ENHANCED STATE ESTIMATION USING DEEP LEARNING FOR QUADRUPED ROBOTICS

by

Xuanning Liu

A thesis submitted to Johns Hopkins University

in conformity with the requirements for the degree of

Master of Science in Mechanical Engineering

Baltimore, Maryland

May, 2024

Abstract

This thesis introduces a novel method for state estimation in quadruped robots that leverages proprioceptive sensor data, including readings from one body-mounted inertial measurement unit (IMU) and four additional IMUs attached to the robot's calf links. This sensor setup captures dynamic movements of both the body and legs, complemented by data from joint encoders. An extended Kalman filter (EKF) integrates these observations to estimate the robot's states relative to the world frame. To eliminate the dependence on motion capture systems or other vision-based sensors, this study employs 1D convolutional neural networks (CNNs) to estimate necessary measurements using only proprioceptive data. Experimental results using real-world data from a Unitree Go1 robot validate the effectiveness of our proprioception-based approach, demonstrating that it achieves accuracy comparable to traditional EKF methods that rely on motion capture systems, thus providing a promising alternative for robotic state estimation without the need for external visual inputs.

Thesis Committee

Axel Krieger (Primary Advisor)
Associate Professor
Department of Mechanical Engineering
Johns Hopkins Whiting School of Engineering

Noah J. Cowan
Professor
Department of Mechanical Engineering
Johns Hopkins Whiting School of Engineering

Acknowledgments

I would like to express my deepest gratitude to a number of individuals and organizations whose support was invaluable during the course of my thesis project.

First and foremost, I extend my sincere thanks to Intelligent Fusion Technology, Inc., where I completed my internship. Special thanks go to my project leader, Dr. Yajie Bao, whose guidance and insights have been fundamental to the success of my project. I am also grateful to Dr. Genshe Chen for his leadership and support throughout this journey.

I am deeply grateful to my academic advisor, Dr. Axel Krieger, for his consistent support and encouragement throughout my Master's program. His insights were invaluable, not only in navigating the academic landscape through strategic course selection but also in providing wise counsel regarding career choices and various other challenges I faced. I would also like to express my sincere thanks to my coop program advisor, Dr. Noah Cowan, for his invaluable feedback and guidance.

I am also grateful to all the friends I have made at Johns Hopkins University, whose camaraderie and support have enriched my academic journey. Finally, I must express my utmost appreciation to my parents, Chunying You and Tao

Liu, who have supported me unconditionally. Their love, encouragement, and belief in my abilities have been my constant source of strength and motivation.

Thank you all for your invaluable contributions to my academic journey.

Contributions

Dr. Yajie Bao, the project leader from Intelligent Fusion Technology, Inc., provided invaluable assistance throughout the project. His guidance was instrumental in refining the neural network concepts central to our research. Additionally, Dr. Bao played a crucial role in debugging the implementation and steering the project's direction. His expertise and insights have significantly shaped the outcome of this work.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

In recent years, quadruped robots have gained significant popularity due to their agility on challenging terrains and their capability to execute complex tasks. Outstanding examples include Boston Dynamics' Spot, which has been used for industrial inspections and public safety (Bouman et al., 2020); ANYmal by ANYbotics, known for its deployment in hazardous environments for maintenance and emergency response (Hutter et al., 2017); and the Unitree Go1, which is acclaimed for personal assistance and entertainment purposes (Romanov, Gyrichidi, and Romanov, 2023). These robots exemplify the advanced mobility and adaptability that quadrupeds can offer across diverse applications, they are integral to advancements in automated systems, where they are employed in diverse scenarios ranging from disaster response and industrial automation to planetary exploration and healthcare assistance. Their ability to maneuver through rubble, climb steep slopes, and interact with environments designed for humans makes them particularly valuable in fields where human intervention is risky or impractical (Biswal and Mohanty, 2021).

Figure 1.1: Examples of quadruped robots: Spot[1], ANYmal[2], and Unitree Go2[3]

Perception, planning, and control form the three foundational pillars of robotic autonomy, each facilitating distinct but interconnected functions that enable robots to operate independently and effectively in diverse environments. Perception involves the real-time acquisition and processing of sensory data to discern the robot's surroundings and its own state within that environment. This capability is critical for recognizing objects, navigating spaces, and understanding terrain. For example, a robotic vacuum uses sensors to detect walls and obstacles, while autonomous vehicles employ Lidar and cameras to map their surroundings and detect other users on the road (Guastella and Muscato, 2021). Planning refers to the process by which a robot uses the information obtained from perception to formulate strategies for achieving its goals. This includes determining the sequence of actions or movements necessary to navigate from one point to another while avoiding obstacles, optimizing paths, or strategizing interaction with objects and other agents. An industrial robot arm, for instance, plans its movements to assemble parts

[1] https://bostondynamics.com/solutions/inspection/thermal/

[2] https://www.dpaonthenet.net/article/199317/Robots-to-team-up-on-trip-to-the-Moon.aspx

[3] https://eu.robotshop.com/products/roboworks-unitree-go2-ent

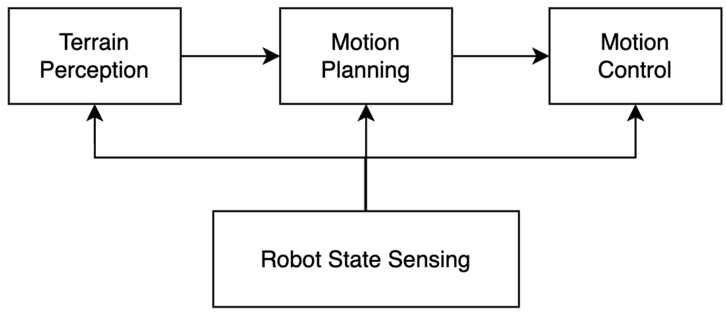

Figure 1.2: Robotics principle flowchart

based on the perceived locations of these components. Control is the execution phase where the robot applies the planned actions to interact with the environment. This involves precise manipulation of the robot's mechanisms, such as motors and other actuators, to achieve desired outcomes. Control systems ensure that movements are carried out smoothly and adjust dynamically to any changes in the environment or internal conditions. For example, drones control their flight patterns through rapid adjustments to rotor speeds, enabling stable flight and agile maneuvers in response to wind or obstacle avoidance commands. Together, these pillars enable a robot to perform complex tasks autonomously by integrating sensory data with decision-making processes and physical actions, thereby adapting to new environments and challenges.

Within the domain of perception, state estimation plays a pivotal role by providing critical information about the robot's location, orientation, and the kinematic states of its various components, such as limbs and joints. Accurate state estimation enables a robot to determine its body configuration in space, assess its stability, and adapt its movements to environmental changes and internal dynamics. Effective state estimation enhances navigational accuracy, allowing the robot to make informed decisions and improve its ability to navigate complex environments. It also supports robust motion planning by providing accurate data on joint angles and limb positions, which is essential for developing dynamic motion plans that can adapt to both planned tasks and unforeseen obstacles. Furthermore, state estimation is critical for maintaining stability and safety, as it provides vital feedback on the robot's balance and stability metrics, which are crucial for preventing falls and ensuring safe interactions with both the environment and human operators (Bloesch, 2017). Additionally, by improving the reliability of internal sensors and processing methods, robots can operate independently in environments where external systems like GPS or visual markers are unavailable.

The challenges of state estimation for quadruped robots are notably distinct due to their complex locomotion dynamics and variable interaction with the environment. Unlike wheeled robots, quadrupeds must constantly adjust to varied terrain types—from flat, predictable surfaces to irregular, off-road conditions—which requires continuously updating their body orientation and foot placement for balance and propulsion (Bloesch, 2017; Bahçeci and Erbatur, 2023). The need to maintain stability while navigating such uneven terrains complicates the state estimation process. Moreover, the integration of multiple sensors to achieve reliable estimation introduces issues of sensor fusion complexity and the potential for conflicting data, particularly under dynamic conditions where rapid movements or external disturbances occur. Additionally, the real-time processing demands for state estimation in quadrupeds are significant, as delays or inaccuracies in data interpretation can lead to falls or navigational errors, posing risks to both the robot and its surroundings.

1.2 Overview of Existing Methods

This thesis examines the field of state estimation for robotics, with a particular focus on quadruped robots. State estimation is critical for autonomous navigation and interaction within various environments, utilizing algorithms to deduce a robot's state—such as position, velocity, and orientation—from sensor data. Traditional methods, primarily various forms of the Kalman filter, have been foundational in addressing the complex dynamics of quadruped robots (Chen, 2012). These methods integrate data from multiple sensors to improve accuracy despite noisy inputs.

Inertial measurement units (IMUs) play a crucial role in this domain by providing acceleration and angular velocity from accelerometers and gyroscopes, particularly useful in environments where external references are unavailable. However, IMUs can suffer from data drift, leading to accumulating errors (Bloesch et al., 2013). Recent advances in machine learning, especially in learning-based sensor fusion using methods like convolutional neural networks (CNNs), offer promising ways to overcome these limitations (BROSSARD and BONNABEL, 2019). These innovative approaches can enhance the robustness and accuracy of state estimations under dynamic conditions.

The Extended Kalman Filter (EKF) remains a staple in robotic state estimation, adept at handling non-linear system dynamics through linearization techniques. Yet, its effectiveness can diminish with large non-linearities or inaccurate model assumptions, paving the way for hybrid approaches that integrate EKF with modern machine learning techniques for improved performance.

This review sets the stage for introducing a novel state estimation framework in this thesis, which combines the precision of traditional methods with the adaptability of machine learning to better support autonomous navigation in quadruped robots. The primary aim of this research is to advance the field of robotic state estimation through the development and validation of a novel deep learning approach, specifically focusing on the following objectives:

- Develop a Proprioception-based Deep Learning Model: To implement a 1D convolutional neural network (CNN) that utilizes proprioceptive sensor data from inertial measurement units (IMUs) to estimate the yaw angles of a quadruped robot. This model aims to leverage the temporal nature of IMU data to enhance the estimation accuracy beyond what is achievable with traditional filtering techniques such as the Extended Kalman Filter (EKF).

- Evaluate Model Performance: To test the developed model using real-world data collected from a Unitree Go1 robot. This evaluation will compare the performance of CNN against traditional state estimation methods that rely on motion capture systems. The criteria for comparison will include accuracy, reliability, and computational efficiency, with the goal of demonstrating that CNN can provide a viable alternative for applications where external visual inputs are unavailable.

Through these objectives, this thesis seeks to contribute a robust solution for state estimation that could potentially reduce the dependence on external sensors and improve the operational versatility of autonomous robots.

1.3 Thesis Structure

The structure of this thesis is designed to provide a comprehensive overview of the research conducted, as well as detailed insights into the methodologies employed and the findings obtained. The thesis is organized into six chapters, as follows:

Chapter 1: Introduction - This chapter introduces the research topic, outlines the research problem, and discusses the significance of the study. It sets the stage for the subsequent chapters by providing a background on the need for advanced state estimation techniques in robotic systems.

Chapter 2: Literature Review - A review of the existing literature related to state estimation techniques in robotics, focusing particularly on methods employing proprioceptive sensors and deep learning models. This chapter highlights the strengths and limitations of current approaches and justifies the need for the proposed research.

Chapter 3: Methodology - Detailed description of the research methodology, including the details of the 1D convolutional neural network, the data acquisition process using the Unitree Go1 robot, and the methods used for data processing and model training.

Chapter 4: Experimental Setup - Explanation of the experimental setup, including the configuration of the physical and software environments, the calibration of sensors, and the procedures followed during the experimental trials.

Chapter 5: Results and Discussion - Presentation and analysis of the results obtained from the experimental validation. This chapter assesses the performance of the CNN model in comparison to traditional methods, discussing the implications of the findings in the context of robotic perception and state estimation. This chapter also outlines potential avenues for future research, suggesting enhancements to the model and additional applications in other areas of robotics.

Chapter 2

Literature review

This chapter explores the current state of research in the field of state estimation for robotics, particularly focusing on quadruped robots. It covers key areas including traditional state estimation approaches, the role of inertial measurement units (IMUs), and recent advancements in learning-based sensor fusion techniques. This review provides a foundation for understanding the context and technical challenges addressed in this thesis.

2.1 Legged Robot State Estimation

State estimation in legged robots is a critical component for achieving autonomous navigation and interaction within complex environments. Unlike wheeled robots, legged robots encounter a variety of terrains that impose unique demands on the estimation process, necessitating sophisticated methods to accurately predict and respond to dynamic conditions.

Legged locomotion introduces complex variables into state estimation due to the intermittent nature of foot contacts and the diverse properties of different terrains. Fahmi, Fink, and Semini, 2021 highlight the impact of soft terrain on state estimation, where traditional methods designed for rigid surfaces fail to account for the additional uncertainties introduced by softer materials, leading to increased estimator drift. This underlines a key challenge: adapting state estimation methods to function reliably across variable terrain types.

Proprioceptive sensors, such as inertial measurement units (IMUs) and joint encoders, are integral to the state estimation of legged robots. These sensors provide crucial data on the robot's motion and orientation, which are vital for maintaining balance and navigating environments without external visual aids. The research by Yang et al., 2019 demonstrates the efficacy of integrating proprioceptive sensor data through the Contact-Centric Leg Odometry (COCLO) method, which enhances estimation precision by focusing on leg contact dynamics. This approach shows promise in improving state estimation accuracy over IMU-centric methods, particularly in challenging terrains.

The fusion of data from multiple sensory sources marks a significant advancement in state estimation technology. For example, the VILENS system described by Wisth, Camurri, and Fallon, 2022 integrates visual, inertial, lidar, and leg odometry data using a factor graph approach. This multimodal fusion not only compensates for the weaknesses of individual sensors but also significantly reduces the estimation errors associated with leg odometry, particularly in adverse conditions. This demonstrates the potential of hybrid sensor systems to provide robust state estimation solutions that can adapt to a range of environmental challenges.

The field of state estimation for legged robots is evolving, with ongoing research focused on overcoming the inherent challenges posed by complex and varied terrains. By leveraging advances in sensor technology and computational methods, such as machine learning and sensor fusion, researchers are paving the way for more autonomous and effective robotic systems capable of operating in unpredictable settings.

2.2 State Estimation Using Proprioceptive Sensors

Proprioceptive sensors, such as IMUs and joint encoders, are fundamental in the state estimation of quadruped robots, particularly when external sensory inputs are limited or unavailable. These sensors provide critical data on internal state dynamics, which are essential for the autonomous operation of robots in diverse environments.

A significant advancement in this area is the development of the invariant extended Kalman filter (EKF) by Barrau and Bonnabel, 2015; Zhang et al., 2023. This method enhances the robustness and accuracy of state estimation in dynamic environments by leveraging the mathematical properties of the invariant EKF. This approach specifically addresses the challenges of limited accuracy and high noise levels typical of proprioceptive sensors, making it a potent tool for complex robotic applications.

Camurri et al., 2017 introduced a probabilistic method to estimate contact and detect impact events using internal force sensors, which marks a pivotal shift from traditional external contact sensors. This technique significantly improves the reliability of state estimation by providing precise information on contact points with the environment, which is crucial for executing dynamic maneuvers and adapting to unexpected terrain changes.

The use of factor graph optimization in state estimation frameworks, as explored by Wisth, Camurri, and Fallon, 2019, offers another robust approach. This method integrates various proprioceptive data sources to maintain accurate robot state awareness, particularly in environments where external sensors are prone to failure or provide unreliable data. Such optimization techniques are vital for enhancing the stability and reliability of state estimations across challenging conditions.

While model-based approaches using proprioceptive sensors are advancing, they still face significant challenges, including the integration of disparate sensor data and the need for precise sensor calibration.

2.3 Learning-Based Sensor Fusion Approaches

Sensor fusion is a critical technique for enhancing the robustness and accuracy of state estimation in quadruped robots. By combining data from multiple sources, these systems can compensate for the limitations of individual sensors, particularly in challenging environments where one type of sensor might be insufficient or unreliable. This section explores several innovative approaches to learning-based sensor fusion in quadruped state estimation.

A prominent example of advanced sensor fusion is the VILENS system mentioned above. The primary goal of VILENS is to ensure reliable operation even when individual sensor modalities fail or underperform due to environmental conditions. Extensive validation has demonstrated that VILENS significantly reduces translational and rotational errors across various terrains, underscoring the benefits of a multimodal fusion approach in achieving high-fidelity state estimation.

The calibration of multiple cameras on a legged robot is another area where sensor fusion plays a vital role. Reinke, Camurri, and Semini, 2019 address this by implementing a factor graph approach to optimize the extrinsic calibration between multiple cameras and the robot base. This method not only improves the accuracy of visual feedback but also enhances overall sensor fusion, contributing to more precise navigation and interaction within complex environments.

Yao and Jia, 2021 present a state estimation algorithm that exemplifies the integration of a wide array of sensors, including IMUs, joint encoders, cameras, and LIDAR. This comprehensive fusion is designed to provide robust state estimation critical for maintaining stable locomotion over varied terrains. By pooling information from these diverse sources, the algorithm enhances the robot's ability to adapt to different environmental conditions, improving both the reliability and accuracy of its navigational capabilities.

Despite the advancements in sensor fusion technologies, several challenges remain, particularly concerning the computational demands and the real-time processing capabilities required to handle data from multiple sensors simultaneously.

Building on the foundational work explored, our approach to state estimation in quadruped robots proposes a novel method that leverages advanced sensor fusion and machine learning techniques. Our goal is to overcome the existing limitations noted in traditional and learning-based sensor fusion methods by introducing an innovative system design that optimizes accuracy, robustness, and computational efficiency. We choose proprioceptive sensors over vision-based methods because they provide consistent and reliable data under conditions where optical sensors might be impaired by poor lighting, occlusions, or reflective surfaces. This reliance on proprioceptive data allows for more stable and continuous state estimation, essential for navigating unpredictable or complex environments where visual cues are limited or deceptive.

Chapter 3

Methodology

3.1 Learning-assisted Multi-IMU Proprioceptive State Estimation

In the pioneering work of Yang et al., 2023, the Extended Kalman Filter (EKF) is employed as the state estimator, significantly enhancing state estimation by incorporating dynamic foot movements and multi-sensor integration. This methodology offers robustness against the oversimplified assumptions inherent in previous models, providing a sophisticated and reliable framework for robotic navigation, especially in complex environments where interaction dynamics are paramount.

However, a notable limitation of the model arises from its reliance on yaw angle measurements provided by a motion capture system. These measurements are critical because yaw angles are not observable solely from the body IMU data (Bloesch, 2017). The dependency on motion capture systems, which are typically unavailable in outdoor environments or general experimental setups, restricts the model's applicability. To address this limitation,

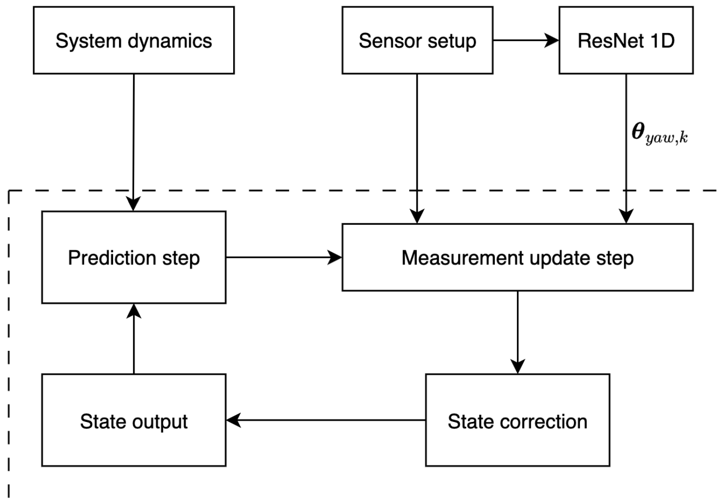

Figure 3.1: The schematic of the proposed method

our method employs a learning-based approach to predict yaw angles from available data sources, including IMUs, joint motor encoders, and foot contact sensors. This predictive capability enables the replacement of direct motion capture measurements, thus fitting seamlessly within the multi-IMU odometry framework and broadening the model's applicability to outdoor and less controlled environments.

This chapter outlines the comprehensive methodology employed in this study to develop and validate a learning-assisted state estimation framework for quadruped robots as shown in 3.1. The methodology is divided into four main sections: system design and data collection, neural network architecture, Extended Kalman Filter (EKF) implementation, and analysis of outputs.

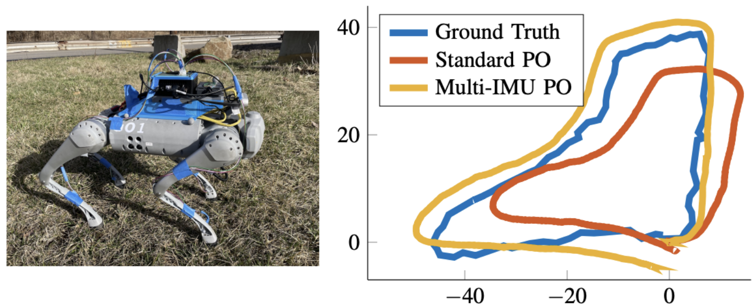

Figure 3.2: Left: A Unitree Go1 robot equipped with foot IMUs. Right: Position estimates while walking over a 160m loop trajectory. (Yang et al., 2023)

3.2 System Design and Data Collection

The system design and setup for the Multi-IMU Proprioceptive Odometry (MIPO) introduced by Yang et al., 2023. focus on enhancing the state estimation of legged robots using proprioceptive sensors, specifically in scenarios where external systems like GPS are not available. Their novel proprioceptive sensing solution adds an additional IMU to each calf link of the robot, just above the foot, alongside the conventional sensors of one body IMU and joint encoders.

The inclusion of these additional IMUs allows the system to determine foot contact modes and detect slips without the need for tactile or pressure-based foot contact sensors. The Extended Kalman Filter (EKF) fuses data from all sensors to estimate the robot's body and foot positions and velocities in the world frame.

The MIPO system achieves accurate, low-drift, long-term position and velocity estimation while requiring minimal additional hardware and computational resources, a significant improvement over conventional methods.

The proposed approach, validated through hardware experiments, is demonstrated to reduce position drift by nearly an order of magnitude compared to traditional proprioceptive odometry (PO) approaches as shown in Figure 3.2.

The MIPO design strategically places additional IMUs without substantially altering the robot's form factor. These IMUs were synchronized to operate at the same frequency as the robot's built-in sensors. Data from the IMUs and other sensors were collected and processed on an Intel NUC minicomputer, which also ran the MIPO algorithm and a nonlinear predictive control locomotion controller.

Building upon the foundational work of Yang et al., 2023, this thesis advances the system design for state estimation in legged robots by refining the proprioceptive odometry process. Enhancing the Multi-IMU Proprioceptive Odometry (MIPO) system, this research integrates a novel array of proprioceptive sensors, strategically augmenting the existing IMU and encoder setup with advanced sensor fusion techniques. This augmented system employs an innovative extended Kalman Filter (EKF) that incorporates dynamic foot movement data and improves upon the integration strategy of multisensor inputs.

Through experimentation and analysis, this enhanced MIPO system demonstrates superior performance in state estimation accuracy and robustness compared to its predecessors. The modifications detailed in this work not only provide a higher fidelity in state estimation but also present a viable solution for legged robots operating in GPS-denied environments.

3.3 Neural Network Architecture

The core of the enhanced state estimation system is a 1D convolutional neural network (CNN) designed to infer yaw angles from integrated sensor data Hong et al., 2020. The CNN architecture is composed of the following layers:

- **Input Layer:** This layer accepts a time series of normalized sensor data vectors, each vector encapsulating IMU and joint encoder readings.
- **Convolutional Layers:** Multiple convolutional layers with ReLU activation functions are used to extract and learn features from the input data. The layers use kernels of varying sizes to capture both short and long-term dependencies in the data.
- **Pooling Layers:** Max pooling layers follow convolutional layers to reduce dimensionality and computational complexity.
- **Dropout Layers:** Dropout is employed intermittently to prevent overfitting by randomly dropping units during the training process.
- **Fully Connected Layers:** Dense layers follow the convolutional and pooling layers to interpret the features extracted and to output the estimated yaw angles.

The network outputs a series of values representing the yaw angle, which is then integrated into the EKF. The CNN is trained using backpropagation with a mean squared error loss function to minimize the difference between the estimated and true yaw angles obtained from ground truth data. For a detailed description of the CNN architecture and training parameters, refer to Section 4.2 and the Table 5.2 in Appendix.

3.4 Kalman Filter Implementation

3.4.1 Kalman Filter

The Kalman Filter operates with two main steps: Prediction and Update (Welch, Bishop, et al., 1995).

Prediction:

$$\mathbf{x}_{k+1|k} = \mathbf{A}\mathbf{x}_{k|k} + \mathbf{B}\mathbf{u}_k + \mathbf{n}_k \tag{3.1}$$

where $\mathbf{A}$ is the state transition matrix, $\mathbf{x}_{k|k}$ is the estimated state from the previous timestep, $\mathbf{B}$ is the control-input matrix applied to the control vector $\mathbf{u}_k$, and $\mathbf{n}_k$ represents the process noise.

Covariance Prediction:

$$\mathbf{P}_{k+1|k} = \mathbf{A}\mathbf{P}_{k|k}\mathbf{A}^T + \mathbf{Q} \tag{3.2}$$

where $\mathbf{P}_{k|k}$ is the covariance of the estimate at time k and $\mathbf{Q}$ is the process noise covariance matrix.

Update:

$$\mathbf{K}_k = \mathbf{P}_{k+1|k}\mathbf{H}^T(\mathbf{H}\mathbf{P}_{k+1|k}\mathbf{H}^T + \mathbf{R})^{-1} \tag{3.3}$$

$$\mathbf{x}_{k+1|k+1} = \mathbf{x}_{k+1|k} + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}\mathbf{x}_{k+1|k}) \tag{3.4}$$

$$\mathbf{P}_{k+1|k+1} = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_{k+1|k} \tag{3.5}$$

where $\mathbf{K}_k$ is the Kalman gain, $\mathbf{z}_k$ is the measurement at time k, $\mathbf{H}$ is the measurement matrix that maps the true state space into the observed space, and $\mathbf{R}$ is the measurement noise covariance matrix. $\mathbf{I}$ represents the identity matrix of appropriate size.

These equations establish a framework for predicting and correcting the state of a dynamic system, utilizing a model of the system's dynamics and measurements. The Kalman Filter iteratively updates its estimates of the state and error covariances to incorporate new measurements, thereby refining the state estimates over time.

3.4.2 Extended Kalman Filter incorporating Multi-IMU

The Multi-IMU Proprioceptive Odometry (MIPO) method enhances the Extended Kalman Filter (EKF) by integrating data from multiple IMUs positioned near the robot's feet, offering significant improvements over the simple Kalman Filter (KF). Unlike the simple KF, which only handles linear dynamics, the EKF adeptly manages non-linear system behaviors crucial for legged robots' varied joint movements and terrain interactions (Moore and Stouch, 2016). This multiple IMU approach not only refines the granularity and accuracy of state estimations but also increases the robustness of the EKF, enabling more precise adaptations to complex environments and dynamic conditions.

In the Multi-IMU enhanced EKF, the state vector is defined as $x = [p; v; \theta; s_j; \dot{s}_j]$, where p is the robot's position, v its linear velocity, θ the orientation angles,

s_j the foot position of leg j, and $\dot{s}_j$the foot velocity. All the positions are represented in the world frame. The filter updates the state with a prediction model using foot IMU data and a novel measurement model that leverages foot IMU data for contact and slip detection.

The process model for the EKF is given by:

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k) + \mathbf{n}_k \tag{3.6}$$

where $\mathbf{f}(\cdot)$ is the prediction function and $\mathbf{n}_k$ represents the process noise.

The EKF's prediction step for state transition from time k to $k+1$ is:

$$\mathbf{x}_{k+1|k} = \begin{bmatrix} \mathbf{p}_k + \Delta t \mathbf{v}_k \\ \mathbf{v}_k + \Delta t(\mathbf{R}(\boldsymbol{\theta}_k)\mathbf{a}_b - \mathbf{g}_w) \\ \boldsymbol{\theta}_k + \Delta t \boldsymbol{\Omega}(\boldsymbol{\theta}_k)\boldsymbol{\omega}_b \\ \mathbf{s}_k + \Delta t \dot{\mathbf{s}} \\ \dot{\mathbf{s}}_k + \Delta t(\mathbf{R}(\boldsymbol{\theta}_k)\mathbf{R}_b^f(\boldsymbol{\phi})\mathbf{a}_f - \mathbf{g}_w) \end{bmatrix} \tag{3.7}$$

where Δt is the time step, $\mathbf{R}(\cdot)$ is the rotation matrix, $\mathbf{a}_b$ and $\mathbf{a}_f$ are the linear acceleration measurements from the body and foot IMUs, $\boldsymbol{\omega}_b$ is the angular velocity from the body IMU, $\mathbf{g}_w$ is the gravitational acceleration, $\boldsymbol{\Omega}(\cdot)$ relates the derivative of the Tait-Bryan angles to the angular velocity, and $\mathbf{R}_b^f(\cdot)$ transforms the acceleration from the foot frame to the body frame.

The EKF measurement model is:

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k, \boldsymbol{\phi}, \boldsymbol{\omega}_f) = \begin{bmatrix} \mathbf{R}(\boldsymbol{\theta}_k)^T(\mathbf{s}_k - \mathbf{p}_k) \\ \mathbf{R}(\boldsymbol{\theta}_k)^T(\mathbf{v}_k - \dot{\mathbf{s}}_k) \\ \dot{\mathbf{s}}_k - \boldsymbol{\omega} \times \mathbf{d} \end{bmatrix} \tag{3.8}$$

where $\mathbf{d}$ is defined as $-d_0 \cdot \frac{\mathbf{n}}{\|\mathbf{n}\|}$, with d_0 being the distance between the foot center and the foot surface, and $\mathbf{n}$ is the contact normal vector.

For further details, refer to the work from Yang et al., 2023.

3.5 Analyzing Outputs

The state estimation process outcomes are assessed to gauge the neural network model's performance in contrast with conventional EKF methodologies lacking CNN integration. Metrics such as the Root Mean Square Error (RMSE) and maximum RSE (maxRSE) are computed to detail the precision of the estimations. These metrics prove the effectiveness of the CNN-augmented approach under diverse operational conditions and its feasibility as a substitute for traditional methods in practical scenarios.

Yaw predictions from the CNN are illustrated, and metrics such as the IMU drift percentage, RMSE, and Mean Absolute Error (MAE) are derived to evaluate performance. Furthermore, the state estimation process yields calculated and plotted data regarding the robot's orientation and position, providing a comprehensive understanding of the robot's navigational status. The extensive evaluation, including statistical testing and visual comparisons against multiple datasets, is critical in affirming the proposed method's robustness for autonomous robotic navigation. For a full exposition of these findings, refer to the Results chapter detailed later in this document.

Chapter 4

Experimental setup

4.1 Experiments and Software Configuration

The experimental setup for this thesis utilizes a dataset collected from an indoor environment to thoroughly evaluate the proposed state estimation techniques. This dataset was derived from experiments conducted in a controlled laboratory space equipped with a high-precision motion capture system, providing accurate ground truth data for validation purposes.

On the software front, the convolutional neural network (CNN) is implemented in Python 3.10, utilizing a 1D-ResNet architecture specifically tailored for time-series analysis. This implementation, based on the ResNet model originally developed by He et al., 2015, has been adapted for time-series analysis by Hong et al., 2020, who optimized it for performance with Python 3.7.5 and PyTorch 1.2.0. The detailed model summaries are facilitated by torchsummary, providing critical insights into the model architecture and performance. The state estimation component is developed in MATLAB, leveraging the Casadi library (version 3.5.5) designed by Andersson et al., 2019 to enable complex algorithmic calculations and optimizations. The Extended Kalman Filter (EKF) used for state estimation, a key component of the system, was extensively developed in the work by Yang et al., 2023. This comprehensive setup ensures that the system's performance can be accurately assessed across a wide range of scenarios, reflecting its potential for practical deployment in autonomous robotic navigation.

4.2 Dataset description

The dataset utilized in this study was derived from experiments conducted using a Unitree Go 1 quadruped robot, as reported by Yang et al., 2023. This robot featured a suite of sensors including a body-mounted MEMS Inertial Measurement Unit (IMU), twelve motor encoders at the leg joints providing real-time torque and angular data, and four foot pressure contact sensors for detailed ground interaction feedback. Each foot of the robot was equipped with an MPU9250 IMU, and an Arduino Teensy board managed the synchronization and transmission of data from these IMUs to the central processing unit.

Data collection occurred indoors in a controlled laboratory setting with a high-precision motion capture (MoCap) system to ensure data fidelity and to provide ground truth for algorithm validation. The robot operated at movement speeds ranging from 0.4 to 1.0 m/s across flat terrain, employing trotting or flying trotting gaits. Sensor data, comprising three-dimensional linear accelerations, angular velocities, torque, and angular measurements, were recorded at a high frequency of 200 Hz. Orientation and position data were captured in quaternion and 3D coordinate forms, respectively, by the MoCap system. All sensors' data were initially resampled to unify the sampling frequency, synchronized, normalized, and archived in rosbag file format, accessible on the author's GitHub page.

The dataset encompassed trials of approximately 45 seconds each on simple, flat terrain. Given the brief duration and the controlled environment, IMU drift—common in mobile robotics—was minimal. However, to simulate deteriorative effects on sensor accuracy over longer periods and varied terrains, Gaussian noise and cumulative Gaussian noise drift were synthetically added to the IMU data. This artificial augmentation helped create a robust framework for evaluating state estimation algorithms under realistic conditions.

In this study, the MoCap system's pose measurements provided a high-fidelity ground truth. The Multi-IMU Proprioceptive Odometry (MIPO), using multiple foot-mounted IMUs, served as the baseline for comparison with other state estimation methods. The integration of body IMU data served as another baseline, where angular velocity measurements were integrated over time to estimate orientation; however, due to IMU drift, this method proved less reliable.

The specifics of the number of trials and initial conditions were not extensively detailed in the source study by Yang et al. This omission limited the depth of information available, which in turn may have influenced the comprehensiveness of our findings. The lack of detailed trial data restricts our understanding of the variability and reproducibility of the results, thus posing a significant limitation on the generalizability and applicability of our conclusions to other settings or conditions.

4.3 Neural Network Details

To address the limitations posed by dataset variability and enhance the model's generalizability across different robotic movements, we implement two neural network-based methods using a 1D ResNet architecture (Hong et al., 2020). Both approaches utilize approximately 3000 randomly selected data points for training. The first neural network method is the Multi-IMU CNN Angle Estimator (MI-CAE), which predicts yaw angles at two consecutive time steps simultaneously, using mocap-derived orientation measurements as ground truth, incorporating the differences between these predictions and IMU integrations directly into the network as a regulatory mechanism. This method is designed to mitigate the cumulative error introduced by IMU drift by focusing on short time intervals between predictions. The second neural network approach is the Multi-IMU CNN angle Correction Enhancer (MI-CCE), which aims to predict a correction factor for the IMU-derived yaw angles, with the ground truth being the discrepancy between the IMU integrations and mocap measurements. The corrected yaw angles are then integrated into the MIPO framework to enhance state estimation further. The efficacy of these methods is subsequently evaluated by comparing the resulting position estimates against the mocap system data, thus ensuring a robust assessment of each approach's ability to improve upon traditional IMU integration techniques in dynamic environments.

We implemented a 1-dimensional Residual Network (ResNet1D) model for the regression analysis of time-series data. The architecture is tailored to process sequences by leveraging a convolutional neural network framework, which is specifically designed to capture temporal dependencies and intricate patterns in the data. The model's input layer is configured to accept a number of features per time step, matched to the dataset's feature count, and utilizes 256 base filters in its convolutional layers. These layers utilize a kernel size of 5 and a stride of 2, optimizing the model to efficiently reduce data dimensionality while retaining critical information. The ResNet1D structure comprises 9 residual blocks, each containing a sequence of convolutional layers, batch normalization, and ReLU activation functions. The output of the model is configured to produce a single continuous value, which presents the predicted yaw angles or the correction factor for IMU yaw angle measurements.

From a training perspective, the model employs the Adam optimizer with an initial learning rate of 0.001 and a weight decay of 0.001 to prevent overfitting. Additionally, a ReduceLROnPlateau learning rate scheduler is integrated to adjust the learning rate based on the performance, specifically reducing the rate if no improvement in loss is observed over 10 epochs. This is complemented by the use of Mean Squared Error Loss (MSELoss) as the loss function. The entire model is trained over 50 epochs.

Chapter 5

Results and Discussion

In this chapter, the performance of two enhanced Multi-IMU Proprioceptive Odometry (MIPO) models for state estimation in quadruped robots is analyzed and compared with existing methods. Each model utilizes convolutional neural networks (CNN) to enhance state estimation accuracy by integrating proprioceptive sensor data from IMUs.

5.1 Comparative Performance Metrics

Table 5.1 presents the performance metrics for the various state estimation methods tested. The Multi-IMU CNN Angle Estimator (MI-CAE) uses CNN-predicted yaw angles derived from multiple IMU measurements for state estimation, resulting in an average drift of 15.65%, a median drift of 15.87%, an RMSE of 0.304946, and a maximum RSE of 0.855203. The Multi-IMU CNN angle Correction Enhancer (MI-CCE) improves upon this by using a CNN to predict an angle correction factor from multiple IMU measurements, which is then used to correct the IMU data before performing state estimation. This method exhibits slightly lower drift percentages and error metrics, with an average drift of 15.65%, a median drift of 15.87%, an RMSE of 0.304946, and a maximum RSE of 0.855203.

The "mocap (MIPO)" method, the original approach presented by Yang et al., 2023, employs angle information from a motion capture system and shows the lowest drift and error metrics among the methods using IMUs, with an average drift of 14.66%, a median drift of 14.87%, an RMSE of 0.286135, and a maximum RSE of 0.788676. In contrast, the "IMU integration" method, which calculates the angle directly from IMU measurements without any correction or enhancement, performs the poorest, with the highest recorded drift percentages and RMSE of 0.312954 and a maximum RSE of 0.981696.

Method	median drift %	RMSE	max RSE	use mocap
MI-CAE	15.87	0.304946	0.855203	No
MI-CCE	14.77	0.290946	0.757274	No
mocap (MIPO)	14.87	0.286135	0.788676	Yes
IMU integration	15.37	0.312954	0.981696	No

Table 5.1: Summary of results for various state estimation methods.

The performance results illustrate that all three specialized methods, MI-CAE, MI-CCE, and MIPO, surpass the IMU integration approach in terms of lower drift percentages and error metrics. Notably, while the MIPO method uses angle information from a motion capture system and achieves the lowest error metrics, the CNN-enhanced methods, MI-CAE and MI-CCE, exhibit similar levels of accuracy without the reliance on external motion capture systems.

5.2 Visual Representations of Model Performance

In this section of the results chapter, a series of figures illustrate the performance of the state estimation models under various conditions. These figures are crucial for visually representing the efficacy of different CNN models and sensor configurations in estimating orientation and position, as well as comparing their output against ground truth data. Figures are grouped according to the model and sensor setup to provide a clear comparative analysis.

5.2.1 Orientation and Position Estimation Results

Figures illustrating orientation and position estimation results are presented to compare the efficacy of multi-IMU setup across different models.

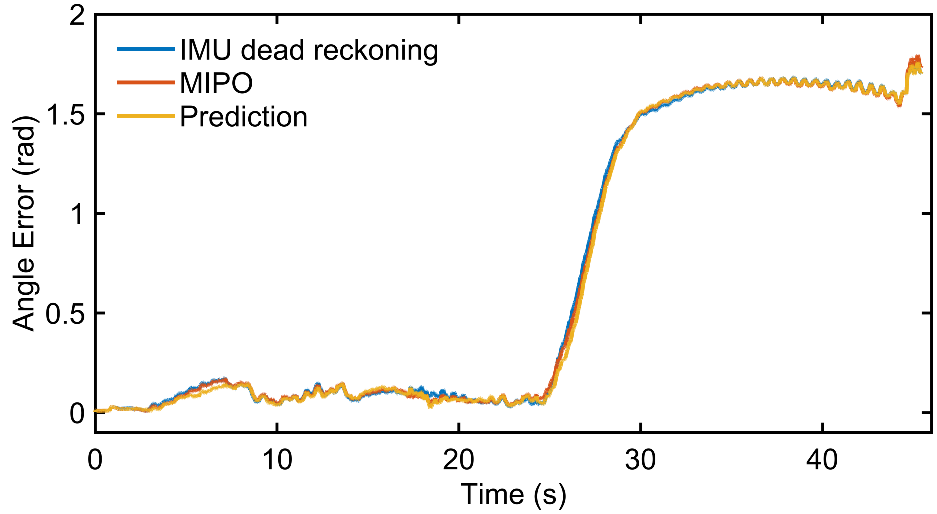

Figure 5.1: Orientation results for Multi-IMU CNN Angle Estimator (MI-CAE)

5.2.2 CNN Prediction Outputs and Yaw Predictions

This subset of figures showcases the CNN prediction outputs and yaw predictions, illustrating the precision of the CNN models in predicting yaw angles against ground truth data.

These figures are integral for demonstrating the models' accuracy and their potential for practical deployment in autonomous robotic navigation without reliance on external motion capture systems.

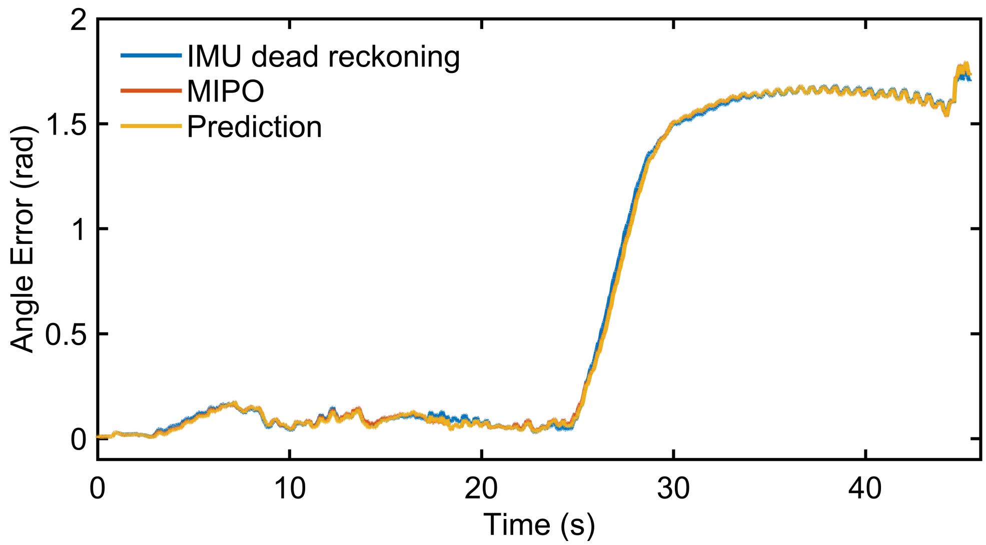

Figure 5.2: Orientation results for Multi-IMU CNN angle Correction Enhancer (MI-CCE)

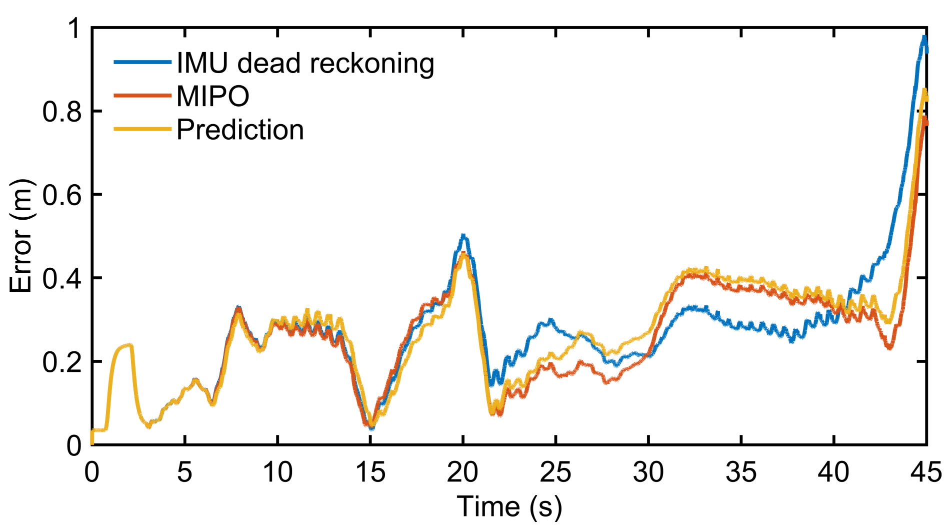

Figure 5.3: Position results for Multi-IMU CNN Angle Estimator (MI-CAE)

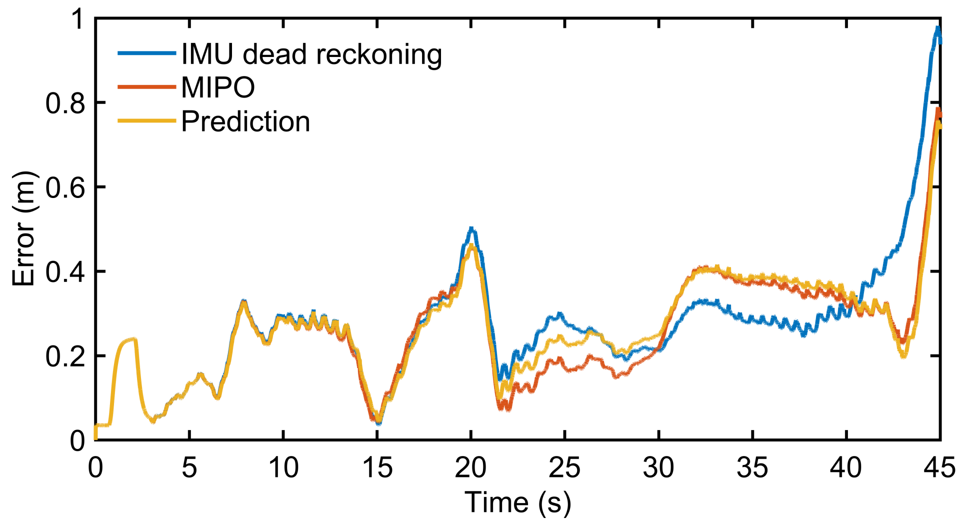

Figure 5.4: Position results for Multi-IMU CNN angle Correction Enhancer(MI-CCE)

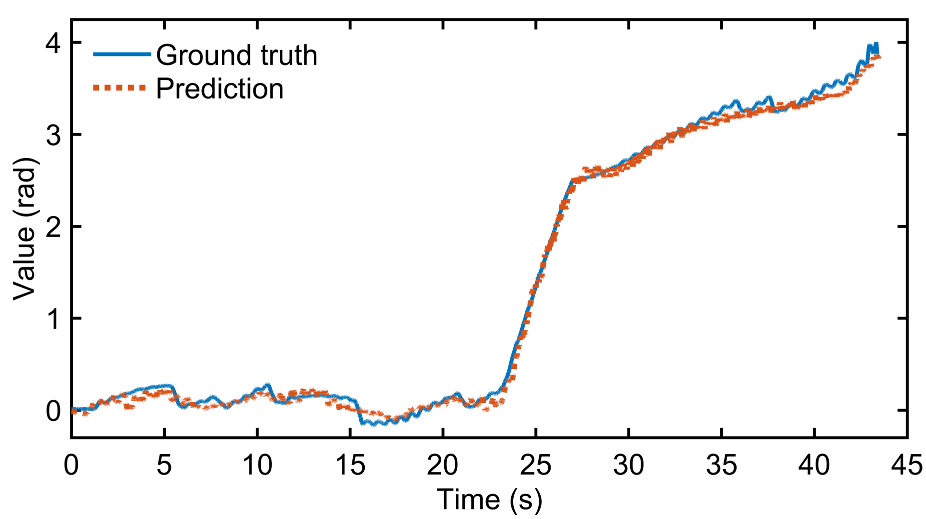

Figure 5.5: Yaw prediction for Multi-IMU CNN Angle Estimator (MI-CAE)

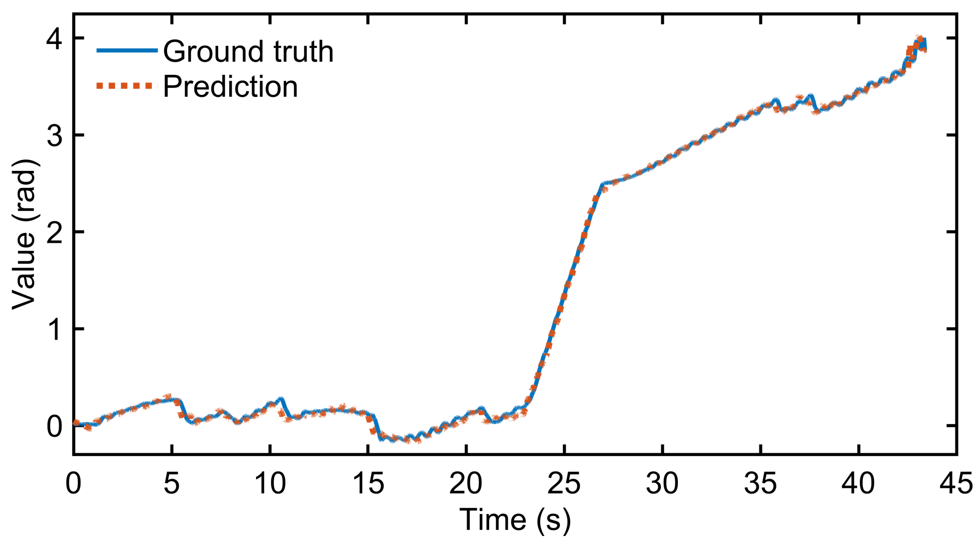

Figure 5.6: Yaw prediction for Multi-IMU CNN angle Correction Enhancer (MI-CCE)

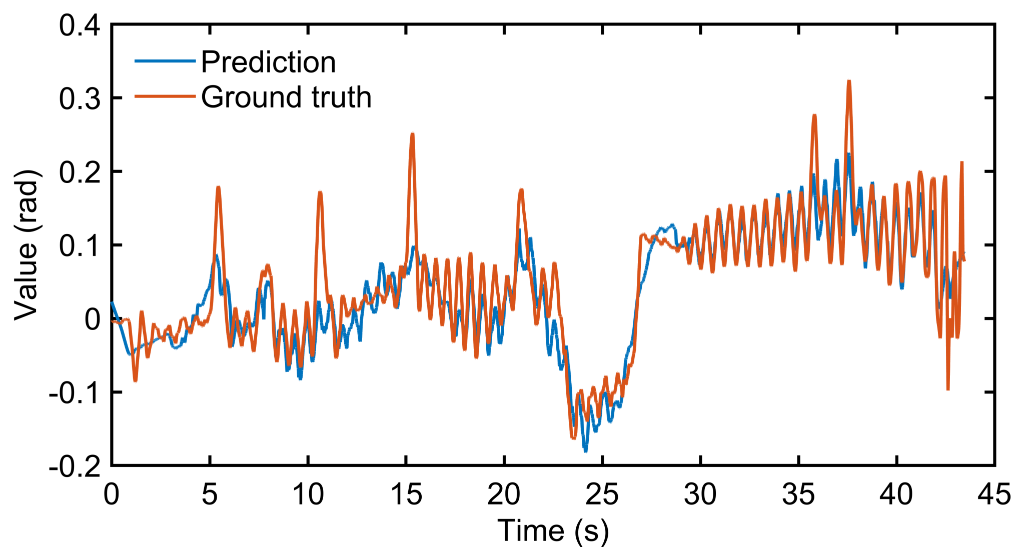

Figure 5.7: Prediction output for Multi-IMU CNN angle Correction Enhancer (MI-CCE)

5.3 Assessment of Methodological Advantages and Challenges

In this section, we delve into the implications of the experimental findings, focusing on the advantages, potential improvements, and limitations of the state estimation methods investigated.

The Enhanced MIPO models, which integrate convolutional neural networks (CNNs) for state estimation, have demonstrated comparable performance to the traditional mocap-based MIPO in terms of error metrics. However, a critical advantage of the CNN-enhanced methods is their operational independence from motion capture systems. This autonomy is particularly beneficial in environments where deploying motion capture technology is unfeasible—such as outdoor or uncontrolled environments—broadening the potential applications of these methods. By eliminating the dependency on external hardware, the Enhanced MIPO models enhance the robustness and versatility of state estimation processes, making them more suitable for diverse real-world applications in autonomous robotic navigation.

While the current CNN implementations provide significant benefits, there is room for improvement in model performance through further optimization. Enhanced results could potentially be achieved by fine-tuning the hyperparameters more extensively or by experimenting with different neural network architectures such as ACNN1D, CNN1D, and CRNN1D. Each of these models offers unique characteristics that could better capture the temporal dynamics and dependencies in IMU data, potentially reducing estimation errors further and increasing the accuracy of state predictions.

The Extended Kalman Filter (EKF), used for integrating sensor data and providing state estimates, also presents opportunities for enhancement. One improvement could be the integration of adaptive filtering techniques, which adjust the filter parameters in real time based on the observed estimation errors. This adaptation helps the EKF to better handle the non-linear dynamics, especially in varied terrains. Additionally, refining the process and measurement noise models could significantly enhance estimation accuracy. This involves recalibrating the noise covariance matrices to better reflect the true sensor behaviors under different operational conditions. Another potential enhancement is the refinement of the mathematical model used in the EKF. This would include more sophisticated modeling of the dynamics involved, particularly in how the robot interacts with its environment, to improve the filter's responsiveness to sensor discrepancies and to mitigate the impact of model uncertainties on the state estimates. These adjustments would enable more robust and accurate performance in real-world navigation tasks, catering to the complexities of autonomous robotic movements.

By addressing these areas, the efficacy of the EKF could be significantly improved, leading to more reliable and accurate state estimations under varying operational conditions. This would further solidify the CNN-EKF framework as a robust solution for state estimation in autonomous systems, paving the way for broader adoption and implementation in more complex navigational tasks.

Due to the limited scope of the project timeline, this study did not include data collection on varied terrains, which is a significant limitation.

Recognizing this, the potential for extending the dataset to include diverse environmental conditions should be a primary focus for future work. Expanding the range of testing scenarios to incorporate different terrains and more complex environmental conditions would not only test the robustness of the state estimation models but also enhance their applicability and reliability for real-world autonomous navigation tasks. This extension would provide a more comprehensive understanding of the models' performance across a broader spectrum of operational contexts, thus contributing valuable insights into their practical utility and limitations.

Appendix

Model Architecture

The detailed architecture of the model is as follows:

Total parameters: 20,418,049

Trainable parameters: 20,418,049

Non-trainable parameters: 0

Input size (MB): 62.59

Forward/backward pass size (MB): 24,925.25

Params size (MB): 77.89

Estimated Total Size (MB): 25,065.73

Table 5.2: Architecture of the ResNet1D Model

Layer (type)	Output Shape	Param #
Conv1d-1	[32, 256, 8691]	75,776
MyConv1dPadSame-2	[32, 256, 8691]	0
BatchNorm1d-3	[32, 256, 8691]	512
ReLU-4	[32, 256, 8691]	0
Conv1d-5	[32, 256, 8691]	327,936
MyConv1dPadSame-6	[32, 256, 8691]	0
BatchNorm1d-7	[32, 256, 8691]	512
ReLU-8	[32, 256, 8691]	0
Dropout-9	[32, 256, 8691]	0
Conv1d-10	[32, 256, 8691]	327,936
MyConv1dPadSame-11	[32, 256, 8691]	0
BasicBlock-12	[32, 256, 8691]	0
...	...	...
Conv1d-106	[32, 1024, 544]	5,243,904
MyConv1dPadSame-107	[32, 1024, 544]	0
BasicBlock-108	[32, 1024, 544]	0
BatchNorm1d-109	[32, 1024, 544]	2,048
ReLU-110	[32, 1024, 544]	0
Linear-111	[32, 1]	1,025

Bibliography

Bouman, Amanda, Muhammad Fadhil Ginting, Nikhilesh Alatur, Matteo Palieri, David D. Fan, Thomas Touma, Torkom Pailevanian, Sung-Kyun Kim, Kyohei Otsu, Joel Burdick, and Ali-akbar Agha-Mohammadi (2020). "Autonomous Spot: Long-Range Autonomous Exploration of Extreme Environments with Legged Locomotion". In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2518–2525. DOI: 10.1109/IROS45743.2020.9341361.

Hutter, M., C. Gehring, A. Lauber, F. Gunther, C. D. Bellicoso, V. Tsounis, P. Fankhauser, R. Diethelm, S. Bachmann, M. Bloesch, H. Kolvenbach, M. Bjelonic, L. Isler, and K. Meyer (2017). "ANYmal - toward legged robots for harsh environments". In: *Advanced Robotics* 31.17, pp. 918–931. DOI: 10.1080/01691864.2017.1378591.

Romanov, Alexey M., Ntmitrii Gyrichidi, and Mikhail P. Romanov (2023). *Enabling Navigation and Mission-Based Control on a Low-Cost Unitree Go1 Air Quadrupedal Robot*. Available at SSRN: https://ssrn.com/abstract=4584863.

Biswal, Priyaranjan and Prases K. Mohanty (2021). "Development of quadruped walking robots: A review". In: *Ain Shams Engineering Journal* 12.2, pp. 2017–2031. ISSN: 2090-4479. DOI: https://doi.org/10.1016/j.asej.2020.11.005. URL: https://www.sciencedirect.com/science/article/pii/S2090447920302501.

Guastella, Dario Calogero and Giovanni Muscato (2021). "Learning-Based Methods of Perception and Navigation for Ground Vehicles in Unstructured Environments: A Review". In: *Sensors* 21.1. ISSN: 1424-8220. URL: https://www.mdpi.com/1424-8220/21/1/73.

Bloesch, Michael (2017). "State estimation for legged robots-kinematics, inertial sensing, and computer vision". PhD thesis. ETH Zurich.

Bahçeci, Beste and Kemalettin Erbatur (19, 2023). "Balance and Posture Control of Legged Robots: A Survey". In: *Journal of Intelligent & Robotic Systems*

108.2, p. 27. ISSN: 1573-0409. DOI: 10.1007/s10846-023-01882-7. URL: https://doi.org/10.1007/s10846-023-01882-7.

Chen, S. Y. (2012). "Kalman Filter for Robot Vision: A Survey". In: *IEEE Transactions on Industrial Electronics* 59.11, pp. 4409–4420. DOI: 10.1109/TIE.2011.2162714.

Bloesch, Michael, Marco Hutter, Markus A. Hoepflinger, Stefan Leutenegger, Stephan Gehring, C. D. Remy, and Roland Siegwart (2013). "State Estimation for Legged Robots: Consistent Fusion of Leg Kinematics and IMU". In: *Robotics: Science and Systems VIII*.

BROSSARD, Martin and Silvère BONNABEL (2019). "Learning Wheel Odometry and IMU Errors for Localization". In: *2019 International Conference on Robotics and Automation (ICRA)*, pp. 291–297. DOI: 10.1109/ICRA.2019.8794237.

Fahmi, Shamel, Geoff Fink, and Claudio Semini (2021). "On State Estimation for Legged Locomotion Over Soft Terrain". In: *IEEE Sensors Letters* PP, pp. 1–1. DOI: 10.1109/LSENS.2021.3049954.

Yang, Shuo, Hans Kumar, Zhaoyuan Gu, Xiangyuan Zhang, Matthew Travers, and Howie Choset (2019). *State Estimation for Legged Robots Using Contact-Centric Leg Odometry*. arXiv: 1911.05176 [eess.SY].

Wisth, David, Marco Camurri, and Maurice Fallon (2022). *VILENS: Visual, Inertial, Lidar, and Leg Odometry for All-Terrain Legged Robots*. arXiv: 2107.07243 [cs.RO].

Barrau, Axel and Silvère Bonnabel (2015). *The invariant extended Kalman filter as a stable observer*. arXiv: 1410.1465 [cs.SY].

Zhang, Zhuqing, Zhang Li, Yong Chen, Dongqi Wang, Jiyu Yu, Rong Xiong, and Yue Wang (2023). "Invariant EKF based State Estimator for Quadruped Robots". In: *2023 42nd Chinese Control Conference (CCC)*, pp. 4311–4317. DOI: 10.23919/CCC58697.2023.10240429.

Camurri, Marco, Maurice Fallon, Stéphane Bazeille, Andreea Radulescu, Victor Barasuol, Darwin G. Caldwell, and Claudio Semini (2017). "Probabilistic Contact Estimation and Impact Detection for State Estimation of Quadruped Robots". In: *IEEE Robotics and Automation Letters* 2.2, pp. 1023–1030. DOI: 10.1109/LRA.2017.2652491.

Wisth, David, Marco Camurri, and Maurice Fallon (2019). "Robust Legged Robot State Estimation Using Factor Graph Optimization". In: *IEEE Robotics and Automation Letters* 4.4, 4507–4514. ISSN: 2377-3774. DOI: 10.1109/lra.2019.2933768. URL: http://dx.doi.org/10.1109/LRA.2019.2933768.

Reinke, Andrzej, Marco Camurri, and Claudio Semini (2019). "A Factor Graph Approach to Multi-camera Extrinsic Calibration on Legged Robots". In: *2019 Third IEEE International Conference on Robotic Computing (IRC)*. IEEE. DOI: 10.1109/irc.2019.00071. URL: http://dx.doi.org/10.1109/IRC.2019.00071.

Yao, Chen and Zhenzhong Jia (2021). *Multi-Sensor State Estimation Fusion on Quadruped Robot Locomotion*. arXiv: 2007.02679 [cs.RO].

Yang, Shuo, Zixin Zhang, Benjamin Bokser, and Zachary Manchester (2023). "Multi-IMU Proprioceptive Odometry for Legged Robots". In: *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. Detroit, USA. ISBN: 978-1-6654-9190-7. DOI: 10.1109/IROS55552.2023.10342061.

Hong, Shenda, Yanbo Xu, Alind Khare, Satria Priambada, Kevin Maher, Alaa Aljiffry, Jimeng Sun, and Alexey Tumanov (2020). "HOLMES: Health OnLine Model Ensemble Serving for Deep Learning Models in Intensive Care Units". In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1614–1624.

Welch, Greg, Gary Bishop, et al. (1995). "An introduction to the Kalman filter". In.

Moore, Thomas and Daniel Stouch (2016). "A Generalized Extended Kalman Filter Implementation for the Robot Operating System". In: *Intelligent Autonomous Systems 13*. Ed. by Emanuele Menegatti, Nathan Michael, Karsten Berns, and Hiroaki Yamaguchi. Cham: Springer International Publishing, pp. 335–348.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun (2015). *Deep Residual Learning for Image Recognition*. arXiv: 1512.03385 [cs.CV].

Andersson, Joel A E, Joris Gillis, Greg Horn, James B Rawlings, and Moritz Diehl (2019). "CasADi – A software framework for nonlinear optimization and optimal control". In: *Mathematical Programming Computation* 11.1, pp. 1–36. DOI: 10.1007/s12532-018-0139-4.