CHAPTER 1
INTRODUCTION

This chapter provides a concise overview of the research in seven sections. It commences by presenting the background, identifying the problem, research objectives, and scope limitations. The chapter then proceeds to cover this thesis's methodology, research method, and contribution. The subsequent chapter will provide a comprehensive explanation.

1.1 Background

In the era of technological developments for autonomous vehicles and robotics, accurate and efficient localization is one of the key challenges. The ability to precisely determine position is essential to support safe autonomous navigation. With the development of sensor technologies such as cameras and LiDAR, combining LiDAR and camera data has become an effective approach to improve localization accuracy, especially by utilizing 3D point cloud data validated with image data to generate accurate depth maps [1]. LiDAR data provides a detailed spatial representation of the environment in the form of a 3D point cloud, while camera data helps provide additional visual information in the form of imagery that enriches the overall understanding of the environment [2].

One dataset that is often used for allocation is KITTI Odometry, which provides high-quality data from cameras and LiDAR to support research in the field of autonomous driving [3]. Here are some methods that use the KITTI Odometry dataset with 3D point cloud as input data:

Table 1. 1 Literature Review Method using LiDAR KITTI Odometry Dataset

LR	Method	Method Description	Accuracy
[4]	CAE-LO (Convolutional Auto-Encoder based LiDAR Odometry)	is to leverage a Convolutional Auto-Encoder (CAE) to automatically detect and describe interest points from LiDAR scans, allowing for more robust and accurate odometry.	0.86%
[5]	SuMa-MOS (Surfel based Mapping - Moving Object Segmentation)	This extension allows the system to segment moving objects from the static environment, which is crucial for accurate mapping and localization in dynamic environments.	0.99%
[6]	SuMa++	SuMa++ uses a network called RangeNet++ for real-time segmentation of LiDAR point clouds, which enables it to generate more accurate and detailed semantic maps.	1.06%
[7]	ULF-ESGVI (Uncertainty-aware Latent	ESGVI optimizes state estimations using Gaussian Variational Inference, while ULF enhances this with uncertainty	1.07%

	Feature-based Exact Sparse Gaussian Variational Inference)	handling, improving robustness to noise and outliers.	
[8]	EfficientLO-Net	It focuses on leveraging the 3D point cloud data generated by LiDAR sensors and incorporates a number of advanced techniques to improve both the speed and accuracy of odometry estimation.	1.92 %
[9]	DeepCLR (Correspondence-Less Architecture)	Is designed for end-to-end point cloud registration without relying on explicit point correspondences.	3.83%
[10]	D3dlo (Deep 3D Lidar Odometry)	LiDAR odometry that processes 3D point clouds to estimate motion without the need to manually define corresponding points.	5.4%

Based on the data in table 1.1, the accuracy value obtained for position accuracy is less than 5.4%. So, another method is needed to improve the accuracy. One recent study proposed a CNN architecture specifically designed to fuse image data and 3D point clouds, which showed a significant improvement in localization accuracy over traditional methods [11].

Although convolutional neural networks (CNNs) have been widely used in image processing for localization purposes, the accuracy achieved is still limited to the complexity of the network architecture and hyperparameter settings [12]. This research focuses on improving localization accuracy by modifying CNN using several additional components, namely Feature Pyramid Networks (FPN), Batch Normalization, and layer freezing techniques. FPN is known to be effective in improving the network's ability to detect objects at various scales [13], while batch normalization helps speed up the training process and improve model stability [14]. Layer freezing aims to lock down the initial part of the model so that the model can focus on the final layers that are more specific to the localization task.

In the context of autonomous robotics, the combination of camera and LiDAR data has been shown to provide better results in 3D mapping compared to the use of single sensor data [3]. However, a big challenge remains on how to optimally combine these two data sources to improve localization and mapping accuracy. This research attempts to address this challenge by applying modifications to CNN to process 3D point clouds resulting from the combination of camera and LiDAR data. Previous research has also shown that by combining batch normalization techniques and architectural modifications such as FPN, the model can capture more details in the surrounding environment which helps improve model performance in complex scenarios [15]. In addition, the implementation of layer freezing techniques allows the model to adapt more quickly to new data without losing generalization of previously learned data [16].

1.2 Problem Identification

Along with the rapid development of autonomous vehicles, the need for systems capable of accurate localization is increasing. Cameras and LiDAR sensors have become the main technologies used in the collection of surrounding environment data. However, combining data from these two sensors in localization tasks still faces various challenges, especially in terms of real-time 3D mapping accuracy [17]. The KITTI Odometry dataset has been widely used in autonomous driving research as it provides visual data from cameras and LiDAR that enables simulation of real environments. However, the use of this data in modeling 3D localization still has limitations, especially in capturing complex environmental details. This problem is even more pronounced when faced with dynamic scenarios, such as object scale differences and position uncertainty [18].

The use of Convolutional Neural Networks (CNN) has shown promising results in image processing for localization tasks. However, the accuracy of CNN models is highly dependent on the complexity of the network architecture and hyperparameter settings. Several studies have shown that while CNNs are quite good at handling camera data, their performance is still suboptimal when used to incorporate data from LiDAR, especially in dynamic 3D environmental scenarios [19]. To overcome these problems, modifications are needed to the CNN architecture that can capture more details from the surrounding environment. This modification can be done by adding Feature Pyramid Networks (FPN) architecture, which has been shown to improve object detection at various scales (Lin et al., 2017). In addition, the application of batch normalization has been shown to improve stability and speed up model training [14]. Layer freezing techniques can also help the model to focus on the final layer, resulting in more accurate predictions in localization tasks [16].

Although various methods have been proposed, the problem of localization accuracy using combined camera and LiDAR data remains a challenge. This research aims to develop a modified CNN model with additional architecture to improve the accuracy of 3D point cloud-based localization on the KITTI Odometry dataset.

1.3 Object

This research aims to improve the accuracy of LiDAR and camera data-based localization using modified Convolutional Neural Networks (CNN). The KITTI Odometry dataset will be used as the experimental basis to test the proposed model. The LiDAR data used will be validated using image data to generate a depth map, which will then be processed by CNN to improve localization performance.

The main focus of this research is to improve the accuracy of the modified CNN by integrating several new architectures. The modifications include the addition of Feature Pyramid Network (FPN) to capture multi-scale features, Batch Normalization (BatchNorm) to improve training stability and speed, and Freeze Layer to reduce computational complexity.

With these modifications, it is expected that the accuracy of 3D point cloud-based localization generated from combined LiDAR and image data can be significantly improved compared to the original CNN. This research will evaluate each component of the modification to determine their respective contributions to the accuracy improvement, as well as explore the potential use of this technique in autonomous vehicle applications and robotics systems.

1.4 Scope Limitation

To keep the experiment from being too long, this thesis limits the works as follows:

1. Data: This research uses the KITTI Odometry dataset which consists of LiDAR data and validated images to generate depth maps. This data is mainly used for autonomous vehicle localization tasks.
2. Research Focus: This research focuses on evaluating the performance of CNN modifications on 3D point cloud-based localization accuracy. The tested CNN architecture includes the addition of FPN to handle multi-scale information, BatchNorm for training stability, and Freeze Layer for computational efficiency.
3. Performance Evaluation: The performance of the modified model will be compared with the standard CNN, with the measured performance parameter being the prediction accuracy in the localization task.
4. Accuracy result: The accuracy of the results achieved in this study is greatly influenced by the limited accuracy of the sensors used, such as LiDAR and cameras. Variability in sensor precision and sensitivity can affect the precision of the data obtained, thus affecting the final results of the model or system developed.

1.5 Methodology

In this thesis, we use fundamental study and experiment based on work-packages (WP). These are the following WP for this thesis:

- **WP 1: Study of literature**
  Collect and study sources and references related to research. A literature review is conducted to comprehend the fundamental concepts related to natural landmark recognition, localization robot, 3D laser range finder, and other relevant techniques.
- **WP 2: Model Selection and Configuration**
  Choosing an appropriate localization method for natural landmark recognition using 3D laser range finder, Model configurations, including hyperparameters and training settings will be defined.

- **WP 3: Implementation of Improvements**
  Developing modifications and enhancements to the localization method to improve its ability to estimating the position of the robot efficiently, finding misidentified landmarks and large angle measurement errors.
- **WP 4: Performance Evaluation**
  Conducting a series of experiments on various natural landmark recognition datasets using suitable localization method. The result will be used to measure the effectiveness of the proposed improvements.
- **WP 5: Result Analysis**
  Analyzing experimental data to evaluate the extent to which the proposed improvements successfully address the issue of estimating the position of the robot efficiently, finding misidentified landmarks and large angle measurement errors.