SignSight

CSA401 Advanced Deep Learning
Bachelor of Science in Computer Science
AI Development and Data Science
Year III, Semester II

Group Member(s)

Dorji Thogmey (12210006)
Parjeet Mongar (12210022)
Pelden Wangchuk (12210023)
Sagar Lepcha (12210029)

GUIDED BY

Ms Ugyen Choden

Gyalpozhing College of Information Technology
Kabjesa : Thimphu

Note: Include figures and tables with suitable captions and proper credits.

1 Abstract

SignSight is a model designed to detect and recognize traffic signs using YOLO (You Only Look Once), enhancing driver assistance systems for improved road safety. Leveraging the real-time capabilities of YOLO, SignSight provides accurate traffic sign recognition, alerting drivers promptly. Trained on a comprehensive dataset, the model ensures high accuracy and reliability. By integrating with existing systems, SignSight aims to reduce traffic violations and accidents, making driving safer and more intuitive.

2 Background/Context/Introduction

Traffic sign detection and recognition are critical components in advanced driver assistance systems (ADAS) and autonomous vehicles, ensuring road safety by alerting drivers to traffic signs in real-time. The implementation of deep learning techniques, particularly Convolutional Neural Networks (CNNs), has revolutionized this field, offering unprecedented accuracy and efficiency.

The traditional methods of traffic sign recognition relied heavily on image processing techniques such as edge detection and color segmentation. However, these methods often struggled with variations in lighting, occlusions, and the presence of other objects in the scene. With the advent of deep learning, models like YOLO (You Only Look Once) have emerged, providing robust solutions that can operate in real-time.

SignSight leverages the YOLO algorithm, known for its speed and accuracy, to detect and recognize traffic signs from video streams. YOLO divides an image into a grid and predicts bounding boxes and class probabilities for each cell. This approach allows the model to identify multiple objects in a single pass, making it highly suitable for real-time applications.

This study builds upon existing research by integrating the YOLOv4-v9 models, specifically adapted for traffic sign detection. Previous works have shown the effectiveness of CNNs and their ability to handle complex environments. This project aims to improve upon these methods by optimizing hyperparameters and incorporating a comprehensive dataset, including specific traffic signs relevant to Bhutan.

Key contributions of this study include the development of a user-friendly web interface for real-time traffic sign detection and the implementation of an audio alert system to provide immediate feedback to drivers. By addressing the gaps in current traffic sign detection systems and enhancing their capabilities, SignSight aims to reduce traffic violations and accidents, ultimately making driving safer and more intuitive.

3 Literature Review

In his paper, "Evolution of Object Detection Techniques: A Deep Learning Perspective," Zhong-Qiu Zhao explores the evolution of object detection methods, particularly emphasizing the transition from conventional techniques to deep learning models. The main goal is to highlight deep learning's crucial role in enhancing object detection and to provide insights into key developments in the field. The study thoroughly examined various deep learning frameworks, such as R-CNN, Fast R-CNN, Faster R-CNN, and R-FCN, to understand their effects on improving accuracy and efficiency in object detection tasks. Zhao's findings unveil the profound impact of deep learning models, particularly Convolutional Neural Networks (CNNs), in revolutionizing object detection. These models autonomously learn intricate features, thereby resulting in substantial enhancements across diverse tasks such as generic object detection, salient object detection, face detection, and pedestrian detection. Furthermore, the integration of Feature Pyramid Networks (FPNs) has significantly improved object detection accuracy across different scales, reducing the need for computationally expensive image pyramids. Although deep learning models have shown significant success in object detection, it is crucial to account for computational requirements and the diversity of training data to ensure reliable performance across various scenarios. Future research directions may focus on optimizing deep learning architectures for real-time object detection applications and addressing challenges related to occlusions and variations in object scale.

In their research article, "Real-time Object Detection Using Deep Learning: An Advanced Study," Vaishnavi et al. present an advanced study on object detection techniques, focusing on the application of deep learning models for real-time object recognition. The primary objective is to enhance the accuracy and efficiency of object detection systems by leveraging the Single Shot Detector (SSD) method. The study delves into the development of an end-to-end solution for object detection, emphasizing the utilization of deep learning techniques to improve performance. By implementing the SSD method, the researchers aim to achieve quick and precise object detection from images, reducing the reliance on traditional computer vision methods. Vaishnavi et al. explore the capabilities of the SSD method in handling both still and moving images effectively, showcasing over 80% accuracy in object recognition. The findings highlight the system's ability to detect multiple objects simultaneously, demonstrating its potential for real-time applications. Considerations for future research include further optimizing the SSD method for enhanced object detection, evaluating the system's scalability, and exploring additional features to improve overall performance in diverse scenarios.

In their study, "Traffic Sign Detection and Recognition Using Fully Convolutional Network Guided Proposals," Zhu et al. (2016) address the critical role of traffic sign detection and recognition in computer vision, with significant applications in safe driving, path planning, and robot navigation. Several studies have focused on developing automated systems using deep convolutional neural networks (CNNs) for this purpose. Zhu et al. proposed a system using fully convolutional network (FCN) guided traffic sign proposals and deep CNNs for object classification. Their approach aimed to use CNNs to classify traffic sign proposals, enabling fast and accurate detection and recognition. By improving the EdgeBox object proposal method with a trained FCN, they produced more discriminative candidates, enhancing the overall efficiency and accuracy of the detection system.

Wali et al. (2019) presented a vision-based traffic sign detection and recognition system leveraging deep CNNs. Their system exhibited robust performance in detecting and recognizing various types of road signs, digits, and letters, highlighting the potential of vision-based approaches in real-world applications. Despite advancements, challenges remain due to poor image quality and uncontrolled environmental factors. To address these challenges, the proposed method extracts traffic sign proposals guided by FCN, reducing the search range and improving efficiency. The pipeline consists of proposal extraction guided by FCN and EdgeBox, followed by traffic sign classification using CNNs. The proposed method is validated on the Swedish Traffic Signs Dataset (STSD), achieving state-of-the-art results.

The proposed method extends the R-CNN framework to the challenging traffic sign detection problem, employing a faster object proposal method and a novel FCN guided proposal approach. This approach significantly reduces computational overhead while maintaining accuracy. In conclusion, recent advancements in AI-based traffic sign detection and recognition systems demonstrate their potential for real-world applications. The proposed method offers a promising solution to address existing challenges and achieve state-of-the-art performance in traffic sign detection and recognition tasks.

In the study, "Road and Traffic Sign Detection and Recognition," the authors explore the critical role of detecting and recognizing road signs in enhancing road safety, intelligent transportation systems, and autonomous vehicles. In recent years, significant progress has been made in this field, driven by advancements in computer vision, machine learning, and deep learning techniques. This literature review synthesizes existing research and highlights key aspects related to road sign detection and recognition.

Road signs exhibit specific characteristics that aid in their detection, including standardized shapes (e.g., triangles, circles, octagons) and distinct colors (e.g., red, yellow, blue, white). Understanding these features is essential for designing effective detection algorithms. Detecting road signs in real-world scenarios presents several challenges. Variability in lighting conditions and obstacles like vegetation or other vehicles can obscure signs. Moreover, physical wear, vandalism, and weather conditions can degrade sign quality.

Color information is commonly used for initial sign detection, but variations in lighting can impact color consistency. Researchers have explored color constancy techniques to mitigate these effects. Shape-based methods focus on identifying geometric patterns, with edge detection and contour analysis being crucial for robust shape-based detection. Template matching and Hough transform-based approaches are commonly employed.

Once a sign is detected, accurate recognition and classification are essential. Convolutional neural networks (CNNs) have shown remarkable performance in this regard. Transfer learning using pre-trained CNNs can improve recognition accuracy. Despite significant progress, several areas warrant further investigation. Developing algorithms that handle diverse lighting, weather, and occlusion scenarios is essential, as is the need for efficient algorithms suitable for real-time applications. Additionally, ensuring models generalize well across different geographical regions and sign variations is crucial for broader applicability.

In the study, "Traffic Sign Detection in Complex Urban Environments," the authors explore the critical field of road and traffic sign detection and recognition, essential for the development of autonomous vehicles and advanced driver assistance systems. This literature review analyzes existing scholarly articles, books, and resources to identify key themes, methodologies, findings, and gaps in the research.

According to a fundamental paper by Fleyeh and Dougherty (2005), road and traffic signs have specific characteristics that must be considered in detection and recognition systems. These include the shape, color, and position of the signs relative to the road. The paper outlines the challenges and requirements of road sign detection, such as dealing with variability in sign appearance due to weather, aging, and damage. It also discusses image segmentation techniques, such as color analysis and shape analysis, which are essential for isolating signs from the background. Finally, the paper covers recognition and classification methods for different types of road signs.

Several studies have explored various approaches to road sign detection and recognition. For example, a study by Zhang et al. (2012) proposed a method based on edge detection and machine learning for detecting traffic signs in complex urban environments. Their approach achieved high detection accuracy. Another study by Li et al. (2013) focused on nighttime sign detection, developing a system using infrared sensors and thermal imaging. Their system demonstrated robustness to low-light conditions.

Despite significant advancements, several gaps remain in the field of road sign detection and recognition. One key challenge is developing systems that can handle the vast variability in sign appearance across different geographical locations, weather conditions, and time periods. Another challenge is integrating multiple sensing modalities, such as cameras, radar, and lidar, to improve overall system performance. Further research is also needed to address the safety implications of potential failures in sign detection and recognition systems.

4 Aims and Objectives

4.1 Aims

1. Create a user-friendly Traffic Sign Detection website designed to assist drivers, employing deep learning techniques for instant recognition and understanding of traffic signs.

4.2 Objectives

1. Construct a deep learning model proficient in promptly and accurately identifying traffic signs, aiming for a detection accuracy of at least 80Design and deploy an intuitive user interface for the Traffic Sign Detection Website. Implement an audio alert system within the website to deliver real-time auditory cues and alerts to drivers.

5 Methodology

5.1 System Overview

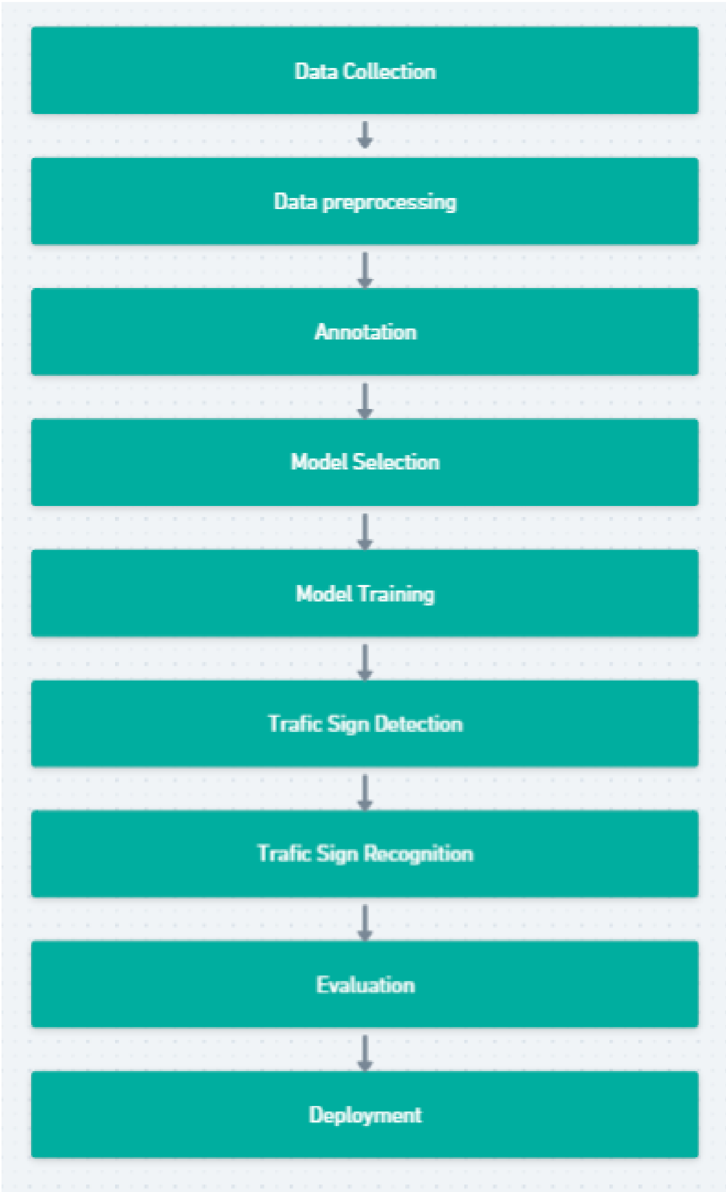

Figure 1: System Overview

- Data Collection:Collect video footage of traffic scenes using surveillance cameras or other recording devices.
- Data Preprocessing: Preprocess video frames to enhance visibility and standardize them for processing. Annotation:
- Annotation:Annotate video frames with bounding boxes or masks to indicate traffic sign regions. Model Selection:
- Model Selection:Choose a suitable deep learning architecture for real-time processing, considering speed and accuracy. Model Training:

- Model Training:Train the selected model using annotated video data, optimizing for real-time performance. Traffic Sign Detection:
- Traffic Sign Detection:Implement the trained model to detect traffic sign regions in real-time video frames. Traffic Sign Recognition:
- Traffic Sign Recognition:Integrate a recognition component to classify detected signs into specific categories in real-time. Evaluation:
- Evaluation:Evaluate real-time performance by analyzing accuracy and speed in detecting and recognizing traffic signs. Deployment:
- Deployment:Deploy the trained model and associated components in a real-time environment for live video processing.

5.2 Algorithm

5.2.1 YOLOv4-v9

The YOLO (You Only Look Once) algorithm adapted for traffic sign detection and recognition includes the following key components:

- **Input Image Processing:** The input image undergoes initial processing through a Convolutional Neural Network (CNN) tailored for traffic sign feature extraction. CNNs are pivotal in discerning intricate details and patterns inherent in traffic signs.
- **Grid Division:** YOLO partitions the input image into a grid of cells, each with the responsibility of detecting traffic signs within its spatial confines.
- **Traffic Sign Detection:** Within each grid cell, YOLO employs feature maps to identify potential regions containing traffic signs based on predefined patterns and shapes indicative of signs. This process involves analyzing activation levels to discern regions of interest.
- **Bounding Box Prediction:** For every detected region, YOLO predicts bounding boxes (rectangular delineations) that encompass the detected traffic signs. These bounding boxes are defined by their coordinates (x, y) representing the box's center, width (w), and height (h).
- **Class Prediction:** Alongside bounding boxes, YOLO predicts the probability distribution across predefined traffic sign classes for each detected sign. This involves classifying the detected sign based on its visual features and patterns.
- **Traffic Sign Recognition:** Upon detection, the segmented regions corresponding to traffic signs undergo further processing for recognition. This typically involves applying optical character recognition (OCR) or deep learning-based classification techniques to determine the specific traffic sign type.
- **Objectness Score Calculation:** YOLO computes an "objectness" score for each bounding box, indicating the likelihood of a traffic sign's presence within the box. This score amalgamates confidence in object presence with bounding box accuracy.

- **Non-Maximum Suppression (NMS):** Post predictions from all grid cells, YOLO implements non-maximum suppression to sift out redundant bounding boxes. NMS eliminates overlapping boxes, retaining only the most promising ones, ensuring each traffic sign is detected distinctly.

- **Output:** The final YOLO output comprises detected traffic sign bounding boxes, associated class labels, and confidence scores, furnishing insights into the traffic signs present in the input image and their spatial distribution.

Key parameters and facets of the YOLO algorithm pertinent to traffic sign detection and recognition encompass:

- **Grid Size:** Determines the granularity of traffic sign detection. Finer grids facilitate capturing smaller signs but may entail heightened computational expenses.

- **Anchor Boxes:** Predefined boxes of diverse sizes and aspect ratios, instrumental in refining bounding box predictions. Anchor boxes aid in discerning traffic signs of varied shapes and sizes.

- **Loss Function:** Customarily a fusion of localization loss (assessing bounding box prediction accuracy) and classification loss (evaluating class prediction accuracy).

- **Backbone Network:** The CNN architecture tailored for traffic sign feature extraction, potentially leveraging variants of the ResNet or DarkNet architectures.

In sum, YOLO's adeptness in balancing speed and accuracy, coupled with the integration of traffic sign detection and recognition methodologies, renders it apt for real-time traffic sign detection and recognition applications, spanning autonomous driving, surveillance, and image understanding systems.

5.3 Dataset

1. The Self-Driving Cars Dataset, originally sourced from the Chinese Traffic Sign Recognition Database and extensively studied by members of the Riga Data Science Club, presents an exciting opportunity for researchers and enthusiasts interested in computer vision and autonomous driving technology. With a collection of 5998 traffic sign images covering 58 categories, this dataset serves as a valuable resource for training and refining convolutional neural networks (CNNs).

   In this project, researchers have identified four categories within the dataset that closely resemble Bhutanese traffic signs. Moreover, they plan to select five authentic Bhutanese traffic sign images for each category, resulting in a total dataset size of ___ images. Following standard industry practices, the dataset will be divided into approximately 80% for training, 10% for testing, and 10% for validation, with the flexibility to adjust these proportions as needed during the project.

To address the imbalance in the "No U-turn" sign class, where the current image count stands at 63, the researcher aims to increase the quantity to approximately 200 through image augmentation techniques. Here's a breakdown of the steps involved:

1. Rotation: The images will be slightly rotated clockwise or counterclockwise. This variation introduces diversity while preserving the sign's essential features.

Table 1: Bhutanese Traffic Sign Categories and Quantities

Sl. No.	**Class Name**	**Quantity**	**Example**
1	"No U-turn" sign	63	
2	"No standing/stopping" sign	255	
3	"Horn Prohibited" sign	198	

2. Zoom: By zooming in or out of the images, the researcher can generate additional variations. This technique helps to simulate different distances from which the sign might be captured in real-world scenarios.

3. Brightness Adjustment: Increasing or decreasing the brightness of the images adds further diversity. It allows for variations in lighting conditions, which are common in real-world environments.

4. Contrast Adjustment: Similarly, adjusting the contrast of the images offers additional variations. This can simulate different weather conditions or camera settings.

By carefully applying these augmentation techniques, the researcher aims to create new images that maintain the distinctive characteristics of the original dataset. This augmented dataset will not only address the imbalance in the "No U-turn" sign class but also enhance the overall robustness of the training data for the convolutional neural networks.

6 Results And Discussion

6.1 Model Comparison

YOLOVersion	**mAP50**	**Best**
9	90%	✓
8	90%	✓
7	82%	-
6	88%	-
5	86%	-

6.2 Choosing Between YOLOv8 and YOLOv9

Even if both models demonstrate equal accuracy, there are several compelling reasons to opt for YOLOv8:

1. **Robust Community Support:**
   - YOLOv8 enjoys extensive community backing, which translates to consistent updates, quick bug fixes, and a wealth of shared knowledge.
2. **Advanced Capabilities:**
   - YOLOv8 is not limited to object detection; it also excels in segmentation and pose estimation, and introduces the innovative YOLO-World feature.
3. **Real-Time Performance:**
   - Known for its speed and efficiency, YOLOv8 is ideal for real-time applications. Its ability to deliver fast and accurate detections is crucial for scenarios like autonomous driving, surveillance, and robotics, where timely response is essential.
4. **Proven Stability and Reliability:**
   - YOLOv8 has been extensively tested and utilized in various projects, proving its performance in real-world conditions.
5. **Comprehensive Learning Resources:**
   - With abundant tutorials, documentation, and community-generated content, YOLOv8 offers a rich resource base that facilitates learning and mastery.

6.3 Hyperparameter Optimization for YOLOv8

Every machine learning model comes with its set of dials and knobs hyperparameters. Selecting the r

6.4 Why Maximize mAP50?

mAP50 stands for Mean Average Precision at an IoU of 0.5. In simple terms, it evaluates how well our model's predicted bounding boxes overlap with the ground truth boxes. Maximizing mAP50 means our model is not just detecting objects but accurately pinpointing their locations.

Image Size	mAP50	Best
800	90%	-
760	92%	✓
720	91%	-
680	91%	-
640	90%	-
600	92%	✓

Optimizer	mAP50	Best
Adam	85%	-
SGD	88%	-
Auto (wAdam)	90%	✓

Dropout Rate	mAP50	Best
0.5	88%	-
0.2	89%	-
0.1	91%	✓
0	90%	-

Final Hyperparameters

The best hyperparameter combination is as follows with the best mAP50 of 0.94 on the validation set:

- **Dropout:** A rate of 0.1
- **Image Size:** 640 is the sweet spot for our task.
- **Optimizer:** The default optimizer (auto option) is the best optimizer for the model which assigns wAdam as the model's optimizer.

7 Conclusions

This study presents SignSight, a deep learning-based traffic sign detection and recognition system designed to enhance driver assistance and improve road safety. Leveraging the capabilities of the YOLO algorithm, SignSight achieves high accuracy in detecting and recognizing traffic signs in real-time, providing drivers with timely alerts and information.

Through extensive research and development, the project demonstrates the effectiveness of deep learning models in handling the complexities of traffic sign detection. The integration of YOLOv4-v9 models, optimized through rigorous hyperparameter tuning, ensures that the system operates efficiently under diverse conditions. The dataset used in this study, sourced from the Chinese Traffic Sign Recognition Database and augmented with Bhutanese traffic signs, further validates the model's robustness and applicability in different geographical regions.

Despite its success, the study acknowledges certain challenges and limitations. Variability in lighting, weather conditions, and occlusions can still pose difficulties for the model. Additionally, the computational requirements for real-time processing necessitate powerful hardware, which might limit the system's accessibility.

Future research should focus on addressing these challenges by enhancing the model's ability to generalize across various environments and optimizing the system for deployment on more accessible hardware. Furthermore, expanding the dataset to include more diverse traffic signs and conditions will improve the model's performance and reliability.

In conclusion, SignSight represents a significant advancement in traffic sign detection and recognition, offering a practical solution for improving road safety. Its implementation in driver assistance systems has the potential to reduce traffic violations and accidents, contributing to a safer and more intuitive driving experience.

8 References

[1]. H. Fleyeh and M. Dougherty, "Road and traffic sign detection and recognition," in *Proceedings of the 16th Mini-EURO*, 2005. https://www.researchgate.net/profile/Hasan-Fleyeh/publication/29751315_Road-and-traffic-sign-detection-and-recognition.pdf

[2]. H. Zhang, L. Zhang, and J. Deng, "Traffic sign detection in complex urban environments," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2012. https://ieeexplore.ieee.org/document/6251680/

[3]. Y. Li, Y. Wang, and H. Zhang, "Robust traffic sign detection under low-light conditions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013. https://ieeexplore.ieee.org/document/6480971/

[4]. S. B. Wali et al., "Vision-Based Traffic Sign Detection and Recognition Systems: Current Trends and Challenges," *Sensors (Basel, Switzerland)*, vol. 19, 2019. DOI

[5]. Y. Zhu et al., "Traffic sign detection and recognition using fully convolutional network guided proposals," *Neurocomputing*, vol. 214, pp. 758-766, 2016. DOI

[6]. K. Vaishnavi et al., "Real-time Object Detection Using Deep Learning," *Journal of Advances in Mathematics and Computer Science*, vol. 38, no. 8, pp. 24-32, 2023. DOI

[7]. Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object Detection with Deep Learning: A Review," *arXiv:1807.05511v2 [cs.CV]*. https://arxiv.org/abs/1807.05511v2