C2SMART
CONNECTED CITIES WITH SMART TRANSPORTATION
A USDOT University Transportation Center

New York University
Rutgers University
University of Washington
The University of Texas at El Paso
City College of New York

COST-EFFECTIVE APPROACH TOWARDS BUILDING A TRAFFIC SIGN DATA INVENTORY USING OPEN STREET IMAGES

September 2023

TECHNICAL REPORT DOCUMENTATION PAGE

<table>
<tr><td>1. Report No.</td><td>2. Government Accession No.</td><td colspan="2">3. Recipient’s Catalog No.</td></tr>
<tr><td colspan="2" rowspan="2">4. Cost-effective Approach Towards Building a Traffic Sign Data Inventory Using Open Street Images</td><td colspan="2">5. Report Date
September 2023</td></tr>
<tr><td colspan="2">6. Performing Organization Code:</td></tr>
<tr><td colspan="2">7. Author(s) Yinhai Wang, Wei Sun, Chenxi Liu</td><td colspan="2">8. Performing Organization Report No.</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address
Connected Cities for Smart Mobility towards Accessible and Resilient Transportation Center (C2SMART), 6 Metrotech Center, 4th Floor, NYU Tandon School of Engineering, Brooklyn, NY, 11201, United States</td><td colspan="2">10. Work Unit No.</td></tr>
<tr><td colspan="2">11. Contract or Grant No.
69A3551747119</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address
Office of Research, Development, and Technology
Federal Highway Administration
6300 Georgetown Pike
McLean, VA 22101-2296</td><td colspan="2">13. Type of Report and Period
Final report, 3/1/22-9/30/23</td></tr>
<tr><td colspan="2">14. Sponsoring Agency Code</td></tr>
<tr><td colspan="4">15. Supplementary Notes</td></tr>
<tr><td colspan="4">16. Abstract
A computer vision algorithm has been applied to facilitate traffic sign detection and recognition tasks and solve labor-intensive issues. As public traffic signs datasets were collected in different regions, the existing models cannot be directly implemented into the US environment. Therefore, this study collected an additional 5,000 traffic signs in Washington data from Google Map API and self-installed dash cameras. In collaboration with Connected Cities with Smart Transportation (C2Smart) and the Washington State Department of Transportation (WSDOT), STARLab has collected the traffic sign data using three test vehicles equipped with onboard devices. The entire traveling route of three test vehicles covers most of the main roads in the Seattle region. Then, these data were manually labeled into 43 classes for training.
To develop a traffic sign detection and recognition model (TSDR), the Faster R-CNN Inception V2 is selected as a base model for detection with an accuracy rate of 98.34%. Existing datasets and collected data were used to develop a customized traffic sign recognition model, which yielded an accuracy of 97.1%.
In addition, an automated pipeline for traffic signs captures, detects, classifies, and stores is developed. By embedding the TSDR model in edge devices, this system allows the ability to uphold privacy standards. Processed data, consisting of an image and type of traffic sign, is transmitted to the server automatically. A sample traffic signs data inventory in Washington state was created. This database is a valuable resource for developing and testing various machine-learning models. Additionally, it can be used for asset management.</td></tr>
<tr><td colspan="2">17. Key Words</td><td colspan="2">18. Distribution Statement
No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161.
http://www.ntis.gov</td></tr>
<tr><td>19. Security Classif. (of this report)
Unclassified</td><td>20. Security Classif. (of this page) Unclassified</td><td>21. No. of Pages
55</td><td>22. Price</td></tr>
</table>

Form DOT F 1700.7 (8-72) Reproduction of completed page author

Cost-effective Approach Towards Building a Traffic Sign Data Inventory Using Open Street Images

PI: Yinhai Wang
University of Washington
0000-0002-4180-5628

Co-PI: Wei Sun
University of Washington
0000-0002-2191-5352

Student: Chenxi Liu
University of Washington
0000-0002-5447-4768

C2SMART Center is a USDOT Tier 1 University Transportation Center taking on some of today's most pressing urban mobility challenges. Some of the areas C2SMART focuses on include:

Urban Mobility and Connected Citizens

Urban Analytics for Smart Cities

Resilient, Smart, & Secure Infrastructure

Disruptive Technologies and their impacts on transportation systems. Our aim is to develop innovative solutions to accelerate technology transfer from the research phase to the real world.

Unconventional Big Data Applications from field tests and non-traditional sensing technologies for decision-makers to address a wide range of urban mobility problems with the best information available.

Impactful Engagement overcoming institutional barriers to innovation to hear and meet the needs of city and state stakeholders, including government agencies, policy makers, the private sector, non-profit organizations, and entrepreneurs.

Forward-thinking Training and Development dedicated to training the workforce of tomorrow to deal with new mobility problems in ways that are not covered in existing transportation curricula.

Led by New York University's Tandon School of Engineering, C2SMART is a consortium of leading research universities, including Rutgers University, University of Washington, the University of Texas at El Paso, and The City College of NY.

Visit c2smart.engineering.nyu.edu to learn more

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

Executive Summary

Traffic signs inventory is essential for asset management as it provides information about traffic sign, such as the installation date, type, and condition. Previously, many road agencies manually collected traffic signs data, which is a time-consuming process and has low accuracy.

A computer vision algorithm has been applied to facilitate traffic sign detection and recognition tasks and solve labor-intensive issues. As public traffic signs datasets were collected in different regions, the existing models cannot be directly implemented into the US environment. Therefore, this study collected an additional 5,000 traffic signs in Washington data from Google Map API and self-installed dash cameras. In collaboration with Connected Cities with Smart Transportation (C2Smart) and the Washington State Department of Transportation (WSDOT), STARLab has collected the traffic sign data using three test vehicles equipped with onboard devices. The entire traveling route of three test vehicles covers most of the main roads in the Seattle region. Then, these data were manually labeled into 43 classes for training.

To develop a traffic sign detection and recognition model (TSDR), the Faster R-CNN Inception V2 is selected as a base model for detection with an accuracy rate of 98.34%. Existing datasets and collected data were used to develop a customized traffic sign recognition model, which yielded an accuracy of 97.1%.

In addition, an automated pipeline for traffic signs captures, detects, classifies, and stores is developed. By embedding the TSDR model in edge devices, this system allows the ability to uphold privacy standards. Processed data, consisting of an image and type of traffic sign, is transmitted to the server automatically.

A sample traffic signs data inventory in Washington state was created. This database is a valuable resource for developing and testing various machine-learning models. Additionally, it can be used for asset management.

Table of Contents

List of Figures

List of Tables

Section 1 Introduction

Subsection 1.1 Project Background and Problem Statement

As critical assets for roadway and infrastructure management, traffic signs play a pivotal role due to their vast variety and differing conditions. Given their significance, constructing a data inventory for traffic signs is an essential task for comprehensive asset management. However, creating such an inventory poses two significant challenges. The ubiquity of traffic signs makes the collection of initial information, such as raw images, an arduous and time-consuming process. Furthermore, once this raw data is gathered, the task of extracting traffic signs and their associated information from the sources requires an innovative solution.

Understanding the context of these challenges and the importance of resolving them necessitates an examination of the Asset Management Plan proposed by the US Department of Transportation (DOT). This plan is predicated on the inclusion of an asset inventory and an assessment of its condition. The inventory acts as the groundwork for various other critical tasks, such as asset lifecycle planning, performance measurement, risk considerations, and investment strategies. As such, the motivation to address the identified research problem can be summarized in the following key objectives: supporting the asset management plan through active traffic sign condition monitoring, aiding in the performance evaluation, maintenance, risk estimation, lifecycle planning, and investment of other assets, increasing the data diversity of the databases, and augmenting the functionality of the advanced traveler information system.

The significance of this problem is further emphasized by the recent FHWA Transportation Asset Management Plan (TAMP). The framework proposed by FHWA is shown in Figure 1-1, which indicates the significant role of data sensing and collection in the asset management. This plan mandates every state to formulate a risk-based asset management plan for pavement and bridge inventories—critical infrastructures of the National Highway System (NHS). The NHS, integral to the nation's economy, defense, and mobility, does not currently encompass traffic signs in its asset management. For instance, in Washington State, which has a vast network of interstate, non-interstate State Highways, and Local Agency lane-miles, traffic sign information would significantly augment the NHS and enhance asset and highway management in the region. Furthermore, the incorporation of a comprehensive traffic sign inventory could foster a deeper understanding of mobility and safety designs across various roadway segments through data integration.

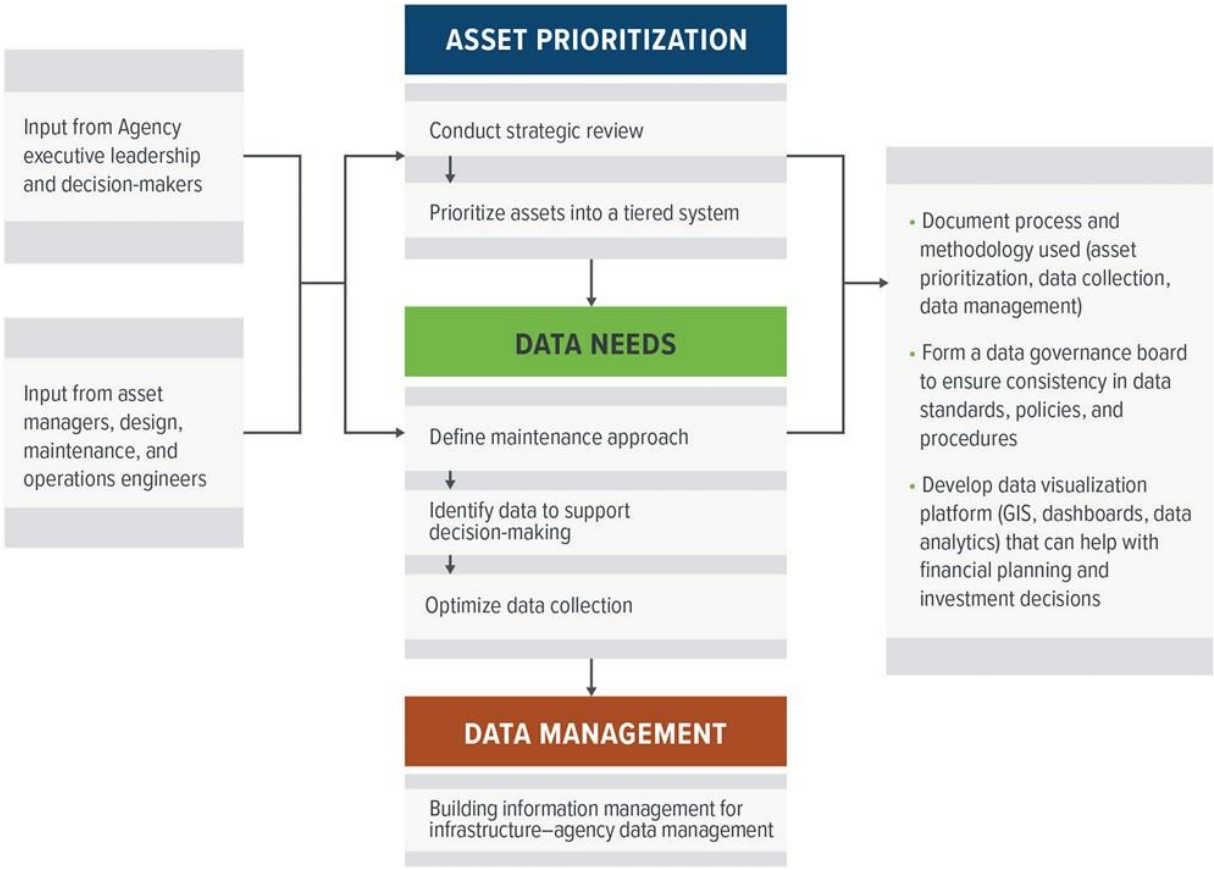

Figure 1-1 FHWA Asset Management Graphic Framework

Subsection 1.2 Project Objectives

The objectives of this research are articulated with the primary focus on the utilization of machine learning (ML) in enhancing traffic sign management. The project embarks on these objectives through three primary components.

Firstly, the research intends to propose a machine learning-based approach to enable cost-effective traffic sign detection. This objective seeks to leverage the power of advanced machine learning algorithms to automate the task of traffic sign detection, minimizing the reliance on labor-intensive, manual methods. Notably, by employing machine learning techniques, the research aims to enhance the accuracy and efficiency of traffic sign detection, making it not only cost-effective but also more reliable.

Secondly, the project aims to create an automated pipeline for the collection and classification of traffic sign data. This task is integral to addressing the previously mentioned challenge of the ubiquity of traffic signs, where conventional methods of data collection prove to be labor-intensive and time-consuming. Through automation, the research strives to expedite the collection process, bringing consistency and precision in data collection. Furthermore, with an automatic classification mechanism in place, traffic signs can be categorized accurately based on predefined classification models. This component will ensure a smooth flow of data from the collection stage to the point of usage, significantly boosting the overall efficiency of the traffic sign management process.

Lastly, the research plans to build a sample data inventory of traffic signs, derived from open imagery along with associated information. This objective seeks to consolidate the data obtained and classified in a structured and accessible format. By using a data inventory, the research will create a comprehensive repository of traffic sign data, which will serve multiple purposes. For one, it can act as a valuable resource for developing and testing various machine learning models. Additionally, this inventory can also support various asset management initiatives, providing key insights for performance evaluation, maintenance planning, risk assessment, and investment strategies related to traffic sign assets.

In extension, the outcomes of these objectives will not only cater to the traffic sign detection and management but also have a broader impact on traffic control, road safety, and infrastructure planning. By enabling efficient traffic sign detection and classification, the research will support informed decision- making, proactive maintenance schedules, and effective asset management, ultimately contributing to safer and better-managed roadways.

Subsection 1.3 Research Scope

The research scope can be described from the following perspectives:

The project consists of seven main tasks:

1. **Review of the State of the Art and Practice (Task 1):** The objective of this initial task is to gather a comprehensive understanding of the current standing and challenges associated with the development of a traffic sign data inventory. To this end, the research team will examine existing public datasets, including traffic signs, and scholarly literature on topics such as traffic sign detection and recognition methods (TSDR), traffic sign data collection, and classification systems. This literature review will facilitate the summation of the strengths and weaknesses of existing methodologies, datasets, and systems.
2. **Data Source Selection and Verification (Task 2):** This task involves identifying the appropriate types of data and sourcing them in alignment with our research objectives. Relying on the literature review completed in Task 1, the team will enumerate all necessary data for the various components of the research - traffic sign detection, recognition, classification, and camera calibration. Subsequently, data sources utilized in the literature will be examined for their size, scale, and quality, leading to the selection of data sources for the project and the development of a data collection plan.

3. **Development of Data Collection System (Task 3):** The aim of this task is to design a data collection system capable of automatically sourcing the necessary data from various platforms. This involves the development of web applications that autonomously download traffic sign images, street images, and relevant information. Moreover, should the need arise, the research team will devise our own data collection devices. The sample data collected by the system in this project will serve as the training and test datasets for subsequent tasks.
4. **Development of a Machine Learning Based Traffic Sign Detection and Recognition Method (Task 4):** This task focuses on creating an automatic traffic sign detection and recognition model based on machine learning. The model will be divided into two parts: traffic sign detection (TSD), which locates the target within a frame, and traffic sign recognition (TSR), which conducts a detailed classification to identify the type of the detected target. Both TSD and TSR models will be trained using the sample data collected in Task 3, and a report will be prepared on their accuracy and efficiency.
5. **System Integration and Sample Data Inventory Creation (Task 5):** This task aims to integrate the various systems and create a sample data inventory. Firstly, a complete data flow will be established by integrating three systems: the data collection system, traffic sign detection and recognition system, and data storage system. Secondly, a data inventory will be designed and built to store the results from the system.
6. **Outreach and Technology Transfer (Task 6):** In this task, the research team will collaborate with transportation agencies, such as USDOT, WSDOT, and Seattle DOT, and industry partners, such as Microsoft, INRIX, Google, and Uber. This collaboration aims to develop an asset management system for traffic signs and obtain additional data support and technological advancements. Through seminars, webinars, and conferences, knowledge, findings, models/methods, prototypes, data inventories, and algorithms from this research will be transferred to transportation agencies, industry partners, researchers, and practitioners.
7. **Final Project Report (Task 7):** A draft final report documenting the literature review, system design, development of the data collection system, traffic sign detection and recognition system development, sample data inventory, and significant research findings will be submitted. After incorporating feedback from C2SMART and USDOT, a final report will be produced and submitted.

The research scope also highlights the need for an additional training dataset from existing traffic sign classification databases for model training, ensuring the robustness of the machine learning model developed. It further stipulates that the sample data inventory will cover a small region in Washington state, serving as a practical demonstration of the research's applicability. Importantly, the data items included in the inventory will depend on the availability of information from the original data sources, ensuring the integrity and accuracy of the data collected and classified.

Subsection 1.4 Solution Introduction

The research project will address the identified problem using the following solutions and capabilities:

- **Collection of Labeled Traffic Sign Images**: The first step towards realizing our research objective is the acquisition of a sizable and diverse set of labeled traffic sign images. These images will be instrumental in the training and calibration of our machine learning model. They will provide the necessary ground truth data to facilitate supervised learning and validation of our algorithms.
- **Development and Training of the Machine Learning Model**: We will subsequently embark on the design and development of a machine learning model, paired with an appropriate neural network architecture. This model will be intensively trained using the labeled traffic sign images collected in the first step. The iterative training process will fine-tune the parameters of the model to optimize its performance in detecting and classifying traffic signs from a variety of images.
- **Development of a Web Application**: Parallel to the model development, we will also be creating a web application capable of automatically retrieving street images, routes, and other related data. This application will provide an effective and streamlined data collection mechanism, feeding our machine learning model with continuous, real-time data.
- **Execution of the Machine Learning Model**: Once the model is trained and the web application is ready, we will run the machine learning model on the collected data. The model will apply its learnt patterns to new data to detect and extract traffic signs and their associated information from the street images. This process transforms raw, unstructured data into structured, actionable information.
- **Design and Creation of a Data Inventory**: Following the extraction of traffic sign data, we will then move towards designing a suitable data inventory structure. This inventory will store and organize the processed data for easy access and analysis. As a part of this process, we will create a sample data inventory for a small region in Washington State as a demonstration of our approach and capabilities.
- **Validation, Analysis, and Documentation**: The final stage of the project involves validation of the machine learning model's performance, analysis of the inventory data, and comprehensive documentation of the entire process. We will verify our model's results, evaluate the effectiveness of our data inventory, and compile all these findings in a detailed project report. This phase will ensure that our solution is robust, effective, and well-documented for future reference and replication.

Section 2 Literature Review

Traffic sign detection and recognition (TSDR) models are an essential component of advanced driver-assistance systems (ADAS) and autonomous vehicles (AV). These models have undergone substantial advancements in the past years, transitioning from conventional image processing techniques to more sophisticated machine learning and deep learning methods. Early models relied primarily on color-based segmentation and shape-based detection to identify traffic signs, leveraging the geometric properties and color information characteristic of traffic signs. As machine learning algorithms evolved, models incorporated techniques such as Support Vector Machines (SVM), Random Forests, and decision trees to enhance recognition accuracy.

In recent years, the advent of deep learning technologies has transformed the field of TSDR. Convolutional Neural Networks (CNNs) have emerged as a popular choice for their ability to automatically learn hierarchical features from raw pixel data. These deep learning models have consistently outperformed traditional techniques, demonstrating their superiority in large multi-class traffic sign datasets like the German Traffic Sign Recognition Benchmark (GTSRB). Furthermore, models such as YOLO (You Only Look Once) and region-based CNN models like Fast R-CNN and Faster R-CNN have shown effectiveness for real-time applications, balancing speed and accuracy. While significant progress has been made, challenges remain due to variable lighting conditions, occlusions, physical wear and tear of signs, and the diversity of traffic signs in shape, color, and symbols.

In the following parts of this section, the research team has conducted a thorough review of existing methods for traffic sign detection and recognition. The aim is to understand their potential applications as well as identify limitations within the context of practical transportation implementations.

Subsection 2.1 Traffic Sign Detection Model

The stage of Traffic Sign Detection is primarily concerned with pinpointing the area within an image or video frame that holds the traffic signs. As traffic signs are typically characterized by distinct colors and shapes, they can be effectively detected through the utilization of color and shape information. This section offers a succinct review of established detection methods, which are generally categorized into color-based, shape-based, and combined color and shape-based strategies.

2.1.1 Color-based Detection

Color-based detection techniques leverage color as a primary feature to identify regions containing traffic signs within images or video frames. A commonly used method in this regard is color segmentation, which helps eliminate irrelevant objects, thereby reducing the search area within the image or video frame.

In a study presented in [1], authors proposed a method to detect red traffic signs in video frames using color as the primary feature. The original video frame is converted to grayscale, with the red channel alone extracted from the original frame. The grayscale frame is then subtracted from the red component of the original frame, and the resulting frame is binarized by applying a threshold. Morphological operations are subsequently performed to eliminate background noise. The system was tested under both daylight and shadow conditions, achieving detection accuracies of 93.3% and 95%, respectively.

Another method, described in [2], involves the use of color distance for traffic sign detection from image sequences. This distance, similar to Euclidean distance between two points, is calculated by determining the difference between two colors. As the color distance decreases, the similarity increases. A color standardization method is proposed in [3] for detecting traffic signs. Here, each pixel in the RGB space is compared with predefined threshold values for each component. Only the component value exceeding the threshold is converted to 255; otherwise, it's set to 0. Following color standardization, the image contains only black, red, green, blue, cyan, magenta, and white. Only the relevant color is retained in the image, and others are discarded to yield a binary image.

While the aforementioned methods use the RGB color space for segmentation (despite its sensitivity to illumination changes), the method presented in [4] performs traffic sign detection in the Hue-Saturation-Intensity (HSI) color space. Owing to the separation of achromatic and chromatic components in this space, detection can be performed even under challenging lighting conditions. The RGB image is converted to HSI color space and segmentation is performed based on Hue and Saturation values. The image is then binarized by setting the pixels of interest to 255 and others to 0. The result of STOP sign detection from [4] is illustrated in Figure 2-1.

Similar techniques are used in [5], but with the Hue-Saturation-Value (HSV) color space for traffic sign detection. The image in the RGB color space is converted to HSV and segmentation is done based on the Hue and Saturation values. Morphological operations such as erosion, dilation, opening, and closing are used to remove small background noise in the segmented image. Thereafter, connected component analysis is carried out on the segmented image to identify potential traffic sign regions satisfying height, width, and area constraints. In [6], the YCbCr color space is chosen for traffic sign detection. The red component is extracted from the YCbCr color space by applying dynamic thresholding on the Cr component value. Objects with an area smaller than a predefined value are eliminated, as they are unlikely to be traffic signs.

Despite higher computational times for color space conversion, segmentation in other color spaces yields better results under various illumination conditions. A comparative study on traffic sign detection in videos using RGB, HSV, and HSI color spaces was presented in [7].

The results obtained under different lighting conditions using these three-color spaces are summarized in Table 2-1. The findings suggest that the HSV color space offers the best detection rate under varying lighting conditions. In order to mitigate the effects of illumination, an automatic white balancing is performed before color segmentation in Figure 2-1. The post-balancing image appears as if it were captured under canonical light.

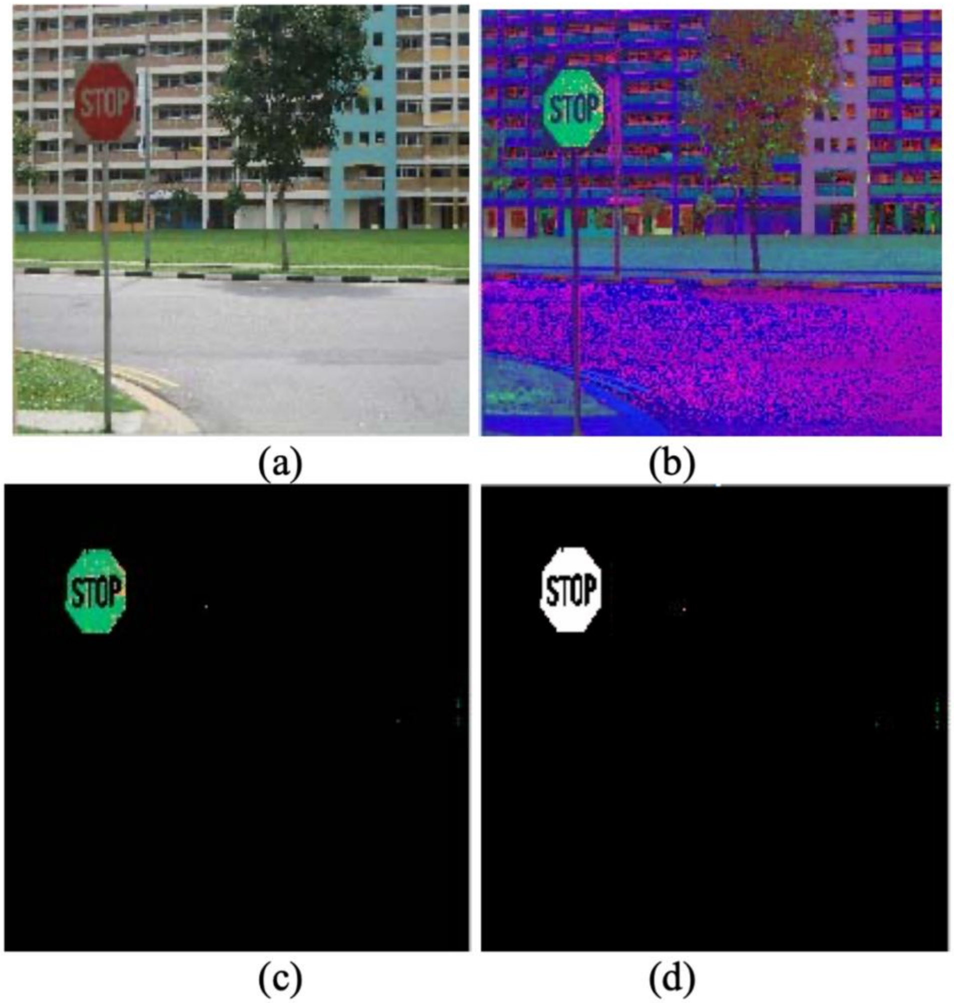

Figure 2-1 Results of Color-based Traffic Sign Segmentation in HSI Color Space [4]

Table 2-1 Detection Rate of RGB, HSV, and HIS color spaces

Color Space	Detection Rate
RGB	88.75%
HSV	95.00%
HIS	91.35%

2.1.2 Shape-based Detection

Shape-based detection methods emerged as a solution to the key challenge faced by color-based detection approaches, namely, fluctuations in ambient illumination. Given that traffic signs usually take on specific geometric shapes such as triangles, rectangles, octagons, and circles, shape-based methodologies are highly applicable. This section will delve into some popular shape analysis algorithms used in these methods.

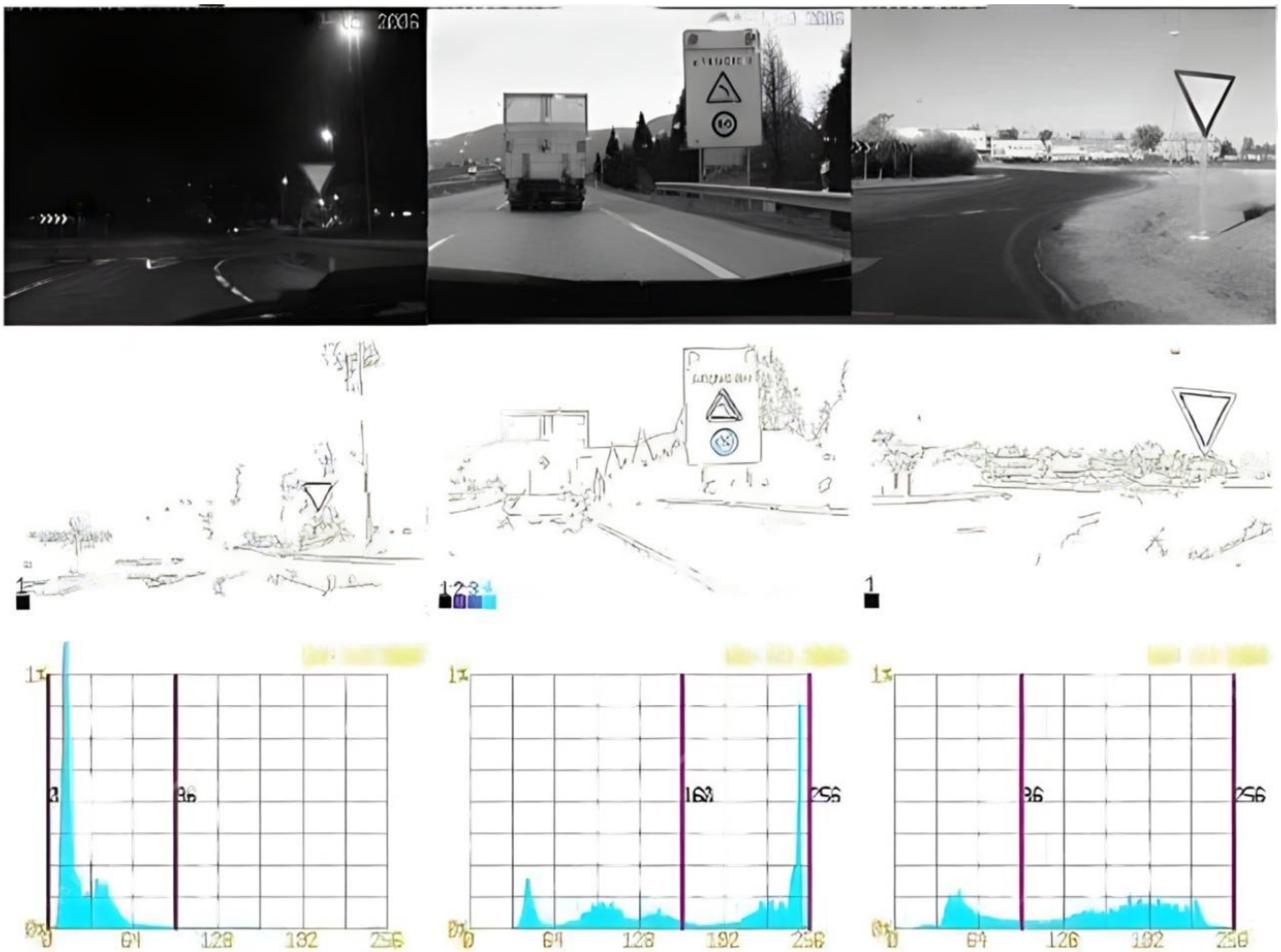

Figure 2-2 Sample Shape-based Traffic Sign Detection based on Contours detection

In the work presented in [9] (See Figure 2-2), Hough transform is utilized for traffic sign detection. The Hough transform is used to detect circular signs based on their circumferences, while for triangular signs, it is used in relation to straight lines. To find the edges in an image, the Canny edge detection method is employed (see Figure 2-2). However, analyzing every contour within an image poses a high computational cost. To address this, certain contours are discarded based on the area and perimeter of these contours, and the Hough transform is then applied to the remaining ones. This approach with the Hough transforms yielded detection rates of 97.2% for speed limit signs and 94.3% for warning signs.

Another noteworthy approach is presented in [10], where speed limit traffic signs are detected using the radial symmetry transform. The source image is first converted into grayscale, followed by the computation of each pixel's gradient. Subsequently, two same-sized vote images are constructed, each encoding information about gradient orientation and magnitude. By combining these two images, a radial symmetry image is formed. Circle detection is then carried out by applying a threshold to this radial symmetry image. It's important to note, however, that the radial symmetry algorithm can only detect regular polygons, rendering it unfit for detecting traffic signs with geometric distortions.

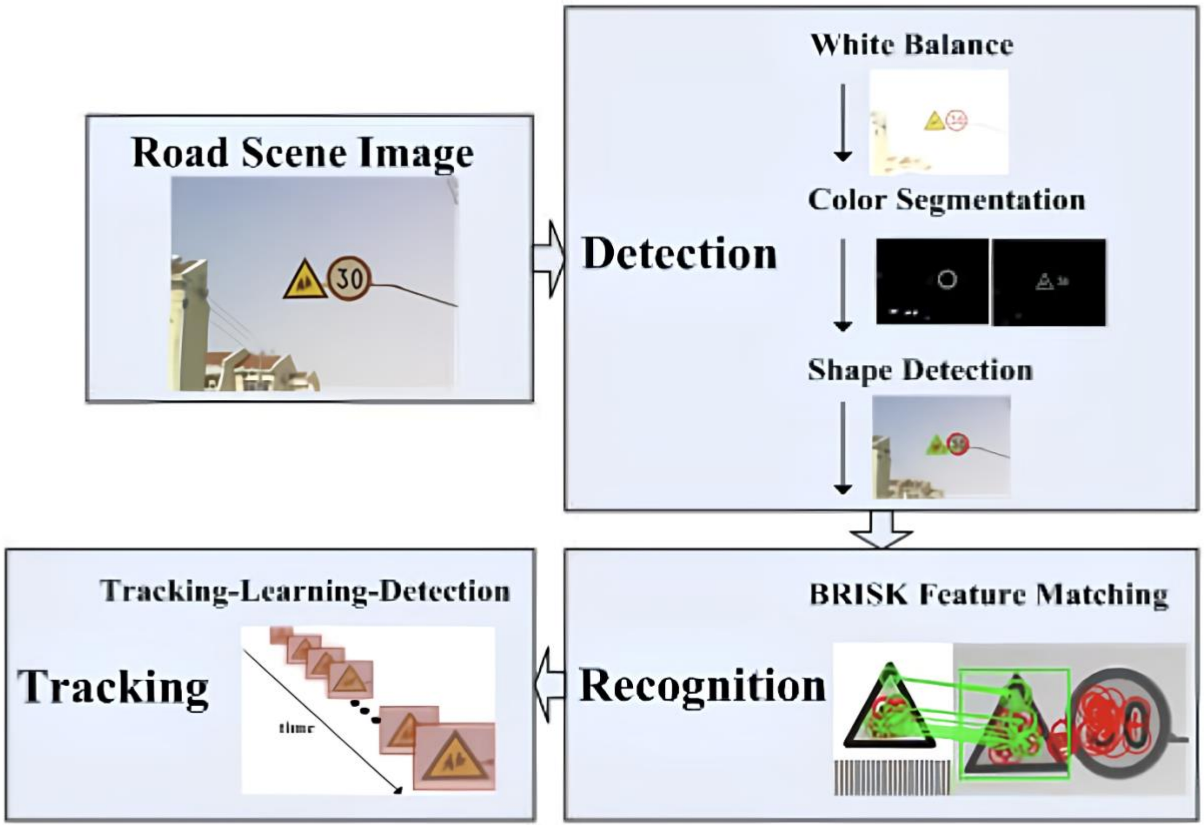

Figure 2-3 A sample framework of Color + Shape based Traffic Sign Detection System [8]

A template-based method was employed in [11] for traffic sign detection. In this approach, a feature image (I) and a feature template (T), both binary images, are used in the detection process. The Distance Transform of the feature image is used to assess the level of match between T and I. Each pixel value in the DT image represents the proximity of that pixel to its closest edge. In order to locate the shape of interest, the template is matched against the DT image. Despite its effectiveness, this approach is associated with a high computational cost.

2.1.3 Color & Shape based Detection

In the realm of traffic sign detection, a combined utilization of chromatic and shape information can mitigate the potential interferences when an image contains objects of similar color and shape. This typically involves two stages: color segmentation in a designated color space, followed by traffic sign detection through shape analysis.

In a study presented in [8], researchers employed both color and shape information to detect traffic signs. The RGB image was segmented using RGB ratios, followed by shape analysis using the Douglas-Peucker (DP) algorithm. The DP algorithm is a contour approximation technique, and the detection process relies on the number of object boundaries. This approach allows for traffic sign detection even in the presence of geometric distortions. The resulting traffic sign detection from [8] is illustrated in Figure 2-3. In a different study, both color and shape were used as features to detect circular traffic signs. The red color segmentation was performed in the HSI color space, and circular signs were detected using the Hough transform for circumference. This approach leverages both chromatic information for initial segmentation and shape analysis for precise detection.

An alternate methodology was proposed in [12], where color segmentation was initially conducted in the HSV color space. Subsequently, boundary boxes were inserted for all regions identified through the color segmentation process. The traffic sign detection was achieved by analyzing features of the bounding boxes such as mean color, size, and the number of pixels enclosed within the boundary box. This method demonstrates how both color and shape information can be used in tandem to successfully detect traffic signs in a variety of conditions.

These methodologies demonstrate the potential of combining color and shape-based detection methods to improve the accuracy of traffic sign detection. By leveraging the strengths of both chromatic and shape information, these combined detection strategies manage to reduce potential interferences and improve overall detection accuracy.

2.1.4 Summary of Traffic Sign Detection Models

In previous sections, the research team reviewed many different traffic sign detection models. Table 2-2 summarizes the features of the methods including the detection type, feature used, color space and detection algorithm.

Table 2-2 Summary of Current Traffic Sign Detection Models

Ref	Year	Detection Type	Feature Used	Color Space	Detection Algorithm
[1]	2016	Color-based	Red Channel	RGB	Thresholding, Morphology
[2]	2010	Color-based	Color Distance	RGB	-
[3]	2016	Color-based	Color Standardization	RGB	Thresholding
[4]	2006	Color-based	Hue, Saturation	HSI	Thresholding
[5]	2016	Color-based	Hue, Saturation	HSV	Morphology, Connected Components
[6]	2010	Color-based	Red Component	YCbCr	Dynamic Thresholding
[7]	2015	Color-based	-	RGB, HSV, HSI	Comparative Study
[8]	2012	Color + Shape	RGB Ratios	RGB	Douglass-Peucker
[9]	2006	Shape-based	Edges	-	Hough Transform, Canny Edge Detection
[10]	2004	Shape-based	Gradient Orientation, Magnitude	-	Radial Symmetry Transform
[11]	1999	Shape-based	Binary Images	-	Distance Transform, Template Matching
[12]	2012	Color + Shape	Mean Color, Size, Number of Pixels	HSV	Bounding Box Analysis

The research team classifies the current methods into three broad categories: color-based methods, shape-based methods, and a combination of color and shape-based methods.

Color-based detection methods primarily leverage the distinct color patterns of traffic signs to identify their presence in a scene. They implement strategies like color segmentation and thresholding in various color spaces (RGB, HSI, HSV, YCbCr) to differentiate traffic signs from the rest of the image. While these methods are intuitive and computationally efficient, they tend to be sensitive to changes in illumination and weather, which can impact color perception. Moreover, false positives can arise from objects in the scene that share the same color as traffic signs.

On the other hand, shape-based detection methods capitalize on the unique geometric characteristics of traffic signs, such as their circular, triangular, or rectangular shapes. Techniques like the Hough Transform, radial symmetry transform, and template matching are often used for this purpose. Although these methods are less sensitive to illumination changes, they can be computationally intensive and might struggle when the sign is partially obscured or damaged.

Recognizing the limitations of relying solely on color or shape information, some research has proposed combining these two aspects. These hybrid approaches often involve a color segmentation stage followed by shape analysis. These methods have shown to reduce false detections and increase overall accuracy, as they can handle a wider variety of scenarios. Nonetheless, they also come with their own set of challenges, including increased computational complexity and the need for effective integration of color and shape data.

In summary, while all three categories of traffic sign detection methods—color-based, shape-based, and combined—have shown promise, they each have their strengths and weaknesses (see Table 2-3). The choice of method often depends on the specific context and constraints of the application, such as computational resources, lighting conditions, and the range of traffic signs to be detected. As research progresses, it is anticipated that more sophisticated models will be developed that further improve the accuracy and robustness of traffic sign detection.

Table 2-3 Summary of Three Kinds of Methods Advantages and Limitations

Methods	Features
Color-based Traffic Sign Detection Methods	**Advantages:** • These methods are simple and intuitive, as color is often a significant and distinguishing feature of traffic signs. • Can be implemented with less computational resources. • Can be highly effective in controlled and favorable lighting conditions.
	Limitations: • Sensitive to illumination and weather changes which can significantly impact color perception. • Different objects in the scene with the same color as traffic signs can lead to false detections. • Conversion to other color spaces may increase the computation time.
Shape-based Traffic Sign Detection Methods	**Advantages:** • Less sensitive to lighting changes as it is based on geometric characteristics. • Can be effective in detecting signs with geometric distortions. • General traffic signs shapes (circle, triangle, etc.) are mostly unique and distinguishable.
	Limitations: • Can be computationally intensive, especially for methods like the Hough Transform.

	• It may not work effectively if the traffic sign is partly obscured or damaged. • Relying solely on shape can result in false detections with objects sharing similar shapes.
Color + Shape based Traffic Sign Detection Methods	**Advantages:** • Combining color and shape information can lead to higher accuracy and fewer false detections. • These methods can handle a wider range of scenarios, as they are not fully reliant on a single type of information. • Can detect signs with both unique colors and shapes, providing a broader range of detection capability.
	Limitations: • The combination of color and shape detection methods can lead to higher computational complexity. • These methods can still be affected by lighting conditions, weather changes, and partial occlusions, although to a lesser extent than methods based solely on color or shape. • Requires a careful balance and integration of color and shape data to ensure accurate results. If the balance is off, it could lead to inaccurate detections.

Subsection 2.2 Traffic Sign Recognition

After the initial detection of traffic signs in images or video frames, the subsequent step is to recognize the type of traffic sign detected. This can be accomplished through either feature matching algorithms or machine learning algorithms. This section briefly describes the different algorithms used for traffic sign recognition.

2.2.1 Feature Matching or Template Matching Algorithms

Feature matching, or template matching, forms one category of traffic sign recognition methodologies. For instance, in [8], the authors employed the scale- and rotation-invariant Binary Robust Invariant Scalable Keypoint (BRISK) features to identify traffic signs. BRISK excels in speed for detection and descriptor computation compared to Speeded Up Robust Features (SURF). BRISK involves two main stages: the detection of keypoints and the formation of binary bit-strings, wherein the first stage identifies points of interest, and the second stage performs pixel-wise brightness comparisons to form the descriptor vector. The template image that provides the highest number of keypoint matches is classified as the target class. In contrast, [12] employs an interest point detector based on the SURF algorithm for traffic sign recognition. SURF serves dual purposes: it detects interest points like blobs and corners in the image, and it forms a feature vector representing these detected interest points. Feature vectors of the target traffic signs are then compared against those stored in a database to identify the most similar match.

In [13], a template matching method is introduced to classify detected traffic signs into specific classes. The method involves comparing detected traffic signs against templates stored in a database, using the normalized cross-correlation (NCC) method to find the match between detected signs and the database- stored templates. A comparison of three feature matching techniques (SIFT, SURF, and BRISK) was undertaken in [14] under two different scenarios: manually segmented road signs compared with database signs, and outputs of the detection system compared against the database-stored signs (see Figure 2-4).

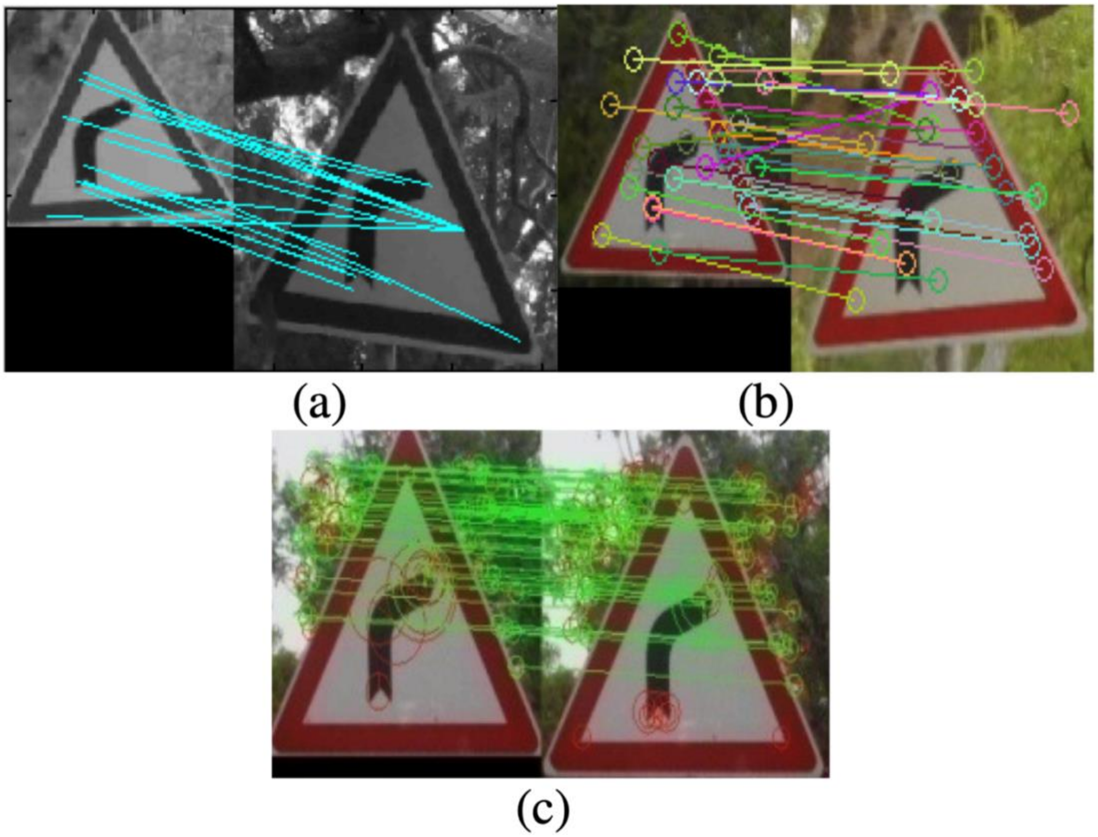

Figure 2-4 Feature Matching Using Three Pre-Defined Descriptors [14]

2.2.2 Machine Learning Approaches

Machine learning-based methods are popular in the realm of traffic sign recognition due to their capability of learning optimal separations between classes. Two dominant methodologies used for traffic sign recognition are artificial neural networks and Support Vector Machines (SVM).

Artificial Neural Networks (ANN), inspired by biological neurons, were utilized in [1] to recognize traffic signs in video frames, achieving a recognition rate of 100% and 94.7% in daylight and shadow environments, respectively. In [4], a multi-layer perceptron neural network trained with the resilient back-propagation algorithm was employed. Meanwhile, a traffic sign recognition system leveraging deep convolutional neural networks was proposed in [15], demonstrating a recognition accuracy of 98.83% at the cost of high computational complexity.

Support Vector Machines (SVM) are a form of supervised learning method. They work by constructing a hyper-plane that decides the class of a given data point. The hyper-plane's maximum margin is defined by the support vectors, which are derived from the data points. In [3], traffic sign recognition is achieved by combining SVM with Zernike moments, which are features invariant to scale, rotation, and translation of images. In [5], the authors employ SVM for traffic sign classification, with SVM being trained using Histogram of Oriented Gradients (HOG) features. Initially used for pedestrian detection, HOG features are invariant to scale variations, making them effective for sign recognition.

To enhance recognition accuracy across different environmental conditions, the authors in [16] utilize a Genetic Algorithm (GA). The GA modifies the original template image using gene coding, represented by 24 bits. The first six bits depict a scaled version of the template, the next six represent the rotated template, and the third and fourth sets of six bits represent the intensity-modified and blurred versions of the template image, respectively. By comparing the detected traffic sign in the image with these modified images, recognition is achieved without the need for a training phase. This system provides an overall accuracy of 94.7%. In [17], traffic signs are detected using Hu moment invariants and neural networks. Hu moments, which are invariant to rotation, scale, and translation, yield high recognition accuracy with relatively low computational complexity. The authors in [18] apply the Local Binary Patterns (LBP) feature extraction technique to classify Chinese traffic signs using SVM. Prior to the recognition stage, the detected regions of interest are resized to 64 x 64 pixels to address scale variations. The LBP approach, robust to grayscale and rotation invariance, is computationally simple. Rotation invariance is achieved via uniform patterns, which also reduces the dimensionality of the feature vector from 256 to 59.

Table 2-4 Summary of Current Traffic Sign Recognition Models

Method	Year	Features	Invariances	Accuracy
BRISK [8]	2012	Binary bit-string	Scale, Rotation	91.25%
SURF [12]	2012	Interest points	Scale, Rotation	90.57%
Template Matching [13]	2016	NCC of template	Scale, Rotation	96.67%
ANN [1]	2016	Neural network features	Depends on Training	94.70%
Deep Convolutional Neural Network [15]	2015	CNN features	Depends on Training	98.83%
SVM with Zernike Moments [3]	2016	Zernike moment features	Scale, Rotation, Translation	100.89%
SVM with HOG Features [5]	2016	HOG Features	Scale	103.47%
GA [16]	2015	Gene coding of template	Scale, Rotation, Intensity	94.70%
Hu Moments and ANN [17]	2010	Hu moment invariants	Rotation, Scale, Translation	Not provided
SVM with LBP Features [18]	2013	LBP Features	Grayscale, Rotation	Not provided

Section 3 Datasets Selection & Verification

The significance of data sources is paramount for the success of the project. High-quality, diverse datasets form the backbone of any machine learning project, and in this context, they are the foundation for building accurate and reliable detection and recognition models. Models learn and draw inferences from the provided data; therefore, the quality and diversity of data directly determine how well the model will perform in real-world scenarios. If the data sources include a broad spectrum of traffic signs, captured under a variety of conditions such as different lighting, distances, angles, and degrees of occlusion, the resulting models will be better equipped to handle complex, real-world situations. Furthermore, with the variation in traffic signs between different regions and countries, data sources should ideally span multiple geographical locations for a more comprehensive representation of global traffic signs. Therefore, the choice and quality of data sources significantly impact the project's success, influencing both the model's training efficiency and its ultimate detection and recognition accuracy.

As a result, the research team conducted an exhaustive review of available public traffic sign datasets, meticulously examining each one based on its diversity, quality, and representativeness of real-world conditions. Three datasets, collected from different geographical regions, were ultimately selected. These datasets offer a rich array of traffic signs that were captured under a variety of lighting conditions, angles, distances, and with varying degrees of occlusion, thus enhancing the potential robustness of the detection and recognition models being trained.

In addition to utilizing publicly available datasets, the team also embarked on a data collection initiative of our own. This involved using Google Street View images, which offer a diverse and comprehensive source of urban and rural road scenes across various countries. Furthermore, to enrich our dataset with first-person perspective images, we used footage from on-vehicle dash cameras, capturing traffic signs in their natural on-road environments. The team then undertook the task of labeling this newly acquired data. By meticulously annotating each traffic sign in the images, the team created a rich, diverse, and highly relevant dataset. This self-collected and self-labeled dataset was amalgamated with the selected public datasets, ultimately forming the cornerstone for training our traffic sign detection and recognition models.

By employing a broad and comprehensive set of data from different sources, the research team has sought to enhance the generalization capability of our models, enabling them to accurately identify and classify traffic signs in diverse and challenging real-world scenarios. This bespoke, multi-source approach to data collection and labeling underscores the team's commitment to achieving the highest possible accuracy and efficiency in traffic sign detection and recognition.

Subsection 3.1 German Traffic Sign Benchmarks (GTSRB)

The German Traffic Sign Recognition Benchmark (GTSRB) dataset 48[19] is a large and diverse dataset specifically designed for the task of traffic sign recognition. It was first released in 2011 for the International Joint Conference on Neural Networks (IJCNN) competition. The research team reviewed the datasets from the following aspects:

Figure 3-1 42 Classes of Traffic Signs Defined in GTSRB Dataset

Figure 3-2 Sample traffic sign data in GTSRB

- **Content:** The dataset contains more than 50,000 images belonging to 43 different classes (see Figure 3-1) of traffic signs (including speed limits, prohibitory signs, dangerous condition signs, etc.). Each class contains varying numbers of examples, reflecting the frequency with which these signs occur in the real world.
- **Quality:** The images in the dataset are all color images, and they have been extracted from larger photographs captured in a variety of lighting conditions and viewpoints. The signs in the images often exhibit variations due to physical damage, lighting effects (like shadows and specular reflections), and occlusions.
- **Labels**: Each image is labeled with the class of the traffic sign it contains. The dataset includes a training set for model development and a test set for model evaluation. This makes it convenient for supervised machine learning approaches.
- **Annotation**: Each image is annotated with a bounding box that indicates the location of the traffic sign within the image. The bounding boxes are tight rectangular boxes that provide a close fit to the traffic sign, facilitating tasks like object detection and region of interest (ROI) extraction (see Figure 3-2).
- **Variety**: The dataset includes a wide variety of traffic signs, captured under different weather conditions (sunny, cloudy, night), different degrees of occlusion, and different distances and angles. This variety helps in developing robust models that can generalize well to new, unseen data.
- **Usage:** The GTSRB dataset has been widely used in many traffic sign recognition researches and competitions. Its large size, quality, and variety of signs make it a benchmark for traffic sign recognition tasks.
- **Limitations:** Despite its strengths, the GTSRB dataset also has certain limitations. The dataset only contains German traffic signs, which may not be universally applicable, especially for countries with significantly different traffic signs. The varying number of examples across different classes may lead to class imbalance problem in machine learning models. The dataset only provides bounding boxes and not pixel-level annotations, which might limit its use in certain applications requiring detailed segmentation. Lastly, it does not include motion blur or weather-related distortions (like fog, snow, or heavy rain), which are common challenges in real-world applications.

Subsection 3.2 Tsinghua-Tencent 100k (TT100K) Dataset

The Tsinghua-Tencent 100K (TT100K) dataset is another noteworthy resource that provides a vast amount of data for training traffic sign recognition models [20]. The dataset is a product of collaboration between Tsinghua University and Tencent, aimed at providing researchers with a high-quality database for street view scene understanding and traffic sign recognition tasks. This dataset represents a variety of geographical locations across China and consists of over 100,000 images, with a resolution of 2048 × 2048 pixels.

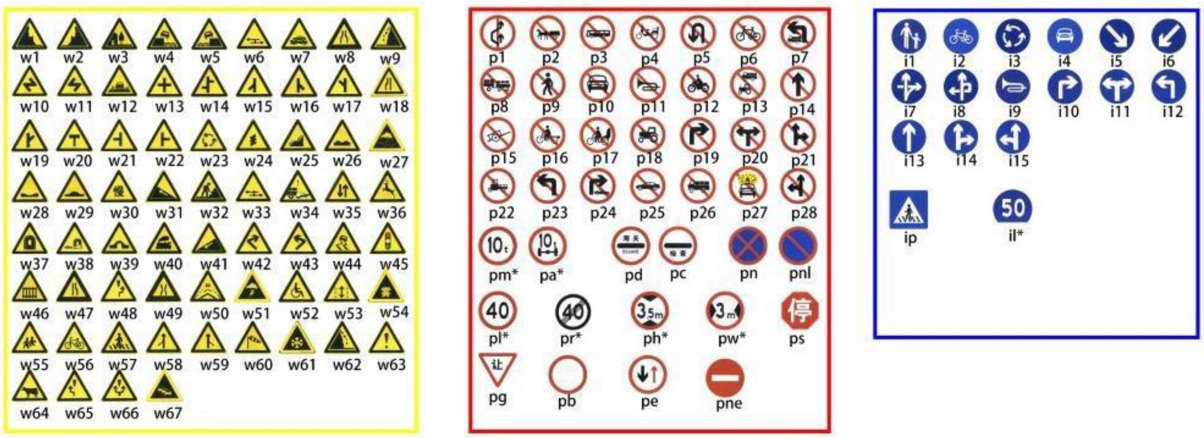

Figure 3-3 Different Classes of Traffic Signs in TT100K Dataset

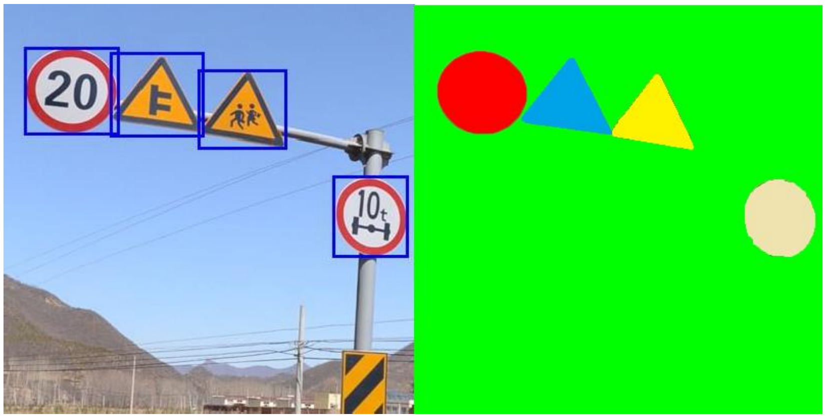

Figure 3-4 Sample Traffic Sign Image in TT100K Dataset

- **Content:** The dataset comprises approximately 100,000 images, making it one of the larger traffic sign datasets currently available. Images were collected from multiple cities across China, providing a broad and diverse collection of traffic scenes and sign appearances.

- **Quality:** The images in the dataset exhibit a wide range of qualities due to variations in lighting conditions, weather, and camera hardware. This range of quality mimics real-world scenarios, providing a suitable platform for developing robust models that can handle various conditions.
- **Annotation:** Each image in the dataset is annotated with bounding boxes for different traffic signs. These annotations encompass over 20 types of traffic signs, providing a varied basis for multi-class traffic sign recognition tasks (shown in Figure 3-3). The annotations are reliable, having been manually verified for accuracy.
- **Variety:** The Tsinghua-Tencent 100K dataset boasts a wide variety of traffic scenes, from highways and city streets to rural roads, encompassing different weather conditions, lighting situations, and occlusion levels. This variety ensures the robustness of models trained on this dataset as it exposes them to a broad spectrum of real-world scenarios.
- **Usage:** Given its rich diversity and comprehensive annotations, this dataset is used widely in various research areas, including traffic sign detection, recognition, and even broader tasks such as scene understanding and autonomous driving system development.
- **Limitations:** Despite its merits, the Tsinghua-Tencent 100K dataset has certain limitations. Firstly, the traffic signs follow Chinese standards, limiting its direct applicability for global or region-specific sign recognition. Furthermore, the dataset primarily includes daytime images, making it less useful for developing models that need to handle nighttime conditions. Lastly, while it does cover various weather conditions, extreme weather scenarios are underrepresented.

In conclusion, the Tsinghua-Tencent 100K dataset, with its diverse content and high-quality annotations, is a valuable asset for traffic sign detection and recognition research. Its limitations provide avenues for supplemental data collection to further enhance model performance under specific conditions.

Subsection 3.3 Laboratory for Intelligent and Safe Automobiles (LISA) Traffic Sign Dataset

The Laboratory for Intelligent and Safe Automobiles (LISA) Traffic Sign Dataset is an influential dataset in the field of autonomous driving and traffic sign recognition [21]. This dataset was collected by the University of California, San Diego (UCSD), and focuses primarily on traffic sign detection, making it a critical resource for developing and validating traffic sign recognition algorithms.

- **Diversity**: The LISA dataset represents the United States driving conditions and includes a total of 47 different traffic sign types, offering good diversity in terms of sign classes. The dataset covers a wide range of scenarios as it was collected in different regions across the United States, under different weather conditions, and at different times of day, thereby providing a varied set of conditions for model training.

- **Quality and Quantity**: The LISA dataset is comprised of approximately 7,000 annotated frames split into six categories, including regulatory signs, warning signs, stop signs, and others. These frames are derived from a total of 20 minutes of 30fps video, offering a good amount of data for model training. However, the quantity might be a limiting factor when compared to other datasets like the GTSRB or TT100K. The image quality varies within the dataset, as it captures real-world driving conditions including various lighting and weather conditions, which can be challenging for model training but also beneficial for building robust models.
- **Annotation**: In the LISA dataset, each image comes with an annotation file in XML format which provides the bounding box location and class label for each traffic sign in the image. The bounding box location includes the coordinates of the upper left and lower right corners, while the class label specifies the type of traffic sign.
- **Challenges**: The LISA dataset presents certain challenges which make it a good testbed for model development. The dataset contains images taken in various lighting and weather conditions, introducing challenges related to changes in illumination and shadows. Some signs are partially occluded, introducing real-world difficulties that models need to handle.

Figure 3-5 Sample Traffic Sign Image in LISA Dataset

In conclusion, the LISA Traffic Sign Dataset, with its real-world challenges and diversity in sign classes and driving conditions, represents a significant resource for developing and evaluating traffic sign recognition algorithms. Despite the quantity being comparatively lower than other datasets, the complexity and realism of the data provide valuable assets for improving the robustness of the trained models.

Subsection 3.4 Google Street View (GSV) Datasets

Google Street View (GSV) offers an invaluable data source for traffic sign detection and recognition due to its vast geographic coverage and diversity. However, using GSV data for this purpose isn't straightforward and requires an in-depth understanding of the key attributes that contribute to its complexity and usefulness. In this project, the research team requested the static API from Google Inc. to obtain the street view images. For example, the following API (https://maps.googleapis.com/maps/api/streetview?size=600x300&location=46.414382,10.013988&heading=151.78&pitch=-0.76&key= AIzaSyCujsaRVLMJaiwfrJ- u5s6tq9Pg YXck8) can obtain the street view image shown in Figure 3-6. In the rest parts of this section, the research team reviewed the dataset from the following aspects:

Figure 3-6 Sample Street View Image Captured by Google Static API

- **Diversity and Geographic Coverage**: Google Street View, with its global footprint, offers images from numerous urban, suburban, and rural environments from different regions and countries. This wide-ranging data offers varied perspectives on traffic signs such as different design conventions, sign placements, weather conditions, and even wear and tear. This variety is key for training robust models capable of operating in diverse conditions. However, it also adds complexity as the model needs to be adaptable to this broad spectrum of input data.
- **Quality and Quantity**: GSV images are typically of high resolution, which is crucial for detailed tasks like traffic sign recognition. Additionally, the sheer volume of available data is unmatched. However, traffic signs could be at different distances, orientations, and states in the images,

introducing additional challenges. The model must be robust enough to recognize signs in any state, whether they're partially occluded, faded, or at a difficult angle.

- **Annotation Challenges**: Google Street View images aren't annotated, which means that bounding boxes and labels must be added manually or by using an automated technique. Automated annotation introduces a bootstrap problem – the users may need an accurate model to automate the annotation process. Manual annotation, on the other hand, is laborious and time-consuming but can yield high-quality labeled data.
- **Legal and Ethical Considerations**: Even though GSV is publicly available, usage restrictions for data mining purposes can vary by region. Further, privacy concerns must be considered, and all data utilization must comply with Google's terms of service. Therefore, it's essential to research and abide by all relevant guidelines before using GSV data.
- **Specific Challenges**: Traffic sign designs can significantly vary across countries or regions, introducing a level of complexity in terms of recognizing different sign designs, colors, and meanings. Lighting conditions, occlusions, and the varying angles of the street view images also add layers of difficulty to the task.

Despite these challenges, Google Street View represents a tremendous resource for traffic sign detection and recognition due to the sheer quantity and diverse range of data available. Overcoming the challenges associated with using GSV data can pave the way for the development of highly versatile and robust detection and recognition models.

Subsection 3.5 Data Collection System Development

To enrich the sources of traffic sign data, the research team designed and implemented a robust data collection system based on dash cameras. This system is displayed in Figure 3-7, featuring a structure designed for efficient data acquisition and processing. It comprises three main components: the Camera Sensor, Edge device, and Back-server, each contributing to a crucial phase of the data collection process.

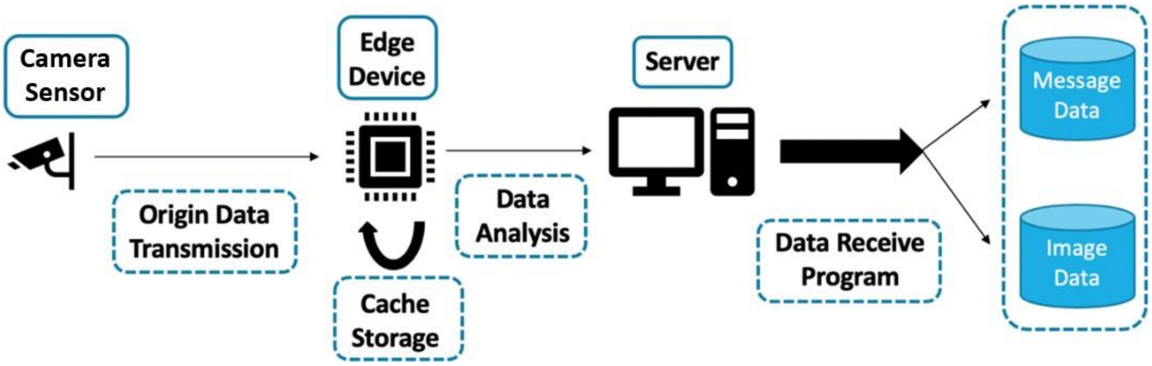

Figure 3-7 The Structure of Data Collection System

Using edge devices for computation brings considerable benefits to transportation systems. Privacy is a critical concern as there is a lot of data transmission in ITS sensing. However, edge computing is a great solution to privacy challenges as data are collected and processed at the edge, and raw data, with private information, is not transmitted to the cloud [22]. In addition, edge computation eliminates high-bandwidth data transmission and heavy post-processing as only objects' representations are sent to the server [23]. As a result, edge computing is successfully integrated as a part of cooperative and real-time ITS systems. For example, an edge device serves as a sensing as a service (SaaS) in the Cooperative and Comprehensive Smart Edge Node for Sensing and Operation (COCO SENSOR) system for practical deployments [24]. In addition, the Edge-based Multi-task Safety-oriented Environmental (Edge-MuSE) sensing system has demonstrated its reliable and precise performance in perception tasks with 92.15% accuracy in visibility estimation and 92.25% in road surface condition classification [25]. For future development, edge devices can be functioned in a multi-thread parallel edge system to increase efficiency and intelligence on the edge [26].

Initially, the dash cameras (shown in Figure 3-8), strategically placed on the test vehicles, capture raw video and image data while navigating various roads. The sample raw data collected from the system is shown in Figure 3-9. These raw data are swiftly transmitted to the Edge device via a high-bandwidth network designed for low latency. Embedded algorithms in the Edge device then spring into action, processing this data to identify and isolate the traffic signs present in the images.

Figure 3-8 The Dash Camera used in the Data Collection System

Figure 3-9 Sample Images Captured by the Data Collection System

Subsequently, a traffic sign recognition model performs a detailed analysis of the isolated traffic signs, classifying them into one of the 43 pre-defined categories, while also assigning a confidence score for each classification. The outputs of this process, which include the detection and classification results as well as the isolated traffic sign image regions, are then transferred to a back-server located in the STAR Lab. Here, they serve as valuable resources for further model training, validation, or testing.

The data collected through this process is bifurcated into two types of databases: one containing message data and the other image data. Given that the recording rate is set at 30 frames per second (FPS), the same traffic sign could be captured hundreds of times during a single drive. To prevent a surplus of duplicated traffic signs in the datasets, the research team conducted manual filtering to expunge identical traffic signs, thereby ensuring the usability of the datasets.

However, the process is not without its challenges. Even though the system is capable of detecting, isolating, and classifying traffic signs, the accuracy of these operations leaves room for improvement. Specifically, when we started the project, our traffic sign recognition (TSR) model, which was pre-trained on the GTSRB and TT100K datasets, struggled to adapt to the unique traffic sign landscape of the U.S., yielding an accuracy of only around 50%. This was mainly due to the significant differences in traffic sign design across Germany, China, and the U.S. Consequently, the TSR model deployed on the Edge device was unable to classify traffic signs with high precision, creating a bottleneck for the team. To mitigate this issue, the research team regularly reviews the collected data in the back server, rectifying errors and continuously improving the quality of the data gathered. This iterative process is an essential part of ensuring the eventual success of the project.

Subsection 3.6 Summary of Datasets Selection & Verification

As part of the research project focused on traffic sign detection and recognition, our team has employed a variety of datasets. Each one offers unique characteristics and strengths, contributing to the robustness of our model training process.

The German Traffic Sign Recognition Benchmark (GTSRB) is a well-known dataset that contains more than 50,000 images of 43 distinct German traffic signs. These images were collected from different real- world settings, and they have been annotated by experts, providing a reliable source for training. The dataset's diversity, in terms of sign types and varied environmental conditions, is its key strength. However, its main limitation lies in its regional specificity, as it only includes German traffic signs.

To complement the GTSRB dataset and introduce international variability, we used the Tsinghua-Tencent 100K (TT100K) dataset, which features traffic signs from China. This dataset is noted for its large volume of traffic sign images, which total over 100,000. The TT100K dataset provides variability in terms of traffic sign design, cultural context, and regional environmental conditions. However, similar to GTSRB, it does not include traffic signs from all regions and its complexity can introduce additional challenges to the model training process.

The Laboratory for Intelligent & Safe Automobiles (LISA) Traffic Sign Dataset brings a necessary North American perspective to our data sources. It contains a variety of American traffic signs captured under diverse conditions, helping our model to better understand and recognize traffic signs commonly seen in the United States. Nevertheless, the dataset size is comparatively smaller than GTSRB and TT100K.

The Google Street View (GSV) dataset serves as an extensive source of global traffic sign data. Google's comprehensive coverage of worldwide roads allows us to extract traffic sign data from a multitude of regions and countries. This vast geographical coverage helps to diversify our dataset further. However, the use of GSV comes with its challenges, such as the high computational resources needed to extract traffic sign data, and the cost involved in accessing the API for data extraction.

Lastly, we supplemented these open-source datasets with our custom data collection system. Equipped with dash cameras installed on test vehicles, this system records and processes traffic sign data in real-time as the vehicle navigates different roads. Despite the challenges with the initial accuracy of the traffic sign recognition model, this approach provides valuable, custom data which greatly enhances the versatility of our model, making it more adaptable to different traffic sign designs and road conditions.

In conclusion, by combining these diverse datasets, including GTSRB, TT100K, LISA, GSV, and our own custom-collected data, we aim to build a comprehensive and robust traffic sign detection and recognition model. This amalgamation of data sources ensures that the model is exposed to a wide variety of traffic signs from different regions, ultimately improving its adaptability and performance.

Section 4 Traffic Sign Detection (TSD) Model Development

This section presents a detailed overview of the traffic sign detection model development undertaken in our project. Traffic sign detection, a crucial step in our workflow, involves identifying and extracting regions of interest (ROIs) from raw video imagery. The aim is to ascertain whether the extracted ROI matches our target object - in this case, traffic signs.

Informed by the comprehensive literature review presented in Section 2, our research team elected to use deep learning-based methods for traffic sign detection. We chose this approach because it effectively encapsulates both color and shape characteristics of traffic signs, resulting in a higher potential for accurate detection.

Nonetheless, we faced a significant challenge. The dearth of publicly available traffic sign data specific to the US limited the applicability of existing, well-developed traffic sign models to our unique use-case scenarios. This issue necessitated the creation of a novel traffic sign detection model, tailored to our specific requirements.

Our team drew from well-established baseline models to build our detection model. Then, through a process of fine-tuning, we customized the model with the help of Google Street View (GSV) data and our self-collected data. This phase of the model development was labor-intensive and required a concerted effort from our lab team members. It was during this stage that we were able to transform the general- purpose models into a traffic-sign-detection-specific tool capable of meeting our project needs.

Upon completion, the performance of our model was thoroughly evaluated using our self-collected dataset. The results were encouraging: the model demonstrated a high detection accuracy rate of 98.34%. This level of performance serves as a testament to the effectiveness of our approach, combining rigorous model development techniques with an intensive fine-tuning process using uniquely gathered datasets. The achievement also underscores our commitment to continually push the boundaries of what's possible in the field of traffic sign detection and recognition.

Subsection 4.1 Detection Model Baseline

The primary objective of a Traffic Sign Detection (TSD) model is to isolate traffic signs from the background within an image. In Task 2, we introduced three publicly available datasets: German Traffic Sign Benchmarks (GTSRB), Tsinghua-Tencent 100k dataset, and LISA Traffic Sign dataset. While these datasets cannot be directly implemented into our project due to regional variances in traffic signage, they can serve as valuable training sets for the TSD model, which can then be utilized for custom data collection from Google Map API and data gathered from self-installed dash cameras.

To create the most effective and efficient TSD model for our needs, we explored and evaluated several TSD model baselines, including Faster R-CNN Resnet 50, Faster R-CNN Resnet 101, Faster R-CNN Inception V2, Faster R-CNN Inception Resnet V2, R-FCN Resnet 101, SSD MobileNet V2, SSD Inception V2, and YOLO V4. The brief overview of the models is shown blow:

Faster R-CNN Resnet 50 and 101: Faster R-CNN, or Fast Region-based Convolutional Network, is a popular model for object detection tasks. The Resnet 50 and 101 versions of this model demonstrated high mean average precision (mAP), indicating strong accuracy. However, they also showed high memory consumption and computational complexity, which might limit their applicability in real-time or resource-constrained settings.

Faster R-CNN Inception V2 and Inception Resnet V2: These variants of Faster R-CNN utilize the Inception architecture, which uses multiple filter sizes in parallel to extract different features at each layer of the convolutional network. While they provided high accuracy similar to their Resnet counterparts, the Inception models were considerably more computationally efficient, especially the Inception V2 version.

R-FCN Resnet 101: Region-based Fully Convolutional Networks (R-FCNs) are known for their efficiency and scalability, outperforming many traditional region-based detectors. The R-FCN Resnet 101 model offered impressive accuracy but consumed less memory compared to Faster R-CNN models.

SSD MobileNet V2 and Inception V2: Single Shot MultiBox Detectors (SSD) can predict the bounding box and class label simultaneously in a single forward pass, making them very efficient. However, the MobileNet V2 and Inception V2 versions of SSD showed lower accuracy compared to other models. Nonetheless, their small model size and fast processing times make them suitable for mobile or edge computing applications.

YOLO V4: You Only Look Once (YOLO) is known for its exceptional speed, making it ideal for real- time detection tasks. While its accuracy was lower than the Faster R-CNN and R-FCN models, YOLO V4 still achieved respectable performance. Notably, it managed to maintain a stable frame rate of 30 FPS on a Jetson Nano device, demonstrating its potential for real-time, on-device traffic sign detection.

Figure 4-1 Sample Detection Results by Faster R-CNN Inception V2 based TSD Model Pretrained by GTSBR Dataset and Finetuned by Self-collected Data.

The performance of these models on traffic sign detection is summarized in Table 4-1.

Table 4-1: TSD Model Baseline Comparison

model	mAP	parameters	memory	total_exec_millis
Faster R-CNN Resnet 50	91.52	43337242	5256.454615	104.0363553
Faster R-CNN Resnet 101	95.08	62381593	6134.705805	123.2729175
Faster R-CNN Inception V2	90.62	12891249	2175.206857	58.53338971
Faster R-CNN Inception Resnet V2	95.77	59412281	18250.446008	442.2206796
R-FCN Resnet 101	95.15	64594585	3509.75153	85.45207971
SSD Mobilenet	61.64	5572809	94.696119	15.14525
SSD Inception V2	66.10	13474849	284.512918	23.74428378
YOLO V4	78.83	50588958	1318.108256	21.4810122

As depicted in Table 4-1, Faster R-CNN Inception V2 delivered the highest performance in terms of accuracy (measured by mean average precision, mAP). However, when considering processing efficiency, YOLOV4 outperformed SSD models in edge device applications, maintaining a stable frame rate of 30 FPS on the Jetson Nano device.

In the data collection process, we selected Faster R-CNN Inception V2 as our base model and fine-tuned it with the GTSBR dataset for enhanced traffic sign data collection. An example of this data collection can be seen in Figure 4-1. This meticulous and data-driven approach to model selection and fine-tuning ensures that our model is highly attuned to our specific data requirements, leading to optimal performance in traffic sign detection.

Subsection 4.2 Model Structure Design

Based on the tests on various baseline models, the research team selected the Faster R-CNN Inception V2 as the backbone network. Faster R-CNN Inception V2 is a widely utilized deep learning architecture for object detection tasks, including traffic sign detection. It merges the merits of the Fast R-CNN architecture with those of the Inception V2 network, offering impressive detection accuracy with efficient computation. The structure of the proposed traffic sign detection model is shown in Figure 4-2.

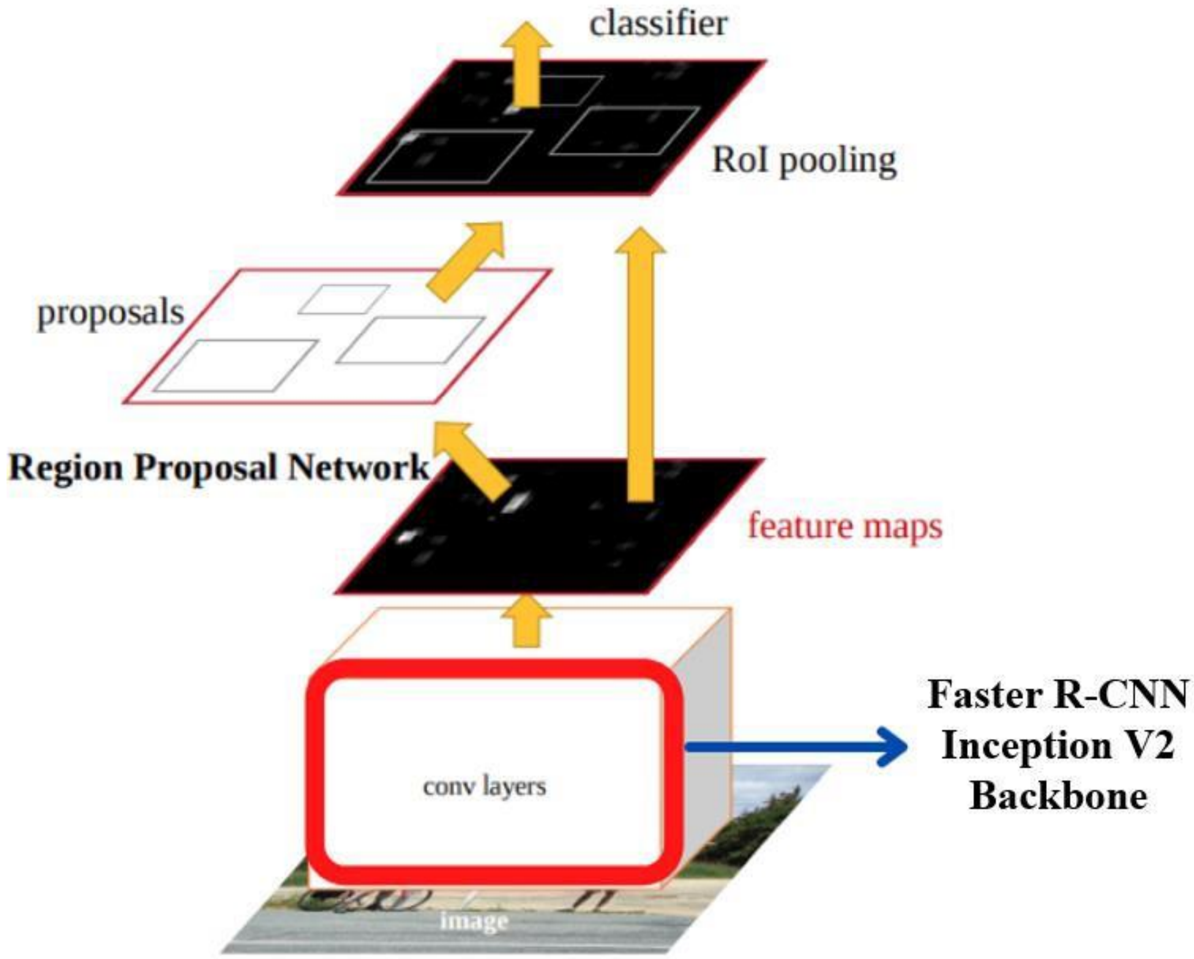

Figure 4-2. Structure of Proposed Traffic Sign Detection (TSD) Model Structure

Inception V2 as the Backbone Network: The backbone network is responsible for extracting useful features from the input image. For Faster R-CNN Inception V2, the Inception V2 architecture is used. Inception V2, a modification of the original Inception (or GoogLeNet), uses batch normalization and separates convolutions for more efficient computation. It consists of multiple inception modules that capture complex spatial and channel-wise patterns in the image data.

Region Proposal Network (RPN): The RPN is a fully convolutional network that scans the image features produced by the backbone and proposes potential object bounding boxes (regions of interest, or ROIs). It generates anchor boxes of different sizes and aspect ratios and classifies them as "object" or "background."

Region of Interest (RoI) Pooling: RoI Pooling is used to convert the proposed regions, which could be of different sizes, into a fixed size so that they can be processed by a fully connected layer. It performs max pooling on inputs of non-uniform sizes to obtain fixed-size feature maps (e.g., 7x7).

Classification and Bounding Box Regression: The feature maps from the RoI pooling stage are passed through fully connected layers to perform two tasks: classify the proposed region into specific classes (in this case, different traffic sign types), and refine the object bounding box coordinates to fit the actual object shape more precisely.

Non-maximum Suppression (NMS): In the final step, NMS is used to eliminate redundant overlapping bounding boxes. For each class, it keeps only the most probable bounding box while suppressing others with significant overlap.

Subsection 4.3 Model Training and Validation

The dataset that being used for training Traffic Sign Detection and Recognition (TSDR) model included three reputable public traffic sign databases, namely the German Traffic Sign Recognition Benchmark (GTSRB), Tsinghua-Tencent 100K (TT100K), and the Laboratory for Intelligent & Safe Automobiles (LISA). Each of these datasets presented its unique merits within the scope of the research. However, they also posed certain limitations when utilized for model training and evaluation. For instance, the GTSRB database solely comprises German traffic signs, rendering it incompatible with the U.S. traffic environment. While the TT100K dataset is valued for its extensive collection of traffic signs, it does not encompass traffic signs from all geographic regions. Finally, the LISA dataset, despite its utility, is relatively small in size and therefore may lack the capacity to cover a comprehensive range of real-world scenarios for effective model testing.

Consequently, there was a discernible need for additional data collection to adequately test the model across an array of traffic signs sourced from diverse regions. This expanded approach ensures that the developed model can adapt more efficiently to a broader spectrum of road conditions and traffic sign designs. It also fosters a more comprehensive model training environment, resulting in more accurate and inclusive traffic sign detection and recognition. Therefore, the team sought to enhance their original data collection with an even more varied and representative compilation of traffic sign data, supporting the continual refinement and expansion of the TSDR model capabilities.

Data Collection System Development

To overcome the limitations of the three public datasets and augment our original data sources, our research team devised an inclusive data collection system. This system aimed to gather an array of traffic sign data across a broad range of contexts, thereby bolstering the Traffic Sign Detection and Recognition (TSDR) model's robustness and accuracy.

Our data collection process harnessed the power of two unique resources. First, we utilized Google Street View (GSV), a vast repository of geotagged imagery from around the world, to collect diverse traffic sign data. Additionally, we employed self-equipped data collection vehicles, fitted with advanced devices like dash cameras and edge-computing devices. These vehicles traversed diverse urban and rural landscapes, capturing high-resolution images of traffic signs under varied conditions.

These two resources, working in tandem, have significantly broadened the scope and variety of our traffic sign data. We are using this enriched data to train the TSDR model, ultimately enhancing its capability to accurately identify and classify traffic signs from a wide array of regions and conditions.

Our proactive approach to data collection, moving beyond the confines of the initial datasets, has yielded a wealth of diverse data. This will undoubtedly strengthen the TSDR model's ability to generalize effectively across different scenarios, thereby improving its reliability and applicability on a wider scale. By integrating these data sources, we are confident that our TSDR model will rise to meet the challenges of real-world traffic sign detection and recognition.

Google Street View (GSV) Dataset

Leveraging Google Street View (GSV) as the primary resource, our data collection process utilized the Google Maps Street View Static API, a broad and virtually limitless repository of geotagged images spanning the globe. GSV's utility lies in its provision of static images, depicting panoramic street views across diverse urban and rural landscapes in various countries. The acquisition of these images involved requesting them from the Google Maps API using specific geographical coordinates. To create a diverse dataset suitable for testing our model, we randomly generated locations within a pre-defined area. We then used the coordinates of these locations to generate URL requests, leading to the automatic saving of corresponding images. This rigorous data collection method resulted in the accumulation of over 10,000 images. Such an extensive collection of data, covering a wide range of traffic signs across numerous regions, significantly bolsters the training dataset for the Traffic Sign Detection and Recognition model. It enhances the model's ability to identify and classify traffic signs accurately and comprehensively. Furthermore, GSV provides images that offer diverse perspectives on traffic signs, reflecting variations in sign placement, weather conditions, distances, orientations, and states. These aspects mimic real-world scenarios, thereby making the data collection process more robust and versatile.

However, it is important to note the challenges and limitations associated with using the GSV dataset. For instance, the images from this dataset are not pre-annotated, necessitating either manual annotation or the application of automated techniques to add bounding boxes and labels. In this project, we manually labeled the GSV dataset into 30 classes, a labor-intensive process, but one that ensured high-quality labeled data. Another consideration is adherence to Google's usage restrictions and terms of service when using GSV data. Lastly, obtaining data from GSV requires considerable computational resources, and the costs associated with requesting data from Google's server via the API can be substantial. These factors must be taken into account when planning data collection strategies using this resource.

Self-equipped Data Collection Vehicles

Secondly, the team utilized a hands-on approach with self-equipped data collection vehicles. This active method offered an opportunity to directly engage with the environment and capture high-resolution images of traffic signs across diverse weather conditions, varied lighting, and numerous traffic scenarios. The use of these vehicles was integral in adding an additional layer of complexity to the data employed for training the model. The vehicles were outfitted with advanced devices, including dash cameras, edge-computing devices, and communication modules. This specialized equipment was strategically deployed to navigate both urban and rural routes, capturing raw video and image data of traffic signs along the way.

The data stream of the proposed self-data collection system. The dash cameras functioned as the initial point of contact in this data collection process, capturing and relaying the raw data. Subsequently, these raw data were transferred via a high-bandwidth local network (i.e., LoRa) to the on-board edge devices. Embedded within these devices were proprietary traffic sign identification and isolation algorithms, designed to process and isolate the relevant traffic sign data from the raw input. This systematic process ensured that the valuable information was extracted effectively for further processing. The next phase involves the classification of the isolated traffic signs. Using a specially designed traffic sign recognition model, the signs are classified into one of 43 predefined categories. In the system's early stages, the classification results required human intervention for corrections. However, after several iterations of model training and fine-tuning, the output classification quality progressively improved. This classification process generates a confidence score that represents the likelihood of a correct classification, adding a quantitative measure to the process. The final output comprises the classification results (presented as text) and the isolated regions of the traffic sign images. The final results from this comprehensive process consisted of the classification results presented as text and the isolated regions of the traffic sign images. These results were subsequently transferred to a STAR Lab back-server for further analysis and data management. This approach allowed the research team to gather real-world, practical images in various conditions, which will likely enhance the robustness and accuracy of the Traffic Sign Detection and Recognition model.

Subsection 4.4 Performance Evaluation

Previous studies have developed traffic sign detection (TSD) models. However, further development is needed to extend the capability of the existing models to be compatible with the US traffic signs. Several TSD model baselines, such as Faster R-CNN Resnet 50, Faster R-CNN Resnet 101, Faster R-CNN Inception V2, Faster R-CNN Inception Resnet V2, R-FCN Resnet 101, SSD MobileNet V2, and YOLO V4, were tested with data from Google maps API and self-collected dash camera to improve existing models.

Table 4-2 and Figure 4-3 presents a comparison between several Traffic Sign Detection (TSD) models, based on Mean Average Precision (mAP), number of parameters, memory usage, and total execution time. Among the models evaluated, Faster R-CNN Inception ResNet V2 achieved the highest mAP of 95.77, indicating the best overall precision-recall trade-off. However, it comes with a significantly larger memory footprint (18250.446008 MB) and a longer total execution time (442.2206796 milliseconds), suggesting that it may not be the most efficient model for real-time applications. On the contrary, SSD Mobilenet and SSD Inception V2, although scoring relatively lower in mAP (61.64 and 66.10 respectively), have significantly fewer parameters and use considerably less memory, implying they might be better choices for systems with limited computational resources. Nevertheless, their execution time, particularly for SSD Mobilenet, is remarkably lower, implying faster detection speeds which could be crucial in real-time scenarios. The YOLO V4 model showcases a decent balance between precision, represented by an mAP of 78.83, and efficiency, with a reasonable number of parameters, lower memory usage, and minimal total execution time. In contrast, Faster R-CNN ResNet 101 and R-FCN ResNet 101 both exhibit a high mAP (95.08 and 95.15 respectively) and a large number of parameters, indicating good model performance but potentially at the cost of increased computational complexity.

Finally, the Faster R-CNN Inception V2 is selected as a base model and finetuned with GTSBR dataset. The TSD model was implemented on the edge device and evaluated on our self-collected dataset. The result shows that the base model performs well in extracting traffic signs from the background with accuracy rate of 98.34%. As traffic signs in the US differ from German, this model performs poorly on classification. Hence, this study collected additional 5000 traffic signs in Washington and manually labeled them into 43 classes which will be used for traffic sign recognition model training.

Table 4-2 TSD Model Baseline Comparison

model	mAP	parameters	memory	total_exec_millis
Faster R-CNN Resnet 50	91.52	43337242	5256.454615	104.0363553
Faster R-CNN Resnet 101	95.08	62381593	6134.705805	123.2729175
Faster R-CNN Inception V2	90.62	12891249	2175.206857	58.53338971
Faster R-CNN Inception Resnet V2	95.77	59412281	18250.446008	442.2206796
R-FCN Resnet 101	95.15	64594585	3509.75153	85.45207971
SSD Mobilenet	61.64	5572809	94.696119	15.14525
SSD Inception V2	66.10	13474849	284.512918	23.74428378
YOLO V2	78.83	50588958	1318.108256	21.4810122

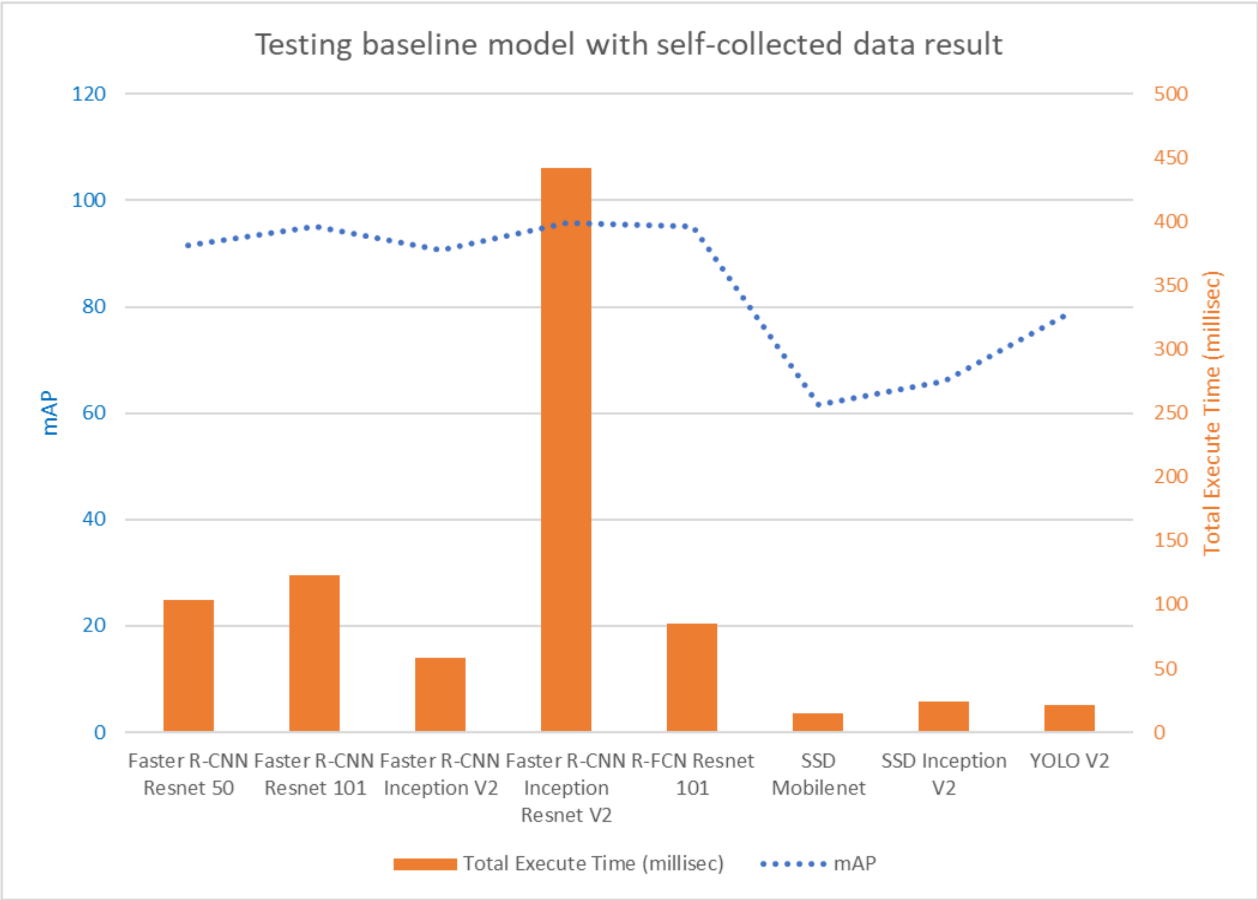

Figure 4-3. Testing TSD baseline model with self-collected data result

Section 5 Traffic Sign Recognition (TSR) Model Development

Despite the initial classification process executed by the Traffic Sign Detection (TSD) model, recognizing and accurately categorizing 42 distinct traffic signs remains a formidable task. This is due to the complexity and variability inherent in the features extracted by the TSD models. To address this challenge and refine the classification process, the research team devised a subsequent Traffic Sign Recognition (TSR) model. This model processes the traffic signs detected by the TSD model and categorizes them into predefined classes, encompassing a total of 42 different traffic signs. These results are then stored for additional processing and as a reference for future queries.

The objective of the proposed TSR model was to comprehensively classify the collected traffic sign images, thereby enhancing the depth and precision of the TSD model output. To realize this goal, the research team constructed an advanced deep Convolutional Neural Network (CNN) model. This model is designed to capture multi-dimensional features from the image data, allowing for a more granular and accurate traffic sign classification. The CNN model employed is not only complex but also adaptable, capable of learning from an extensive range of data inputs to refine its categorization abilities over time.

Subsection 5.1 Model Structure Design

The architecture of the TSR model is comprehensively visualized in Figure 5-1. This diagram illustrates the three stages in the TSR model, from the initial input of the traffic sign data through to the detailed classification output.

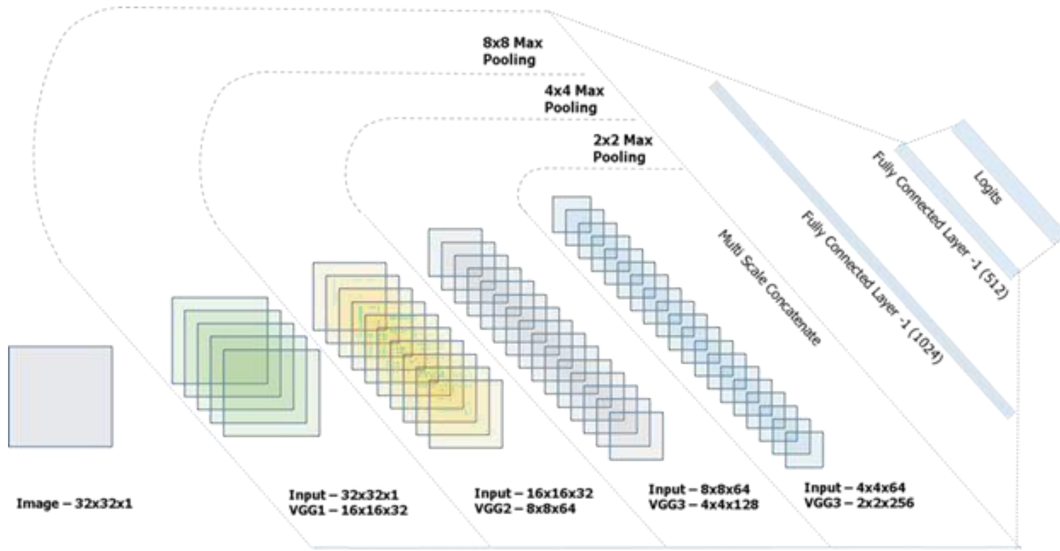

Figure 5-1. The Architecture of Traffic Sign Recognition Model

Convolutional Layer: Initially, a sketch of the RGB image of a detected traffic sign, as identified by the TSD model, is fed into four distinct convolutional layers. The first two layers comprise 32 filters each, while the subsequent two layers incorporate 64 filters each. All these layers employ a kernel size of 3x3 and leverage the ReLU (Rectified Linear Unit) activation function. This function plays a crucial role in transforming the summed weighted input into the output for each node.

Pooling Layer: Following the convolutional layers, the feature maps progress through two pooling layers, each characterized by a pool size of 2x2. These pooling layers function to diminish the feature maps' dimensionality, specifically by reducing their spatial size by a factor of 2. The result is a more efficient and streamlined model, optimizing both the computational load and the model's predictive performance.

Dropout Layer: Simultaneously, during the training phase, two dropout layers are implemented. These layers are designed to disable 25% of the connections between layers intermittently. This tactic is a proactive measure against overfitting, promoting a more robust generalization capability within the model.

Flattening and Fully Connected Layer: Finally, the feature maps are flattened to reshape them into a format suitable for the fully connected layers. They then navigate through two of these layers. The initial layer is equipped with 512 neurons and continues to utilize the ReLU activation function. Meanwhile, the second layer encompasses 43 neurons, a number that directly correlates with the various traffic sign classes. The final layer adopts the softmax activation function, which generates a probability distribution over the classes. The softmax function thus ensures that the model's output represents a valid probability distribution, where the values lie between 0 and 1, and the sum of all probabilities is equal to 1. This approach facilitates the efficient classification of the traffic signs into their respective categories.

Subsection 5.2 Performance Evaluation

While the Traffic Sign Detection (TSD) model proposed in this research paper excels at distinguishing traffic signs from the background, it lacks precision when it comes to the task of detailed traffic sign classification. To overcome this limitation, the paper introduces a complementary Traffic Sign Recognition (TSR) model, designed specifically to handle the granular classification of traffic signs.

However, a key challenge was that the existing systems were primarily trained using traffic sign data collected in Germany and China. As such, these models were ill-suited to the distinct characteristics of the U.S. traffic environment. To address this, the research team proposed a customized TSR model, trained using a comprehensive amalgamation of five datasets - namely, GTSRB, TT-100K, LISA, GSV, and the self-collected dataset. The use of these diverse datasets aimed to increase the model's capability to generalize effectively across different traffic environments.

The training process employed the Adam optimizer and categorical cross-entropy loss function to optimize model performance. The training data was split with an 80-20 ratio, where 80% was used for training, and the remaining 20% was used for validation. The model was trained for 30 epochs with a batch size of 64, a configuration designed to balance computational efficiency and model performance.

Upon testing on the GTSRB and self-collected datasets, the TSR model yielded a high accuracy of 97.1%, demonstrating its effectiveness in accurately recognizing and classifying traffic signs. The test results are illustrated in Figure 5-2, while Figure 5-3 visualizes the output of the model's recognition process. Overall, the TSR model shows promising results in the more nuanced task of traffic sign recognition, complementing the capabilities of the TSD model and contributing towards the overarching goal of developing a comprehensive and efficient traffic sign detection and recognition system.

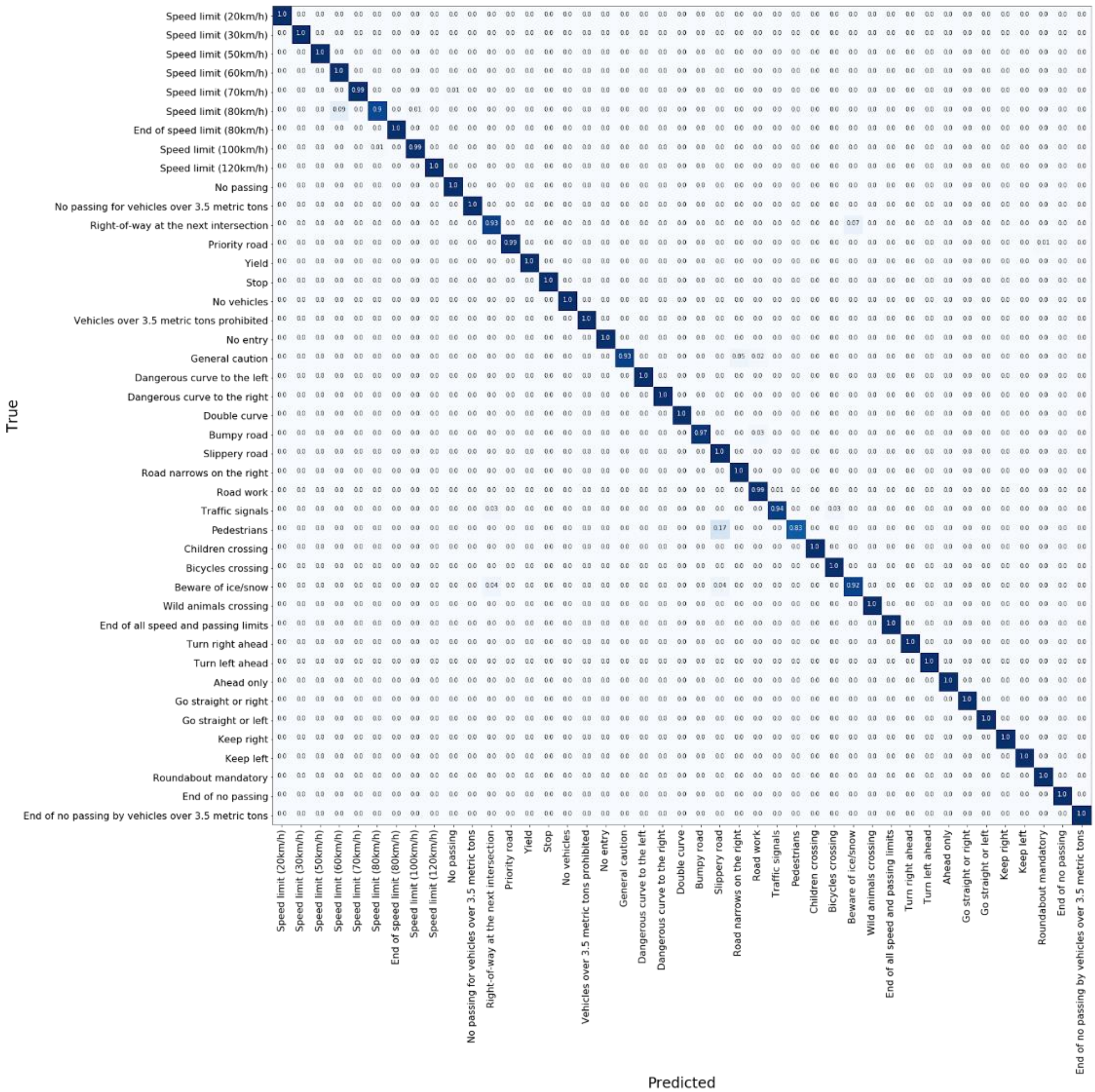

Figure 5-2. Test Results for TSR models

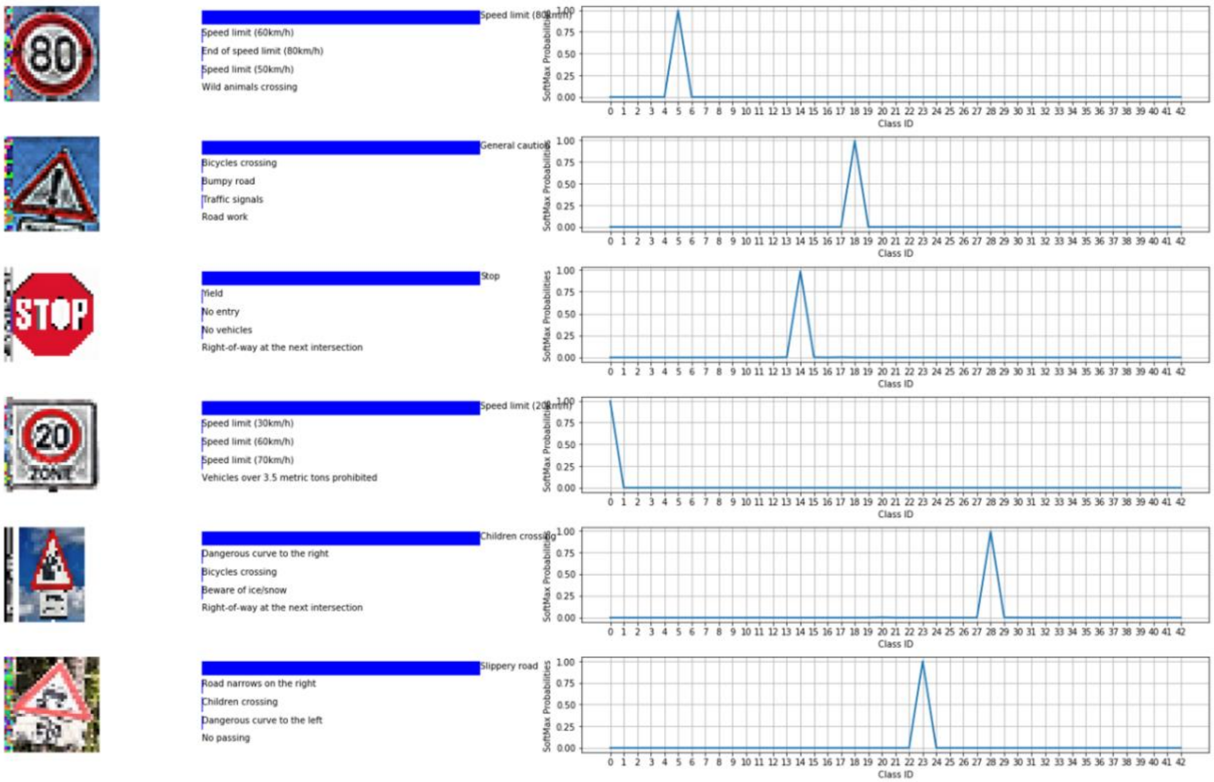

Figure 5-3. Visualized Traffic Sign Recognition Results

Section 6 Data Inventory Buildup

In this research, we devised an advanced data collection system that autonomously captures, detects, classifies, and stores traffic signs. Our initial step involved outfitting test vehicles with bespoke camera sensors tailored for traffic sign capture. This raw data, once captured, is then relayed to an Edge Device using a local area network.

What sets this system apart is the integration of the TSDR model within the Edge Device. When the dash camera captures videos and images, the Edge Device promptly processes this data for traffic sign detection and recognition. Consequently, only pertinent data — specifically, the messages and images related to the traffic signs — are forwarded to the server, ensuring efficient use of bandwidth. A notable advantage of this approach is its ability to uphold privacy standards. By omitting raw data that might contain sensitive information like faces or license plates, the system reduces potential privacy concerns.

Communication between the Edge Device and the server leverages the User Datagram Protocol (UDP). This protocol is apt for scenarios where real-time error correction isn't a critical requirement, offering a balance between speed and data integrity. Once received, both the image and message data are systematically archived in server folders, segmented into 43 distinct categories, as depicted in Figure 6- 1.

To ensure the highest quality of collected data, our team maintained a rigorous review process. Throughout the project's duration, data results were consistently scrutinized, and any discrepancies were corrected on the back server. This meticulous approach culminated in the creation of a curated and organized inventory of traffic sign samples. This invaluable repository not only serves as a foundation for developing sophisticated machine learning models but also aids in pivotal asset management tasks such as maintenance scheduling, risk evaluation, and strategic investment planning.

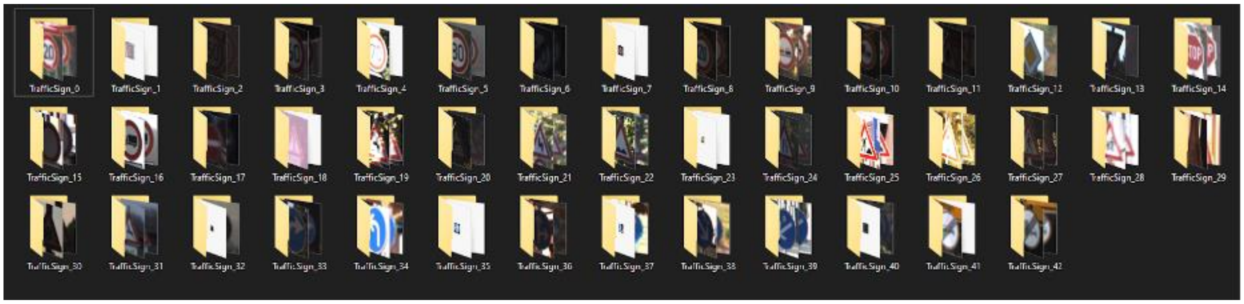

Figure 6-1. Processed Traffic Signs Data in the STARLab Server

Section 7 Outreach and Technical Transfer

The proposed research can help advance cooperative perception algorithms of the roadside and on-board units for driving assistance. Particularly, this project developed a prototype system to improve the effectiveness and accuracy of the driving assistance system for roadway traffic and parking assistance.

The research outcomes are disseminated in the following ways:

1. The proposed Traffic Sign Detection and Recognition (TSDR)system has been accepted by ASCE ICTD 2023 Conference. The work has been presented as a poster presentation on June 15th in Austin, Texas.

2. The proposed Traffic Sign Detection and Recognition (TSDR) system has been accepted by the Transportation Research Board (TRB) 2024 Annual Meeting for presentation. The work will be presented as a poster presentation in January 2024 in Washington D.C.

3. The proposed traffic sign data collection system has been implemented on three test vehicles for real-time data collection. The research paper has been done; we will revise it for journal publication.

4. Results and findings of the prototype system were disseminated to transportation agencies including WSDOT, the City of Bellevue, and PacTrans through webinar. The sample dataset will be uploaded to the CIP system.

Section 8 Conclusion and Discussion

This study has developed a traffic sign detection and recognition model. Existing traffic sign detests, such as GTSRB, Tsinghua-Tencent 100k, and LISA were used for traffic sign detection, however, the recognition model did not perform well on the U.S. traffic sign data. This study collected additional 5000 traffic signs in Washington data from self-collected installed dash cameras and Google Maps Street View. The data were manually labeled into 43 classes and used to fine-tune the model for higher accuracy. An automated pipeline for the collection and classification is developed by using edge devices. The prototype system can provide cost-efficiency and preserve privacy in data collection. Lastly, a sample data inventory is built in a structured and accessible format, which will be beneficial for asset management activities.

In addition to asset management, traffic sign data inventory also benefits traffic control, road safety, and infrastructure planning. Self-collected data can serve as a valuable resource for future model training, validation, or testing of TSDR model development. As the system stores both text and images of detected traffic signs, the conditions of the object can be investigated visually. Future works can use this database to develop an algorithm for traffic sign damage detection automatically.

References

[1]. Supreeth, H.S., & Patil, C.M. (2016). An approach towards efficient detection and recognition of traffic signs in videos using neural networks. 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 456-459.

[2]. Qin, F., Fang, B., & Zhao, H. (2010). Traffic Sign Segmentation and Recognition in Scene Images.2010 Chinese Conference on Pattern Recognition (CCPR), 1-5.

[3]. Xing, Ma & Chunyang, Mu & Yan, Wang & Xiaolong, Wang & Xuetao, Chen. (2016). Traffic sign detection and recognition using color standardization and Zernike moments. 5195-5198. 10.1109/CCDC.2016.7531926.

[4]. Nguwi, Y.Y., & Kouzani, A.Z. (2006). Automatic Road Sign Recognition Using Neural Networks. The 2006 IEEE International Joint Conference on Neural Network Proceedings, 3955-3962.

[5]. Romdhane, N. B., Mliki, H., & Hammami, M. (2016, June). An improved traffic signs recognition and tracking method for driver assistance system. In 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS) (pp. 1-6). IEEE.

[6]. Y. R. Fatmehsari, A. Ghahari and R. A. Zoroofi, "Gabor wavelet for road sign detection and recognition using hybrid classifier", 2010 International Conference on Multimedia Computing and Information Technology (MCIT), pp. 25-28, 2010.
[7]. W. Farhat, H. Faiedh, C. Souani and K. Besbes, "Effect of color spaces on video segmentation performances," 2015 World Symposium on Computer Networks and Information Security (WSCNIS), Hammamet, Tunisia, 2015, pp. 1-5, doi: 10.1109/WSCNIS.2015.7368293.
[8]. Zheng Z., Zhang H., Wang B., Gao Z. (2012), Robust traffic sign recognition and tracking for advanced driver assistance systems, International IEEE Conference on Intelligent Transportation Systems (ITSC), 704 - 709. 10.1109/ITSC.2012.6338799
[9]. Garrido, Miguel & Sotelo, Miguel-Angel & Martm-Gorostiza, E.. (2006). Fast traffic sign detection and recognition under changing lighting conditions. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC. 811 - 816. 10.1109/ITSC.2006.1706843.
[10]. Barnes, Nick & Zelinsky, A.. (2004). Real-time radial symmetry for speed sign detection. 566 - 571.10.1109/IVS.2004.1336446.
[11]. Gavrila, D.M. (1999). Traffic Sign Recognition Revisited. In: Förstner, W., Buhmann, J.M., Faber, A., Faber, P. (eds) Mustererkennung 1999. Informatik aktuell. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-60243-6_10
[12]. E. Oruklu, D. Pesty, J. Neveux and J. E. Guebey, "Real-time traffic sign detection and recognition for in-car driver assistance systems," 2012 IEEE 55th International Midwest Symposium on Circuits and Systems (MWSCAS), Boise, ID, 2012, pp. 976-979.
[13]. Farhat, Wajdi & Faiedh, Hassene & Souani, Chokri & Besbes, Kamel. (2016). Real-time recognition of road traffic signs in video scenes. 125-130. 10.1109/ATSIP.2016.7523081.
[14]. Z. Malik and I. Siddiqi, "Detection and Recognition of Traffic Signs from Road Scene Images," 2014 12th International Conference on Frontiers of Information Technology, Islamabad, 2014, pp. 330-335.
[15]. Qian, Rongqiang & Zhang, Bailing & Yue, Yong & Wang, Zhao & Coenen, Frans. (2015). Robust chinese traffic sign detection and recognition with deep convolutional neural network. 791-796. 10.1109/ICNC.2015.7378092.
[16]. M. Kobayashi, M. Baba, K. Ohtani and L. Li, "A method for traffic sign detection and recognition based on genetic algorithm," 2015 IEEE/SICE International Symposium on System Integration (SII), Nagoya, Japan, 2015, pp. 455-460, doi: 10.1109/SII.2015.7405022.
[17]. M. S. Hossain, M. M. Hasan, M. Ameer Ali, M. H. Kabir and A. B. M. Shawkat Ali, "Automatic detection and recognition of traffic signs," 2010 IEEE Conference on Robotics, Automation and Mechatronics, Singapore, 2010, pp. 286-291, doi: 10.1109/RAMECH.2010.5513177.
[18]. Zhang, H. & Wang, B. & Zheng, Z. & Dai, Y.. (2013). A novel detection and recognition system for Chinese traffic signs. Chinese Control Conference, CCC. 8102-8107.
[19]. Für Neuroinformatik, I. (2019). German Traffic Sign Recognition Benchmark.
[20]. Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li and S. Hu, "Traffic-Sign Detection and Classification in the Wild," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 2110-2118, doi: 10.1109/CVPR.2016.232.
[21]. A. Mogelmose, M. M. Trivedi and T. B. Moeslund, "Vision-Based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems: Perspectives and Survey," in IEEE Transactions on Intelligent Transportation Systems, vol. 13, no. 4, pp. 1484-1497, Dec. 2012, doi: 10.1109/TITS.2012.2209421.

[22]. Zhou X, Ke R, Yang H, Liu C. When Intelligent Transportation Systems Sensing Meets Edge Computing: Vision and Challenges. Applied Sciences. 2021; 11(20):9680. https://doi.org/10.3390/app11209680
[23]. H. Yang, J. Cai, C. Liu, R. Ke, Y. Wang, Cooperative multi-camera vehicle tracking and traffic surveillance with edge artificial intelligence and representation learning, Transportation Research Part C: Emerging Technologies, Volume 148, 2023, 103982, ISSN 0968-090X, https://doi.org/10.1016/j.trc.2022.103982.
[24]. Liu, C., Yang, H., Ke, R., Sun, W., Wang, J., & Wang, Y. (2023). Cooperative and Comprehensive Multi-Task Surveillance Sensing and Interaction System Empowered by Edge Artificial Intelligence. Transportation Research Record, 2677(9), 652–668. https://doi.org/10.1177/03611981231160174
[25]. Liu, Chenxi & Yang, Hao & Zhu, Meixin & Wang, Feilong & Vaa, Torgeir & Wang, Yinhai. (2023). Real-Time Multi-Task Environmental Perception System for Traffic Safety Empowered by Edge Artificial Intelligence. IEEE Transactions on Intelligent Transportation Systems. PP. 1-15.10.1109/TITS.2023.3309100.
[26]. R. Ke, C. Liu, H. Yang, W. Sun and Y. Wang, "Real-Time Traffic and Road Surveillance With Parallel Edge Intelligence," in IEEE Journal of Radio Frequency Identification, vol. 6, pp. 693-696, 2022, doi: 10.1109/JRFID.2022.3211262.