援システムは、使用されるセンサーによって、アクティブセンサーベースの駐車(超音波またはレーザーベース)、ビジョンベースのシステム、またはビジョンとアクティブレンジセンサーの組み合わせに分けることができます[4]。

Preprint submitted to 2018 IEEE Intelligent Vehicles Symposium (IV).
Received February 14, 2018.

Vision and Dead Reckoning-based End-to-End Parking for Autonomous Vehicles

***Abstract*— In this paper a combined vision and dead reckoning-based parking system for end-to-end driving is proposed. Standard autonomous parking frameworks contain multiple modules with each module having its own limitation. On the other hand, the proposed parking framework consists of a single end-to-end module, which reduces these inherent limitations. In the proposed deep learning-based parking system, a novel iterative two-stage learning framework is utilized to predict the steering angles and gear status using a front and back mounted monocular camera. In the first stage of the proposed framework, the encoder-decoder architecture is used to predict an initial estimate of the steering angle trajectory from multiple frames of the front or the back monocular camera. The camera used for steering estimated is selected using the gear status estimate. The gear status is predefined during initialization and estimated subsequently in the second stage of the proposed framework. In the second stage of the proposed framework, the initial estimate of the steering angle trajectory along with the vehicles heading angle, and absolute position is given as an input to the long short-term memory network to estimate the optimal steering angle and gear status. The proposed framework is validated on an acquired dataset. A comparative analysis with baseline algorithms and detailed parametric analysis are performed. The experimental results show that the proposed framework is better than the baseline end-to-end algorithms.**

I. Introduction

Significant achievements have been made in the field of autonomous driving and advanced driver assistance systems (ADASs). Most of the major automobile manufacturers have adopted intelligent parking assist system (IPAS), bringing parking assist and semi-autonomous parking as a mainstream automobile technology in the market. The earlier parking assist technology have the ability to reverse park into an empty parallel parking space using a single man-oeuvre (with $\pm 70cm$ front and back clearance). These systems also have the ability to detect the parking space from distances of $0.5 - 1.5m$. Present IPAS technology, an upgrade of the previous version, are capable of multi man-oeuvre (with $\pm 40cm$ front and back clearance); reverse/forward parking into bay parking space. In spite of being commercially available, the parking assist system have inherent limitations and have not completely solved the parking problem. Moreover, they require human intervention at some level [2], [3] as well as structured parking lots. Hence, in order to develop fully autonomous self-parking system this paper introduces combined vision and dead reckoning-based (DR) parking system. The main objective of the paper is to develop a learning-based framework to overcome the limitation of the present autonomous parking system (dependency on structured parking lot, adjacent vehicle etc.). The main contribution of the paper is as following:

- A novel two-stage deep learning based end-to-end parking system(Figure 2). capable of parking in structured as well as unstructured parking areas.
- Use of encoder-decoder architecture for end-to-end driving. In baseline end-to-end driving frameworks [Section IV], the features extracted from the image are given as input to the regression network. On the other hand, in the encoder-decoder architecture, salient features in the decoder output map are given as input to the regression network. This is shown to improve the estimation accuracy (Section V).
- Use of vehicle heading and distance traveled derived from DR to enhance the prediction accuracy of the proposed deep learning framework. (Sec III)

The remainder of the paper is divided as follows. Section II gives an overview of review work and limitation of the present parking technology. Section III delineates the proposed algorithm developed for end-to-end parking. Section IV explains about the dataset preparation for training and validation of the proposed deep learning based end-to-end parking algorithm. Section V gives a comparative study of the proposed learning framework with other baselines end-to-end learning framework as well as parametric variation of the proposed framework. Finally, in Section VI, the paper is concluded by listing the main contributions of the paper.

II. Literature Survey

Most of the major automobile manufacturers (e.g. Toyota, BMW, Ford, Volkswagen, Mercedes-Benz etc.) roll-out their automobile equipped with semi-autonomous parking or IPAS system. The fundamental across the various semi-autonomous or IPAS technology remains similar. Firstly, the system begins with detection of suitable parking space. Next, the system formulates the best approach to get into the space informing the driver about the safe distance of obstacles around the vehicle. In order to perform the above-mentioned parking procedure semi-autonomous parking or IPAS system consists of multiple range based sensor and camera, mounted on the front and rear bumper of a vehicle to detect most of the variable involved. Hence, the parking system can be divided into multiple sub-modules such as environment perception, path generation, control and collision avoidance. Each of these modules is challenging research problems by themselves and require individual attention by the research community. As per the sensors used semi-autonomous or parking assist systems can be divided into active sensor-based parking (ultrasonic or laser-based), vision-based sys-

Preprint submitted to 2018 IEEE Intelligent Vehicles Symposium (IV).
Received February 14, 2018.

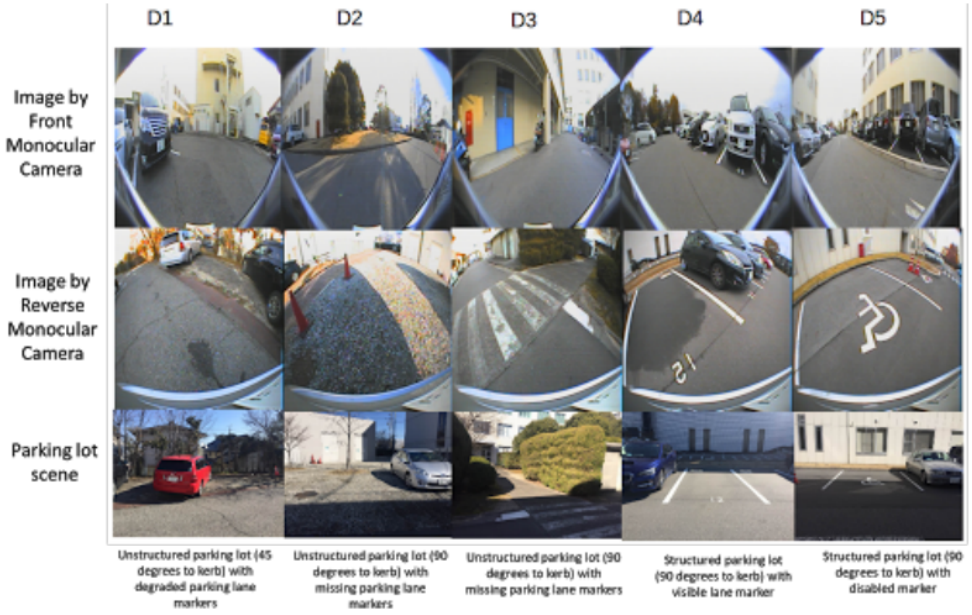

図1: エンドツーエンドの駐車を学習するために、フロントカメラ(上段)とバックカメラ(中段)で撮影された画像のコレクション。(最後の行) 手持ちカメラから撮影されたそれぞれの駐車場の画像。

超音波またはレーザーベースの駐車支援は、駐車支援に使用されるセンサーとして最も一般的であるが [5], [6]、独自の制限 [7]、例えば、超音波センサーベースのシステムは、短距離に制限され、環境内の特定の物体を知覚することが困難である。また、レーザーを用いたレンジセンサーは高精度であるが、高コストで寿命が短いため、使用が制限される。一方、ビジョンベースの駐車支援システムは、駐車レーンマーカーや隣接する駐車車両の存在のような構造化された環境を必要とする[8]、[9]、[10]。しかし、図1)に見られるように、駐車レーンマーカーが欠落しているか、隠されていることが多く、場合によっては隣接する駐車車両が存在しないこともある。さらに、視覚システムは照明の変動や環境ノイズの影響を受けやすい[11]。

本論文では、現在の半自律型駐車支援システムの上記の限界に対処するために、自律走行車のためのディープラーニングを用いた視覚と推測航法に基づく駐車システムを提案する。近年、ディープラーニングフレームワークへの関心が高まっており、人間レベルの精度で画像認識とセグメンテーションの分野で最先端の結果が得られている[11], [12], [13]、自律走行[14], [15], [16], [17]。従来の自律走行フレームワークと比較して、深層学習ベースのフレームワークは、画像から直接操舵角を予測する単一モジュールまたはエンドツーエンドの学習フレームワークで構成されている。このようなシステムでは、環境認識、経路計画、障害物回避、制御など、複数のモジュールが不要になる。エンドツーエンド学習ベースのアルゴリズムは、エンドツーエンドで完全に学習可能であるため、最小限の人的労力で済む。

出力としてmand。したがって、エンドツーエンドの学習は、異なるモジュールを明示的にモデル化する必要がないため、魅力的である。しかし、ディープラーニングに基づくエンドツーエンドの運転に関する既存の文献は、高速道路や公共道路での運転に限られている[14]、[15]、[16]、[17]。本研究では、エンドツーエンドの深層学習フレームワークを自律駐車に拡張する。さらに、従来の駐車支援システムの問題点、特に構造化された環境を必要とする視覚ベースの駐車支援の問題点にも取り組む。

III. Algorithm

本論文の主な目的は、前後に取り付けた魚眼カメラを用いて得られた時間同期された連続画像観測から、専門家の操舵角のシーケンスとギアの状態を予測することである。さらに、移動距離と車両の方位(推測航法)もステアリング角度の予測に使用される。提案するネットワークは、エンコーダ・デコーダ段階とLSTM段階からなる2段階のエンドツーエンド学習フレームワークである。図2は、提案するフレームワークの各ステージの詳細なアーキテクチャを示す。

第一段階では、セマンティックセグメンテーション[18]に使用されるエンコーダ・デコーダアーキテクチャに基づく深層学習フレームワークを、前後に取り付けられた魚眼カメラから得られた画像を用いて、操舵角を予測するように修正する(図2)。前方カメラ画像は前進運動中のステアリング角を予測するために使用され、後方カメラ画像は後進運動中のステアリング角を予測するために使用される。カメラの選択はギアの状態に依存する。初期フレームでは、車が前進していると仮定し、フロントカメラを使用する。この後のフレームワークでは、歯車の状態を

Preprint submitted to 2018 IEEE Intelligent Vehicles Symposium (IV).
Received February 14, 2018.

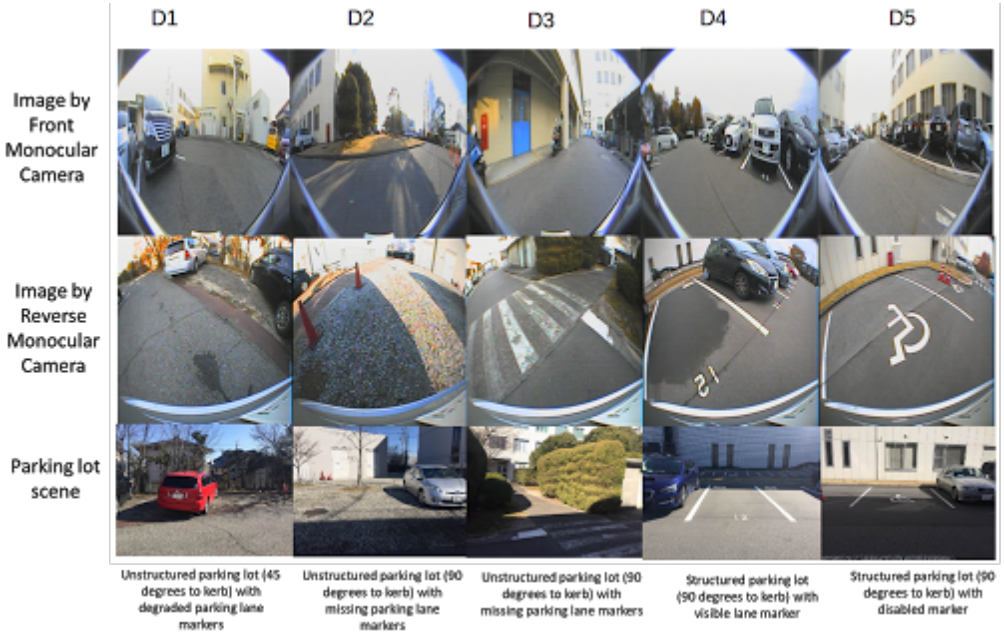

Fig. 1: A collection of images taken by front camera (top row) and back camera (middle row) for training end-to-end parking. (Last row) Image of the respective parking lot captured from hand held camera.

tems or a combination of vision and active range sensors [4]. Ultrasonic or laser-based parking assist are most common [5], [6] sensor used for parking assist, however, they are prone to their own limitation [7]; for example, the ultrasound sensor-based systems are limited to short ranges and have difficulties in perceiving certain objects in the environment. In addition, laser-based range sensor is accurate, however, their high cost and short life limit their use. On the other hand, vision-based parking assistance systems require a structured environment like the presence of parking lane markers or the adjacent parked vehicle [8], [9], [10]. However, as observed in Fig. 1), often the parking lane markers are either missing or occluded and in some case there is no adjacent parked vehicle. Additionally, vision systems are susceptible to illumination variation and environmental noise [11].

In this paper, we propose a vision and dead reckoning-based parking system using deep learning for autonomous vehicles to address the above-mentioned limitations of current semi-autonomous or parking assist systems. Recently there has been an increase in interest in the deep learning framework as it has provided state of the art results in the field of image recognition and segmentation with human-level accuracy [11], [12], [13] and autonomous driving [14], [15], [16], [17]. Compared to traditional autonomous driving frameworks, the deep learning-based frameworks consists of a single module or an end-to-end learning framework which directly predicts the steering angle from an image. Such systems eliminate the need for multiple modules such as environment perception, path planning, obstacle avoidance and control. End-to-end learning based algorithm requires minimal human effort being fully end-to-end trainable; taking image observation as input and steering control command as output. Hence, end-to-end learning is appealing, as it removes the need to explicitly model the different modules. However, existing literature on deep learning-based end-to-end driving is limited to highway and public road driving [14], [15], [16], [17]. In this work, we extend the end-to-end deep learning framework to autonomous parking. Additionally, we address the issues with traditional parking assist systems, especially the vision-based parking assistances which need a structured environment.

III. Algorithm

The main objective of this paper is to predict the sequence of expert steering angles and the gear status, given the time synchronized sequential image observations obtained using a front and back mounted fish-eye camera. Additionally, the distance traveled and vehicle heading (dead reckoning) are also used to predicted the steering angles. The proposed network is a two stage end-to-end learning framework consisting of an encoder-decoder stage and LSTM stage. Figure 2, shows the detailed architecture of the different stages of the proposed framework.

In the first stage, the encoder-decoder architecture based deep learning framework, used for semantic segmentation [18], is modified to predict the steering angle (Fig. 2) using images obtained from the front or back mounted fish-eye camera. The front camera images are used to predict the steering angles during the forward motion, and the back camera images are used to predict the steering angles during the reverse motion. The choice of camera is dependent on the gear status. For the initial frames, we assume the car to be moving forward and use the front cameras. For the subsequent framework, we estimate the gear status using the

Preprint submitted to 2018 IEEE Intelligent Vehicles Symposium (IV).
Received February 14, 2018.

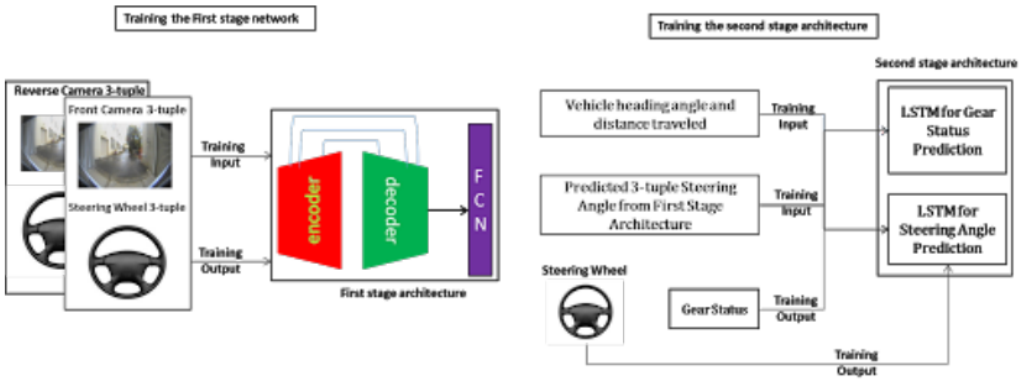

図2:学習ステップの概要。

提案するフレームワークの第2段階

第一段階の操舵角を予測するために、[18]で提案されたようなディープセグメンテーションネットワークフレームワーク、すなわちUNETが、その単純さと他のベースラインセグメンテーションネットワーク[12]、[13]に対する優位性から選択された。図2に示すように、ディープラーニングフレームワークの第1段階は、UNETと洗練ネットワークを用いて定式化される。より具体的には、2つのネットワークを結合するために、U-Netデコーダブロックの出力が洗練ネットワークへの入力として与えられる。精密化ネットワークは3つの完全接続層から構成される。最初の2つの完全連結層は、それぞれ512ユニットと最後の256ユニットを含む。最初の2つの完全連結層はReLu非線形を使用し、最後の層は線形活性化関数を使用する。損失モデルとしての平均二乗誤差は、モデルの重みを学習するためにAdam optimizer (1e-5)と共に使用される。

第2段階では、LSTMネットワークを用いて第1段階の操舵角予測を精緻化し、ギヤの状態を推定する。LSTMネットワークへの入力は、車両の推測航法測定値とともに、ステージ1の操舵角予測値のシーケンスである。次に、アルゴリズムの学習とテストのステップを詳しく説明する。

1) Training Step:

a) エンコーダ・デコーダの学習: エンコーダ・デコーダベースの完全連結ネットワークは、3タプルの画像観測、すなわち(I_{i-1}, I_i, I_{i+1})を学習入力とし、3タプルの操舵角、すなわち(s_{i-1}, s_i, s_{i+1})を出力として学習する。典型的な駐車操作は、前方走行と後方走行の両方から構成される。ステアリング角予測の精度を高めるため、画像取得に2台のカメラを別々に利用する。前方マヌーバの間、3タプルの画像観測は、フロントマウントカメラを使用して取得される。一方、リバースマヌーバでは、バックマウントカメラを用いて3タプルの画像観測を取得する。3タプルの画像観測は、常に車両CANBUSから得られる3タプルの操舵角と同期していることに注意。両カメラから取得した画像は、第1段階のネットワークの学習に使用される。

学習データが与えられると、第1段階のネットワークにおける画像からステアリングへのマッピング関数は、教師ありの方法で学習される。

より具体的には、提案するエンコーダ・デコーダに基づく完全連結フレームワークで近似された回帰関数のパラメータθを最適化する。

b) LSTMの学習: b)LSTMの学習:ネットワークの第2段階では、デッドレコニング(DR)測定値を用いて、第1段階のステアリング角度推定値を精緻化することで、提案するディープラーニングフレームワークの性能を向上させる。より具体的には、第一段階の操舵角推定値とDR測定値が、操舵角の最適推定値を生成するLSTMネットワークへの入力として与えられる。さらに、LSTMネットワークは、第一段のカメラ選択のためのテスト段階で使用されるギアの状態も予測する。

DR測定値は、車両の移動距離(すなわちd)に対応し、4輪回転速度、車速、ギア状態g、ヨーレートϕ ¤からなる時間同期された缶バスデータを用いて導出される。提案するLSTMネットワークは、第1段階のディープネットワークから予測されたs^を、移動距離dと絶対方位角ϕとともに取る。LSTMネットワークは2つの出力からなり、最初の出力はg^を予測するLSTMベースの分類器である。g^は2つの状態のみの分類器から構成されるため、LSTMモデルがこのタスクに適している。次に、LSTMモデルは線形活性化ユニットと平均二乗誤差損失モデルで$s_{\sim f}$を予測する。

2) テスト段階: 2段階の提案ディープネットワークフレームワークを学習した後に得られたパラメータを用いて、時間同期したI,dとϕを$s_{\sim f}$(最終予測制御操舵角)とg^(予測ギア状態)に対応付ける。テストの概要を図 3 に示す。

初期化中、提案アルゴリズムは、車両がギア状態g = 1で前進していると仮定して初期化される。その結果、初期化中、第1段階のネットワークはフロントカメラを使用してs^を推定する。

ステアリング角の初期推定値 s^ と車両 DR の測定値が第 2 段階の LSTM ネットワークの入力として与えられる。LSTMは、ギアの状態g^とともに最適なステアリング角$s_{\sim f}$を推定する。次に、ギアの状態は、入力用の背面カメラまたはフロントカメラを選択するために、第1ステージのネットワークによって再帰的に使用される。g = 1はフロントカメラ、g = 0はバックカメラを意味する。

Preprint submitted to 2018 IEEE Intelligent Vehicles Symposium (IV).
Received February 14, 2018.

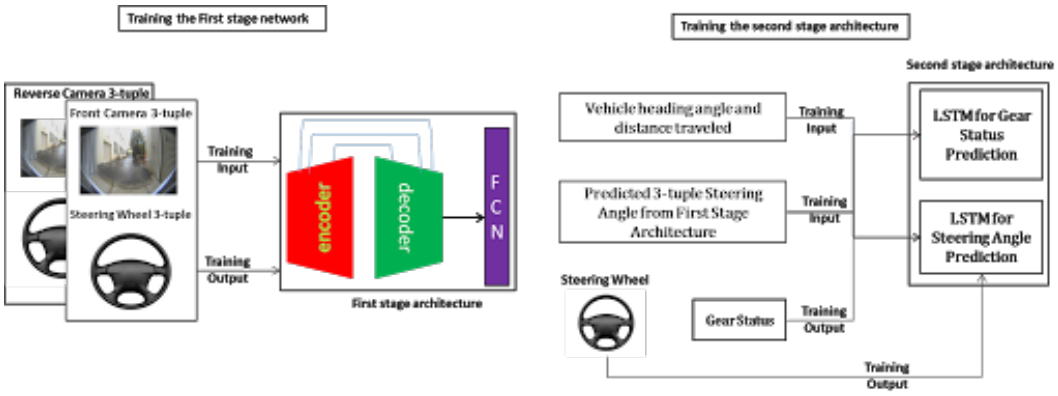

Fig. 2: An overview of the training step.

second stage of the proposed framework.

To predict the first stage steering angles, the deep segmentation network framework as proposed in [18] i.e. UNET was selected due to its simplicity and advantage over other baseline segmentation networks [12], [13]. As shown in Fig. 2, the first stage of the deep learning framework is formulated using the UNET and a refinement network. More specifically, to combine the two networks, the output of the U-Net decoder block is given as the input to the refinement network. The refinement network consists of 3 fully connected layers. First two fully-connected layers contain 512 units each followed by 256 in the last. The first two fully connected layer uses ReLu nonlinearity; whereas the last one uses linear activation function. Mean square error as loss model is used with Adam optimizer (1e-5) for learning the weights of the model.

In the second stage, the LSTM network is used to refine the first stage steering angle prediction and estimate the gear status. The input to the LSTM network is a sequence of stage one steering angle predictions along with the vehicle dead reckoning measurements. We next explain the training and testing steps of the algorithm in detail.

1) Training Step:

a) Encoder-Decoder Training: The encoder-decoder based fully connected network is trained using 3-tuple image observation i.e.(I_{i-1}, I_i, I_{i+1}) as training input and 3-tuple steering angle i.e. (s_{i-1}, s_i, s_{i+1}) as output. A typical parking maneuver consists of both forward driving and reverse driving. To enhance the accuracy for steering angle prediction, we utilize two separate cameras for image acquisition. During the forward maneuvering, the 3-tuple image observations are acquired using the front mounted cameras. On the other hand during the reverse maneuvering, the 3-tuple image observations are acquired using the back mounted cameras. Note that the 3-tuple image observations are always synchronized with the 3-tuple steering angles obtained from the vehicle CANBUS. The images acquired from both the cameras are used to train the first stage network.

Given the training data, the image-to-steering mapping function in the first stage network is trained in a supervised manner. More specifically, the parameter θ of the regression function approximated by proposed encoder-decoder based fully connected framework is optimized.

b) LSTM Training: In the second stage of the network, the dead reckoning (DR) measurements are used to enhance the performance of proposed deep learning framework by refining the first stage steering angle estimates. More specifically, the first stage steering angle estimates along with DR measurements are given as an input to the LSTM network which generates an optimal estimate of the steering angle. Additionally, the LSTM network also predicts the gear status, which is used during the testing phase for the first stage camera selection.

The DR measurements correspond to the distance traveled (i.e. d) by the vehicle; derived using time-synchronized can bus data comprised of four-wheel rotational speed, vehicle speed, gear status g and yaw rate $\dot{\psi}$. The proposed LSTM network takes predicted $\hat{s}$, from the first stage deep network along with the distance traveled d and absolute heading angle ψ. The LSTM network consists of two outputs, first is LSTM based classifier to predict the $\hat{g}$. As $\hat{g}$, is composed of two state only classifier based LSTM model is better suited for this task. Secondly, LSTM model predicts $\hat{s_f}$ with linear activation unit and mean square error loss model.

2) Testing Phase: The parameter obtained after training the two-stage proposed deep network framework, is used to map the time synchronized I,d and ψ to $\hat{s_f}$ (final predicted control steering angle) and $\hat{g}$ (predicted gear status). The overview of the testing is shown in Fig. 3.

During initialization, the proposed algorithm is initialized with the assumption that the vehicle is moving forward with gear status $g = 1$. Consequently, during initialization, the first stage network uses the front camera to estimate $\hat{s}$.

The initial estimate of the steering angle $\hat{s}$ along with the vehicle DR measurements are given as an input to the second stage LSTM network. The LSTM estimates the optimal steering angle $\hat{s_f}$ along with the gear status $\hat{g}$. The gear status is then recursively used by the first stage network to select the back or the front camera for the input. $g = 1$ implies front camera and $g = 0$ implies back camera.

Preprint submitted to 2018 IEEE Intelligent Vehicles Symposium (IV).
Received February 14, 2018.

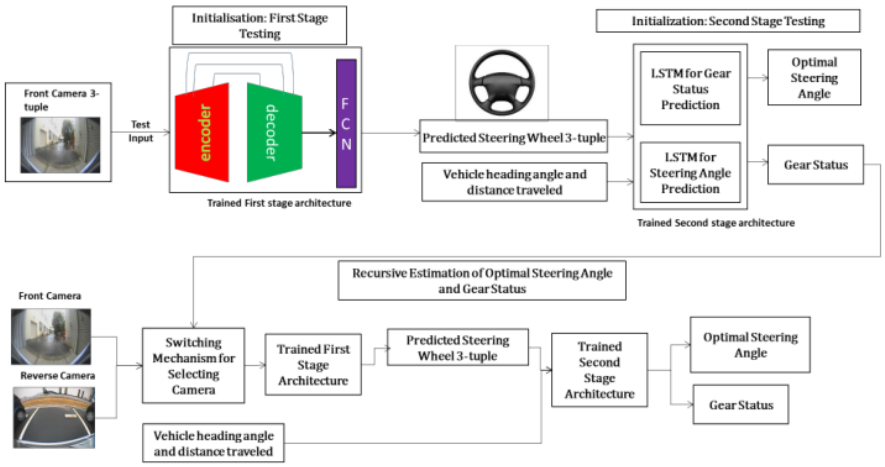

図3:テストステップの概要。

IV. 実験結果

まず始めに、提案する深層学習フレームワークの学習とテストに使用したデータセットを紹介する。最後に、実験セットアップと結果の考察を述べる。ベースラインモデルとの比較分析を行う。さらに、2段階ネットワークのバリエーションによるパラメータ分析も行う。すべてのトレーニングおよび検証は、以下の仕様のシステムで実行された:64ビットIntel Core i7-6850K CPU @ 3.60GHz×12, GeForce GTX 1080, RAM 64GB、テンソルフローバックエンド付きkerasを使用。

A. 実験データセット

自律的な逆駐車のためのエンドツーエンドの視覚認識ベースのディープラーニングフレームワークを学習するために、エキスパートドライバが実証した駐車は、トレーニングとテストのためのデータセットを準備するために使用された。エキスパートドライバーは、様々なデータセットを作成するために、多数の駐車場を実行するよう求められた。D1、D2、D3、D4、D5の5つのデータセット(図1)をそれぞれ2つのシーケンスから構成し、トレーニング用とテスト用に用意した。各データセットは、時間同期された前後カメラ画像観測値I, ϕ, s, dから構成される。各データセットは別々のトレーニングシーケンスとテストシーケンスに分割された。図1は、5つのデータセットすべてについて、フロントカメラとバックカメラで撮影した画像である。データセットD1、D2、D3は、非構造化駐車場、すなわち、劣化した白線と欠落した白線のある駐車場を用いて作成した(図1、最終行)。したがって、D1-D3は非構造化駐車場を表し、D4-D5は構造化駐車場データセット(よく定義された白線(図1、最後の行))を表す。

提案アルゴリズムの性能を検証するために、各データセットの学習シーケンス(すなわちD1.・・・D5)に対して提案モデルを学習させ、最後に学習後に得られた最適化パラメータθを用いて、対応するデータセットのテストシーケンスに対するテストを行った。例えば、D1の訓練シーケンスで提案モデルを訓練することによって得られる最適化されたパラメータθは、D1のテストシーケンスでテストされた。

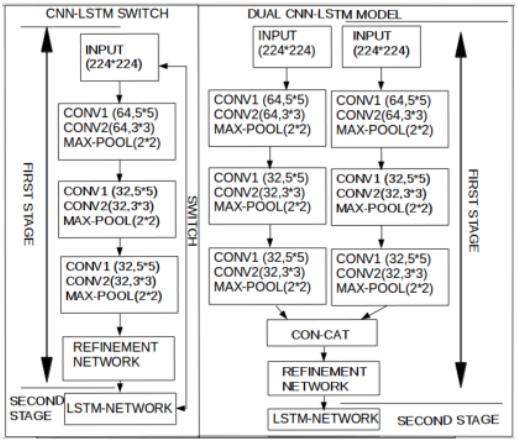

図4:パラメータ分析に使用した深層学習ベースのフレームワークのCNNアーキテクチャ。

B. 比較分析

提案モデルとの比較のために、4つのベースライン深層学習ベースのエンドツーエンドネットワークを選択した。画像ピクセルをステアリング角にマッピングするために[11]で使用されたCNNアーキテクチャ(すなわちVGG16)は、最初のベースラインを表す。残差ベースのCNNアーキテクチャ(RESNET50)は、[20]で提案されているように、2番目のベースラインを表す。3番目と4番目のベースラインでは、VGG-16とRESNET50の最終畳み込み層から抽出した特徴マップを抽出し、木ベースの回帰器(すなわち、VGG16 + ET & RESNET50 + ET)[21]を追加で学習させた。

提案するフレームワークで2台のカメラを利用することの利点を実証するために、ベースラインモデルはフロントカメラ「単独」を使用して学習された。さらに、これは先行研究[15]、[21]で使用された自動運転ネットワークを模倣するためにも行われる。比較のために、5つのデータセットすべてについて、4つのベースラインモデルの性能を調べた。

Preprint submitted to 2018 IEEE Intelligent Vehicles Symposium (IV).
Received February 14, 2018.

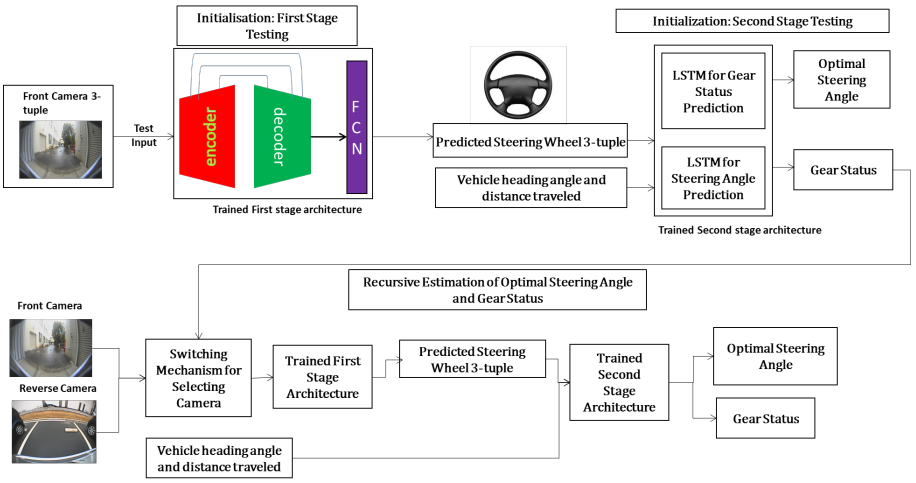

Fig. 3: An overview of the testing step.

IV. Experimental Results

To begin with, firstly, we introduce the dataset used for training and testing the proposed deep learning framework. Finally, we delineate the experimental setup and result discussion. A comparative analysis is performed with baseline models. Additionally, we also perform a parameter analysis with variations of the two-stage network. All the training and validation was performed on the system with the following specification:- 64-bit Intel Core i7-6850K CPU @ 3.60GHz×12, GeForce GTX 1080, RAM 64 GB using keras with tensor flow backend.

A. Experimental Dataset

In order to learn end-to-end visual perception based deep learning framework for autonomous reverse parking, expert driver demonstrated parking was used to prepare the dataset for training and testing. The expert driver was asked to perform numerous parking to create the various dataset. Five datasets (Fig. 1) i.e. D1, D2, D3, D4 and D5 consisting of two sequences each was prepared for training and testing respectively. Each dataset consists of time synchronized front and back camera image observation I,ψ,s and d. Each dataset was partitioned into separate training and testing sequences. Figure 1 shows the image captured by front and back camera for all the five datasets. Dataset D1, D2 and D3 were prepared using unstructured parking lot i.e. the parking lot with degraded white line and missing white line(Figure 1, last row). Hence D1-D3 represents unstructured parking lot and D4-D5 represents structured parking lot dataset (well defined white lines (Figure 1, last row)).

In order to validate the performance of the proposed algorithm, the proposed model was trained on the training sequence of each dataset (i.e. D1....D5) and finally the optimized parameter θ obtained after training was used to test on the testing sequence of the corresponding dataset. For example, the optimized parameter θ derived by training the proposed model on training sequence of D1 was tested on the testing sequence of D1.

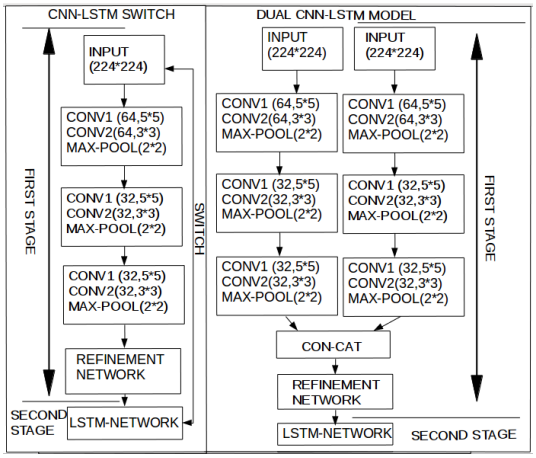

Fig. 4: CNN architecture of the variant deep learning based frameworks used for the parameter analysis.

B. Comparative Analysis

Four baseline deep learning based end-to-end networks were selected for the comparison with the proposed model. The CNN architecture used by [11] (i.e. VGG16) to map the image pixel to steering angle represents the first baseline. The residual based CNN architecture (RESNET50) as proposed in [20] represents the second baseline. For the third and fourth baseline, we extracted feature maps extracted from the final convolutional layer of the VGG-16 and RESNET50 to train an extra trees-based regressor (i.e.VGG16 + ET & RESNET50 + ET) [21].

To demonstrate the benefits of utilizing two cameras in the proposed framework, the baseline models were trained using the front camera "alone". Moreover, this is also done to imititate the automated driving networks used in the prior work [15], [21]. The performance of the four baseline models were studied on all the five datasets for comparison. The different models are quantitatively compared by measuring the mean Euclidean distance between the predicted steering

Preprint submitted to 2018 IEEE Intelligent Vehicles Symposium (IV).
Received February 14, 2018.

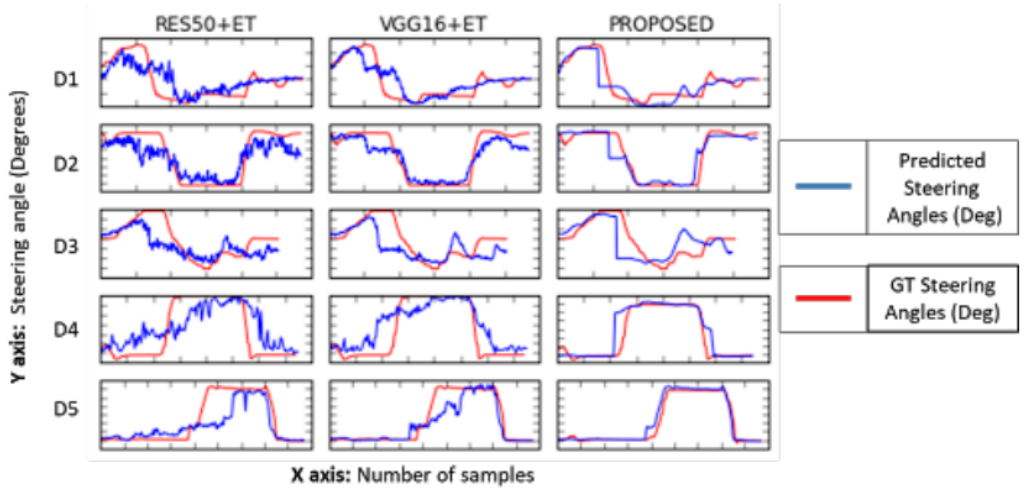

図5:RES50+ET、VGG16+ET、Proposedを用いた各データセットの予測値s ～fの軌跡(青線)とグランドトゥルースの軌跡s(赤線)。

表1: 表1:異なるデータセットにおける、予測されたステアリング角とグランドトゥルースのステアリング角の平均ユークリッド距離(度)。

Data.	RES50	RES50+ET	VGG16	VGG16+ET	**Prop.**
D1	93.74	23.44	157.35	17.08	9.22
D2	64.02	56.00	30.13	41.0	7.89
D3	93.74	105.66	212.78	131.11	11.93
D4	64.47	43.75	157	65.35	38.61
D5	166.17	56.29	29.37	41.46	17.70

表II:異なるステアリング角の予測値とグランドトゥルースとの間の平均ユークリッド距離(度)。

datasets

Data.	**Prop.**	CNN-LSTM-スイッチ	Dual CNN-LSTM
D1	9.22	29.49	22.28
D2	7.89	28.28	31.69
D3	11.93	31.98	155.86
D4	38.61	54.54	133.25
D5	17.70	143.30	92.13

角度とグランドトゥルースのステアリング角度。

図5と表1で得られた結果は、提案ネットワークが異なるデータセットにおいて、ベースラインアルゴリズムよりも優れていることを示している。ベースラインと比較して、性能の向上は以下のことに起因する:

- ステアリング角の初期推定値がLSTMによって精緻化される2段アーキテクチャ。
- エンコーダ・デコーダのアーキテクチャを利用して、U-Netの出力マップが完全連結ネットワークへの入力として与えられる、操舵角の初期推定値を得る。
- 予測のために2台のカメラを切り替える。

C. パラメータ解析

提案モデルとそれぞれのベースラインモデルの比較の後、パラメトリック依存性を研究するために、第一段階での提案モデルアーキテクチャを変化させた。

パラメトリック解析では、提案したフレームワークの2つのバリエーションを使用した(図4)。最初のバリエーションでは、カメラ切り替えメカニズムを排除し、U-Netをフロントカメラ画像とバックカメラ画像用の2つのCNNブランチに置き換える。CNNの枝によって抽出された特徴は連結され、第一段階で洗練ネットワークに与えられる。第2ステージでは、LSTMによって初期ステアリング角が精緻化される。このモデルをデュアルCNN-LSTMモデルと呼ぶ。

2つ目のバリエーションでは、提案フレームワークのU-NetをCNNに置き換え、切り替えメカニズムを保持したまま、画像特徴を抽出する。このモデルをCNN-LSTMスイッチングモデルと呼ぶ。2つの変形モデルのCNNアーキテクチャを図に示す。

4.

パラメトリック解析で得られた結果を図6と表2に示すが、提案モデルの方がパラメトリックバリエーションよりも優れていることがわかる。パラメトリック解析により、U-Net、スイッチング機構、LSTMフレームワークを用いることの利点が示された。U-Netの利点はCNNLSTM-Switchingモデルとの比較で明らかであり、提案モデルのU-Netと変形モデルのCNNを除いて、これらのモデルはどちらも類似している。

提案モデルでは、スイッチング機構を用いて前方の ”または ”後方 ”を選択するため、U-Netとスイッチング機構の優位性は、Dual CNN-LSTMモデルとの比較において明らかである。一方、変形モデルでは、両方のカメラ画像が使用される。

V. Conclusion

本論文では、前方または後方に取り付けた単眼カメラを用いて、ステアリング角度とギアの状態を予測するために、ビジョンとDRを組み合わせた新しい2段エンコーダ・デコーダアーキテクチャを提案する。提案モデルは、限られた学習データセットにおいて、ベースラインのエンドツーエンドモデルと比較して、より良い性能を示す。

Preprint submitted to 2018 IEEE Intelligent Vehicles Symposium (IV).
Received February 14, 2018.

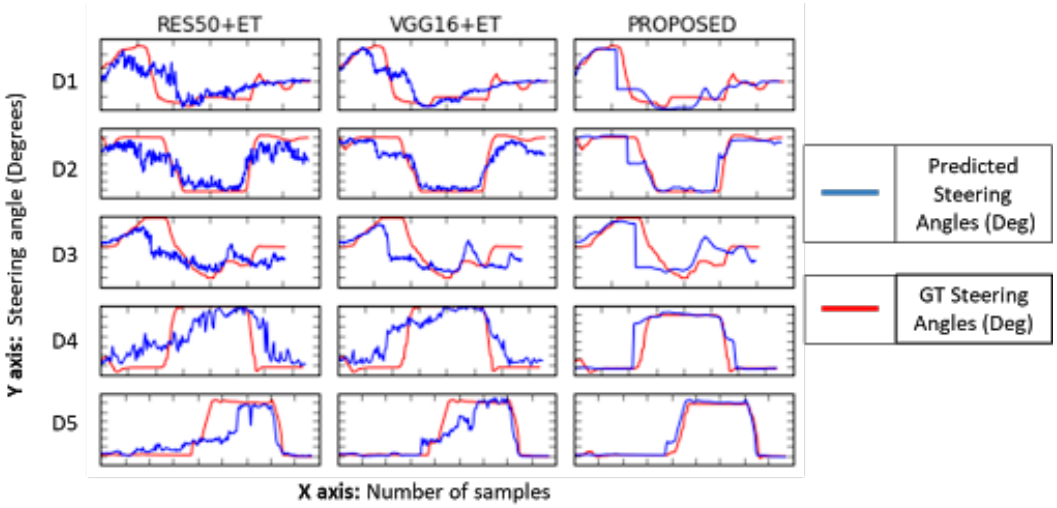

Fig. 5: Trajectories of predicted $\hat{s_f}$ (blue line) along with the ground truth trajectory s (red line) for each dataset using RES50+ET, VGG16+ET and Proposed.

TABLE I: Mean Euclidean distance (degrees) between the predicted and ground truth steering angles for the different datasets.

Data.	RES50	RES50+ET	VGG16	VGG16+ET	**Prop.**
D1	93.74	23.44	157.35	17.08	9.22
D2	64.02	56.00	30.13	41.0	7.89
D3	93.74	105.66	212.78	131.11	11.93
D4	64.47	43.75	157	65.35	38.61
D5	166.17	56.29	29.37	41.46	17.70

TABLE II: Mean Euclidean distance (degrees) between the predicted and ground truth steering angles for the different datasets

Data.	**Prop.**	CNN-LSTM-Switch.	Dual CNN-LSTM
D1	9.22	29.49	22.28
D2	7.89	28.28	31.69
D3	11.93	31.98	155.86
D4	38.61	54.54	133.25
D5	17.70	143.30	92.13

angles and the ground truth steering angles.

The results obtained in Figure 5 and Table 1 show that the proposed network is better than the baseline algorithms across different datasets. Compared to the baseline, the improved performance can be attributed to the following:

- Two-stage architecture where the initial estimate of the steering angle is refined by the LSTM.
- Utilizing the encoder-decoder architecture to obtain an initial estimate of the steering angle, where the U-Net's output map is given as input to the fully connected network.
- Switching between two cameras for prediction.

C. Parameter Analysis

After the comparison of the proposed model with the respective baseline models, the proposed model architecture at the first stage was varied to study the parametric dependency. Two variations of the proposed framework was used in the parametric analysis (Figure 4). In the first variation, we eliminate the camera switching mechanism and replace the U-Net with two CNN branches for the front and back camera images. The features extracted by the CNN branches are concatenated and given to the refinement network in the first stage. The initial steering angle is refined by the LSTM in the second stage. This model is called as *Dual CNN-LSTM model*.

In the second variation, we replace the U-Net in the proposed framework with CNN to extract the image features, while retaining the switching mechanism. This model is termed as the *CNN-LSTM switching model*. The CNN architecture for the two variant models are shown in Figure 4.

The results obtained in the parametric analysis are shown in Fig 6 and Table 2, show that the proposed model is better than the parametric variations. The parametric analysis demonstrates the advantages of using the U-Net, the switching mechanism and the LSTM framework. The advantages of the U-Net is evident in the comparison with the CNN-LSTM-Switching model, as both these models are similar except the U-Net in the proposed model and the CNN in the variant model.

The advantages of the U-Net and the switching mechanism are evident int the comparison with Dual CNN-LSTM model, as the switching mechanism is used to select the front "or" back in the proposed model. On the other hand, both the camera images are used by the variant model.

V. Conclusion

In this paper, a combined vision and DR based novel two-stage encoder-decoder architecture is proposed to predict the steering angles and gear status using front or back mounted monocular camera. The proposed model shows better performance compared to the baseline end-to-end models with

Preprint submitted to 2018 IEEE Intelligent Vehicles Symposium (IV).
Received February 14, 2018.

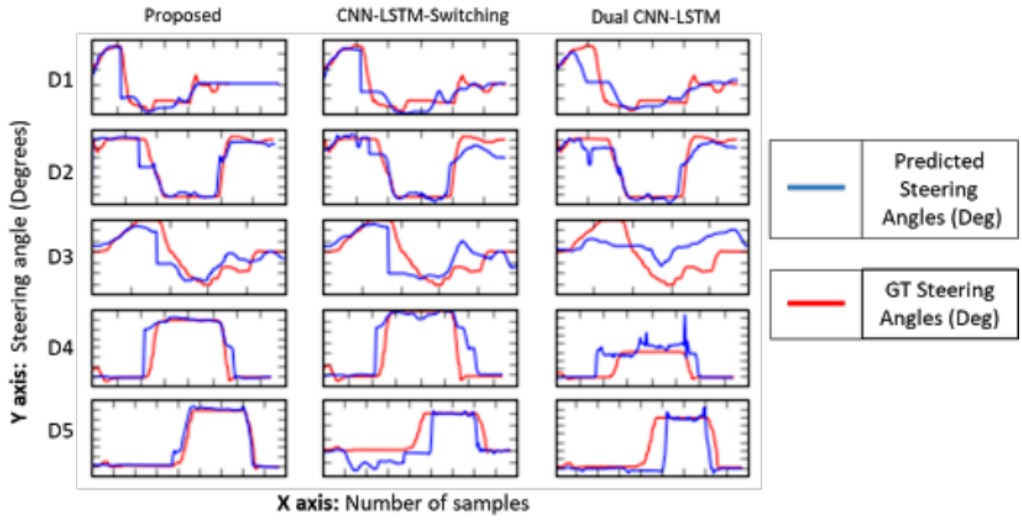

図6: 提案モデル、分離モデル、複合モデルを用いた各データセットの予測値s $\sim_f$ の軌跡とグランドトゥルースの軌跡 s。

提案モデルの有効性を検討するため、提案ネットワークの第1段階でもパラメータ変動はほとんど行わなかった。提案モデルは、ステアリング軌道とギアの状態を予測するのに有効であり、システムを完全に自律化し、マルチマンオーブを行うことができることがわかった。提案するディープラーニングに基づくエンドツーエンドの駐車場は、完全に見える、部分的に見える/見えない駐車場の白線がある非構造駐車場や、駐車場に駐車場がない場合でも使用できる。

REFERENCES

[1] W. Wang, Y. Song, J. Zhang, and H. Deng, "Automatic parking of vehicles: A review of literature," International Journal of Automotive Technology, vol. 15, no. 6, pp. 967-978, 2014.

[2] The Hybrid That Started it All, Mar. 2014. [Online]. Available: http://www.toyota.com/prius/

[3] BMW 7 Series. Park Assist, Mar. 2013. [Online]. Available: http://www.bmw.com/

[4] X. Du and K. K. Tan," Autonomous reverse parking system based on robust path generation and improved sliding mode control," IEEE Trans. Intell. Transp. Syst., vol. 16, no. 3, pp. 1225-1237, Jun. 2014.

[5] J. Pohl, M. Sethsson, P. Degerman, and J. Larsson, "A semi-automated parallel parking system for passenger cars," Proc. Inst. Mech. Eng. D, J. Autom. Eng., vol. 220, no. 1, pp. 53-65, Jan. 2006.

[6] H. G. Jung, Y. H. Cho, P. J. Yoon, and J. Kim, "Scanning laser radar based target position designation for parking aid system," IEEE Trans. Intell. Transp. Syst., vol. 9, no. 3, pp. 406-424, Sep. 2008.

[7] P. Degerman J. Pohl M. Sethson "Ultrasonic sensor modeling for automatic parallel parking systems in passenger cars," SAE 2007 World Congress & Exhibition, Detroit, MI, U.S.A., 16th19th April, 2007 .

[8] K. Fintzel R. Bendahan C. Vestri S. Bougnoux T. Kakinami "3D parking assistant system," Proc. IEEE Intell. Veh. Symp.¡/em¿ pp. 881-886 2004.

[9] N. Kaempchen U. Franke R. Ott "Stereo vision based pose estimation of parking lots using 3d vehicle models" ¡em¿Proc. IEEE Intell. Veh. Symp.¡/em¿ vol. 2 pp. 459-464 2002.

[10] C. Wang,, Hengrun Zhang, Ming Yang, Xudong Wang, Lei Ye and Chunzhao Guo. "Automatic parking based on a bird's eye view vision system," Advances in Mobility Theories, Methodologies, and Applications, vol. 2014 pp. 847406-1-847406-13 2014.

[11] Alex Krizhevsky, IlyaSutskever, and Geoffrey E Hinton,"Imagenet classification with deep convolutional neural networks,"in Advances in neural information processing systems (NIPS), 2012, pp. 1097-1105.

[12] M. Thoma, "A survey of semantic segmentation," CoRR, vol. abs/1602.06541, 2016.

[13] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. G. Rodriguez, "A review on deep learning techniques applied to semantic segmentation," CoRR, vol. abs/1704.06857, 2017.

[14] C. Chen, A. Seff, A. Kornhauser, and J. Xiao. Deepdriving: Learning affordance for direct perception in autonomous driving. In Proceedings of the IEEE International Conference on Computer Vision, pages 2722-2730, 2015.

[15] 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.

[16] H. Xu, Y. Gao, F. Yu, and T. Darrell. End-to-end learning of driving models from large-scale video datasets. arXiv preprint arXiv:1612.01079, 2016.

[17] Lu Chi and Yadong Mu. "Deep Steering: Learning End-to-End Driving Model from Spatial and Temporal Visual Cues". In: arXiv preprint arXiv:1708.03798 (2017).

[18] Y. LeCun, U. Muller, J. Ben, E. Cosatto, and B. Flepp. Offroad obstacle avoidance through end-to-end learning. In NIPS, pages 739-746, 2005.

[19] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MIC- CAI, pages 234-241. Springer, 2015.

[20] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages,770778, 2016.

[21] Vijay John, Seiichi Mita, Hossein Tehrani Niknejad, Kazuhisa Ishimaru,"Automated driving by monocular camera using deep mixture of experts,IV 2017, 10.1109/IVS.2017.7995709.

Preprint submitted to 2018 IEEE Intelligent Vehicles Symposium (IV).
Received February 14, 2018.

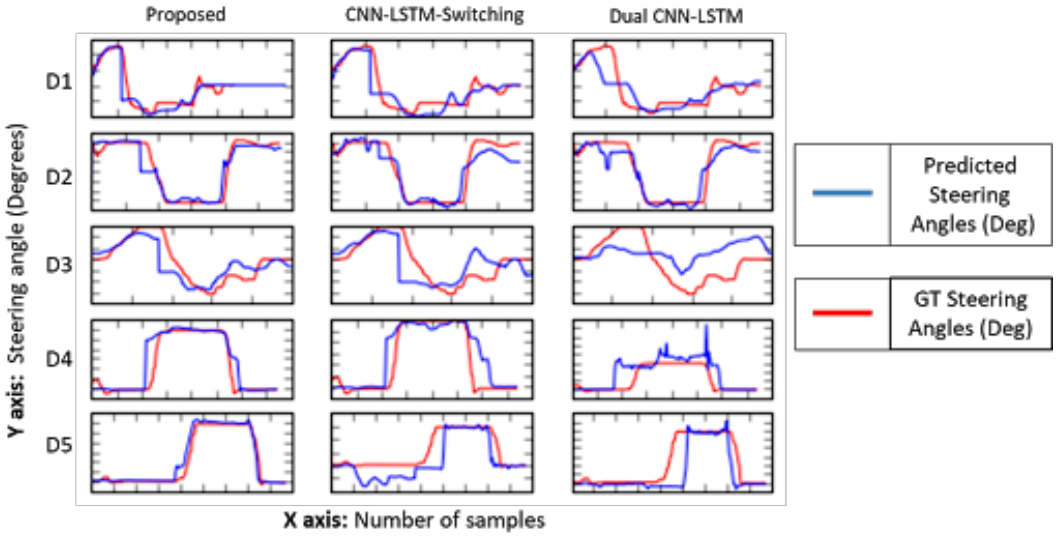

Fig. 6: Trajectories of predicted $\hat{s_f}$ along with the ground truth trajectory s for each dataset using proposed model, separate and combined model.

limited training dataset. Few parameter variation was also performed at the first stage of the proposed network to study the effectiveness of the proposed model. The proposed model was found effective in predicting steering trajectory and gear status making the system fully autonomous and capable of multi man-oeuvre. The proposed deep learning based end-to-end parking can be used even for an unstructured parking lot with fully visible, partially visible/occluded parking white line or even in case if the parking lot has no parking line.

REFERENCES

[1] W. Wang, Y. Song, J. Zhang, and H. Deng, "Automatic parking of vehicles: A review of literature," International Journal of Automotive Technology, vol. 15, no. 6, pp. 967-978, 2014.

[2] The Hybrid That Started it All, Mar. 2014. [Online]. Available: http://www.toyota.com/prius/

[3] BMW 7 Series. Park Assist, Mar. 2013. [Online]. Available: http://www.bmw.com/

[4] X. Du and K. K. Tan," Autonomous reverse parking system based on robust path generation and improved sliding mode control," IEEE Trans. Intell. Transp. Syst., vol. 16, no. 3, pp. 1225-1237, Jun. 2014.

[5] J. Pohl, M. Sethsson, P. Degerman, and J. Larsson, "A semi-automated parallel parking system for passenger cars," Proc. Inst. Mech. Eng. D, J. Autom. Eng., vol. 220, no. 1, pp. 53-65, Jan. 2006.

[6] H. G. Jung, Y. H. Cho, P. J. Yoon, and J. Kim, "Scanning laser radar based target position designation for parking aid system," IEEE Trans. Intell. Transp. Syst., vol. 9, no. 3, pp. 406-424, Sep. 2008.

[7] P. Degerman J. Pohl M. Sethson "Ultrasonic sensor modeling for automatic parallel parking systems in passenger cars," SAE 2007 World Congress & Exhibition, Detroit, MI, U.S.A., 16th19th April, 2007 .

[8] K. Fintzel R. Bendahan C. Vestri S. Bougnoux T. Kakinami "3D parking assistant system," Proc. IEEE Intell. Veh. Symp.¡/em¿ pp. 881-886 2004.

[9] N. Kaempchen U. Franke R. Ott "Stereo vision based pose estimation of parking lots using 3d vehicle models" ¡em¿Proc. IEEE Intell. Veh. Symp.¡/em¿ vol. 2 pp. 459-464 2002.

[10] C. Wang,, Hengrun Zhang, Ming Yang, Xudong Wang, Lei Ye and Chunzhao Guo. "Automatic parking based on a bird's eye view vision system," Advances in Mobility Theories, Methodologies, and Applications, vol. 2014 pp. 847406-1-847406-13 2014.

[11] Alex Krizhevsky, IlyaSutskever, and Geoffrey E Hinton,"Imagenet classification with deep convolutional neural networks,"in Advances in neural information processing systems (NIPS), 2012, pp. 1097-1105.

[12] M. Thoma, "A survey of semantic segmentation," CoRR, vol. abs/1602.06541, 2016.

[13] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. G. Rodriguez, "A review on deep learning techniques applied to semantic segmentation," CoRR, vol. abs/1704.06857, 2017.

[14] C. Chen, A. Seff, A. Kornhauser, and J. Xiao. Deepdriving: Learning affordance for direct perception in autonomous driving. In Proceedings of the IEEE International Conference on Computer Vision, pages 2722-2730, 2015.

[15] 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.

[16] H. Xu, Y. Gao, F. Yu, and T. Darrell. End-to-end learning of driving models from large-scale video datasets. arXiv preprint arXiv:1612.01079, 2016.

[17] Lu Chi and Yadong Mu. "Deep Steering: Learning End-to-End Driving Model from Spatial and Temporal Visual Cues". In: arXiv preprint arXiv:1708.03798 (2017).

[18] Y. LeCun, U. Muller, J. Ben, E. Cosatto, and B. Flepp. Offroad obstacle avoidance through end-to-end learning. In NIPS, pages 739-746, 2005.

[19] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MIC- CAI, pages 234-241. Springer, 2015.

[20] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages,770778, 2016.

[21] Vijay John, Seiichi Mita, Hossein Tehrani Niknejad, Kazuhisa Ishimaru,"Automated driving by monocular camera using deep mixture of experts,IV 2017, 10.1109/IVS.2017.7995709.