CS405 Machine Learning
Research Final Report

TSDR: Traffic-Sign Detection and Recognition

Langchu Huang 12213009 | Yutong Wu 12213012 | Yuhang Huang 12213015 | Tianye Wu 12211201

***Abstract*— Traffic Sign Detection and Recognition (TSDR) is a critical component in autonomous driving technologies and intelligent transportation systems. This study focuses on enhancing the accuracy, robustness, and efficiency of TSDR by leveraging the YOLO framework with innovative advancements. Transfer learning was employed to fine-tune YOLOv5 models on the TT100K dataset, addressing the challenges posed by limited data. The Squeeze-and-Excitation (SE) attention module was integrated to recalibrate channel-wise feature responses, significantly improving detection performance under diverse conditions. Comprehensive analysis of parameter configurations further optimized model performance, and the final model was successfully deployed on both host systems and AI edge devices, achieving real-time processing at 60rames per second.**

Experimental results demonstrated that the proposed methods achieved state-of-the-art detection accuracy, showcasing the model's applicability in real-world scenarios. This work not only highlights the potential of advanced deep learning techniques in traffic sign detection but also provides a solid foundation for future explorations in intelligent transportation and autonomous systems.

I. Background and Significance

A. Introduction

1) What is TSDR?:

Traffic Sign Detection and Recognition(TSDR) refers to the technologies and processes used to identify and interpret various road signs and signals. It is a critical component of Intelligent Transportation Systems (ITS), Advanced Driving Assistance Systems (ADAS) and autonomous driving. TSDR enables real-time recognition and understanding of road signs, making it an essential technology for modern transportation systems.

2) Contributions:

TSDR plays a vital role in **Advanced Driving Assistance Systems (ADAS)**, **Autonomous Driving**, and **Intelligent Transportation Systems** by ensuring compliance with traffic rules and enhancing road safety. In ADAS, it helps vehicles provide timely warnings and navigation instructions. For autonomous driving, TSDR ensures vehicles can detect and recognize various types of signs, such as warnings, prohibitions, and instructions, contributing to a safe and efficient driving environment. Moreover, TSDR supports intelligent transportation systems by improving traffic flow and enabling the development of smart cities.

3) Driving Technological Advancements:

The development of TSDR promotes research in key technologies such as image recognition, deep learning, and pattern recognition. By applying these technologies to traffic sign detection, TSDR not only validates their effectiveness and adaptability but also provides valuable research experience for broader image recognition applications. Its implementation fosters innovation and accelerates progress in fields beyond transportation, making it a cornerstone for advancements in **computer vision and intelligent systems**.

B. Significance

1) Safety:

A report published in 2015 on Global Status Report On Road Safety by World Health Organization (WHO) showed that over 1.2 million people die across the globe annually due to road accidents [1].Traffic signs play a crucial role in ensuring road safety by providing drivers with essential information such as road conditions, speed limits, potential hazards and so on. Real-time detection and recognition of traffic signs can alert drivers to comply with traffic regulations, thereby reducing the likelihood of accidents. For example, autonomous driving systems rely on traffic sign recognition to determine whether actions such as slowing down, stopping, or yielding are required.

2) ADAS and Autonomous Driving:

Traffic sign detection and recognition is a key component in ADAS and autonomous driving. They must be able to recognize and interpret all traffic signs on the road to ensure the vehicle's safe operation and to make appropriate decisions in complex traffic environments. Without accurate traffic sign recognition, an autonomous system cannot reliably understand its surroundings, limiting its widespread use.

3) Intelligent Transportation Systems:

ITS aim to enhance traffic efficiency, reduce congestion, and minimize accidents. By using traffic sign recognition technology, traffic management systems can monitor road conditions in real time, adjust traffic signals, and provide traffic information. This not only improves traffic flow but also helps traffic authorities conduct precise traffic flow analysis and management.

C. Motivations

1) Technology Evolution:

Traffic Sign Detection and Recognition (TSDR) leverages technologies such as computer vision, deep learning, and neural networks to detect and classify traffic signs. As these technologies continue to advance, TSDR systems become more accurate, efficient, and adaptable to various driving environments and conditions. At the same time, the relentless pursuit of optimization in autonomous driving algorithms also contributes to advancements in underlying technologies, fostering the development of the entire technical ecosystem.

With the increase in computational power and the efficiency of algorithms, TSDR can now operate in real-time. This is critical for autonomous vehicles and Advanced Driver-Assistance Systems (ADAS), which rely on the rapid and accurate detection of traffic signs.

2) Commercial Value:

TSDR, as an important part in ADAS, autonomous driving, and ITS, has immense commercial value. Several studies from Mordor Intelligence Table III shows that, The Advanced Driver Assistance Systems Market size is estimated at USD 49.65 billion in 2024, and is expected to reach USD 107.47 billion by 2029, growing at a CAGR of 16.70% during the forecast period (2024-2029) [2]. The Autonomous Car Market size is estimated at USD 41.10 billion in 2024, and is expected to reach USD 114.54 billion by 2029, growing at a CAGR of 22.75% during the forecast period (2024-2029) [3]. The Smart Transportation Market size is estimated at USD 33.38 billion in 2024, and is expected to reach USD 46.36 billion by 2029, growing at a CAGR of 6.79% during the forecast period (2024-2029) [4].

II. Analysis of Current Research Status

A. Datasets and Benchmarks

1) Foreign:

Road-sign-detection Dataset [5]:

Contains 877 images across 4 classes for road sign detection. Bounding box annotations are provided in the PASCAL VOC format.

GTSDB (German Traffic Sign Detection Benchmark) Dataset [6]:

Released by Karlsruhe Institute of Technology (KIT), Germany. A large-scale dataset with 5183 images of 43 distinct traffic sign types from real-world traffic scenarios in Germany. Split into 4080 training images and 1103 testing images. Each image includes bounding box annotations and the corresponding sign type. The dataset also provides evaluation metrics for comparing model performance.

BTSD (Belgium Traffic Sign Dataset) Dataset [7]:

Released and hosted by VISICS, ESAT, KU Leuven. Contains 900 images of traffic signs, categorized into 4 main classes: Prohibited, Hazard, Mandatory, and Other, with 42 sub-classes.

LISA (LISA Traffic Sign Dataset) [8]:

Consists of images and video frames for traffic sign detection, with annotations for 47 types of traffic signs. The dataset includes 6610 frames with 7855 annotations. Image sizes vary from 640x480 to 1024x522 pixels. Annotations include sign type, position, size, occlusion status, and whether the sign is on a secondary road. The complete dataset includes video frames along with images.

2) Chinese:

TT100K Dataset [9]:

A large-scale traffic sign dataset containing 100,000 images. The dataset includes more than 200 classes of traffic signs, with a diverse range of traffic signs from multiple countries. Annotations are in bounding box format, providing detailed information on each sign's location and class. The dataset is designed for both traffic sign detection and recognition tasks.

B. Data Challenges

As the examples shown in Figure 1 from TT100K dataset [9], the data issues in traffic sign datasets are mainly caused by the following factors [10]:

1) Illumination Issues:

Complex lighting variations caused by sunlight, vehicle headlights, streetlights, etc., can result in traffic signs being overexposed, underexposed, or obscured by shadows. Lighting variations affect the color, edges, and contrast of the target. Since recognizing color, edges, and contrast is critical for object detection algorithms, optimized preprocessing methods are necessary for addressing illumination issues in both traditional and neural network architectures.

For instance, analyzing the histogram distribution and statistical moments of the brightness channel can help understand the lighting conditions of an image. Contrast enhancement algorithms and illumination-invariant color models (e.g., HSV) can be used to sharpen edges and preserve color information under low-light conditions.

2) Perspective Issues:

The optical axis of installed cameras may not align perfectly with the target traffic signs, which can lead to image distortion. For example, a circular sign may appear elliptical, and edge lengths may look inconsistent.

The Generalized Hough Transform can largely handle distorted circles. Using an accumulator array and further

TABLE I
MARKET SIZE AND CAGR ESTIMATES (2024-2029)

Market Size and CAGR Estimates (2024-2029)			
Market	**Estimated Size (2024)**	**Estimated Size (2029)**	**CAGR (2024-2029)**
ADAS	USD 49.65 billion	USD 107.47 billion	16.70%
Autonomous driving	USD 41.10 billion	USD 114.54 billion	22.75%
ITS	USD 33.38 billion	USD 46.36 billion	6.79%

Fig. 1. Examples of traffic sign data issue in TT100K.

filtering can resolve distorted or broken circles. For scale changes, scale-invariant features like SIFT and SURF can be employed. Oriented FAST and Rotated BRIEF (ORB) are widely used rotation- and scale-invariant features. Locally Encoded Shape Histograms (LESH) serve as a scale-invariant, shape-based object detection feature. Combining these methods with robust classifiers can address perspective distortions effectively.

3) Motion Blur:

Blurring issues can stem from low resolution, long distances, or camera motion, which weaken the clear edges in the image. Hence, restoration processes are required before detection.

Various algorithms have been developed to extract stable and clear edges from frames, such as optical flow and point spread function modeling. Frequency domain filtering using estimated motion models is also a method for motion blur restoration.

4) Partial Occlusion:

Traffic signs are often occluded by objects like lamp posts, billboards, trees, etc. However, their regular shapes and symmetry allow for detection even under partial occlusion.

Part-based object detection methods, local binary patterns, histograms of oriented gradients, and Randomized Hough Transform are effective algorithms for dealing with occluded signs (e.g., broken edges). Prediction-based tracking algorithms focus on edges rather than frames, resulting in lower error probabilities. Keypoint tracking and kernel-based tracking methods are commonly used for tracking occluded objects.

CCTSDB (Chinese City Traffic Sign Database) Dataset [11]

Contains more than 10,000 images with over 60 categories of Chinese traffic signs. Includes real-world traffic sign images from urban environments in China. Annotated with bounding boxes and detailed labels for each traffic sign. The dataset is intended for traffic sign recognition and detection, focusing on challenges specific to urban Chinese environments, such as diverse traffic signs and variable lighting conditions.

C. Models Architecture and Principle

1) Traditional Computer Vision Method:

Color and Shape Analysis

Based on the traditional computer vision method, traffic sign detection can be achieved by the color and shape analysis (Figure 2 [12]). Color segmentation process eliminates the unnecessary objects and hence it reduces the search area of the image or video frame. The color distance is defined, similar to the Euclidean distance between two points and it is calculated by taking the difference between the two colors, and as the color distance decreases the similarity increases.each pixel in the RGB space is compared with the threshold values defined for each component. Only the component value larger than the threshold is converted to 255 otherwise it is set to 0. Hence, the picture is eliminated, through thresholding in color spaces like RGB, HSV, or this time, HSV can achieve a maximum detective speed of 95% .

As for the shape analysis, it can eliminate the issue in color based detection is the ambient illumination variation since generally traffic signs are triangle, rectangle, octagon and circular using Hough transform, detection rates of 97.2% and 94.3% (Figure 3) are achieved for the speed limit and warning signs respectively [13].

There are also some other color and shape analysis(Figure 4 Figure 5 [14]):

When it comes to the recognition part, it can be done either by using feature matching algorithms or by using

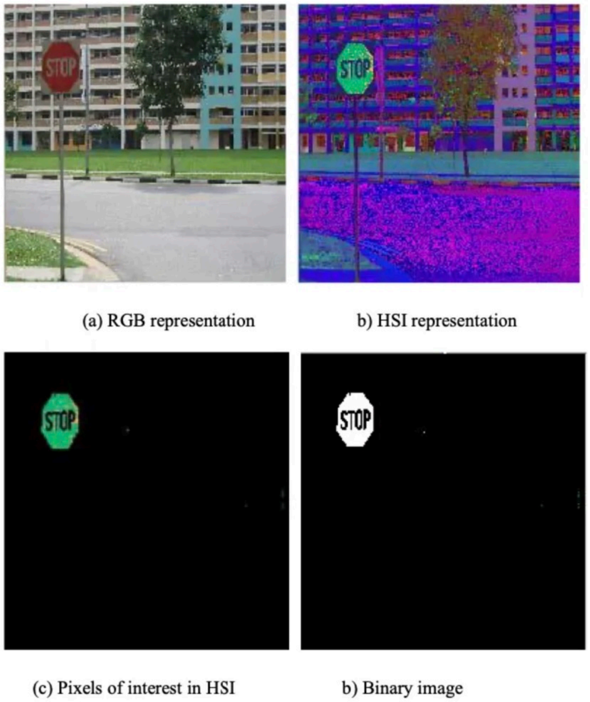

Fig. 2. Color and Shape Analysis.

Color Space	Detection Rate
RGB	88.75%
HSV	95%
HSI	91.35%

Fig. 3. Three Kinds of Color Detection.

machine learning algorithms.Artificial Neural Networks are built with the inspiration of biological neurons. Intelligent behaviour of human beings is attributed to the biological neural network. This kind of system is implemented artificially by using Artificial Neural Network (ANN). The authors [15] in used the auto-associative neural network to recognize the traffic signs in video frames and achieved a recognition rate of 100% and 94.7% in both daylight and shadow environment

Feature-Based Methods

The SVM and Histograms of Oriented Gradients (HOG) (Figure 6) [13] based detection structure was first proposed to detect pedestrians and has been commonly used in different detection problems in the past decade. This structure utilizes HOG-like features to express the objects and treats the object detection problem as an SVM classification problem, in which each candidate is classified into objects or backgrounds.

The SVM based detection structure has been successfully applied in TSD problems.The introduction of HOG-like features is the key of the success of SVM based detection methods. The HOG feature [16] is the most popular feature used in different detection problems. Using classical HOG features, the HOG+SVM based detection methods [17], [18] can achieve high detection results. Using different HOG features can generate more vectors for SVM classification. If different HOG features can express objects well, the performance may be improved. An exhaustive scanning process is used in the SVM-based detection process, which is a time-consuming process for scanning a high-resolution image. Hence, most SVM based detection methods have a ROI extraction process, which can largely reduce the scanning regions saving detection time. In [18], color enhancement and MSERs based method was utilized to extract ROIs and applied SVM+HOG detector to classify the ROIs as objects or backgrounds. Many ROI extraction methods for SVM detectors have been described in the color or shape based detection parts in this review. Furthermore, methods based on SVM and HOG also play an important role in the classification of shapes, normal signs and occluded signs [19].

Ensemble learning

Viola and Jones' AdaBoost and cascade based detection structure (VJ) [17] has been proved very efficient in some object detection problems, such as face detection, car detection, license plate detection, etc. This structure has also been successfully applied in different TSD applications.Combined with some types of rectangular features, an AdaBoost based learning method and a cascade structure, the VJ structure can select features with the AdaBoost method for object expression and then detect objects in a cascade process.

The selection of features is crucial for AdaBoost based TSD detectors. The Haar-like feature [16] is the most popular feature used in different detection problems. The Haar-like feature can express the gray level difference of traffic signs. [17] Considering that Haar-like features have connected dipoles, Baróet al.proposed the dissociated dipoles feature, which is a more general rectangular feature. Using uncon- nected two dipoles, the dissociated dipoles feature can pro- duce more features to express traffic signs. Multi-Block Local

Binary Pattern (MB-LBP) feature [18] is another popular used rectangular features. Liuet al. [19] designed multi-block normalization LBP (MN-LBP) features to express different types of features. The designed MN-LBP feature can be trained to find the common features of different types of traffic signs.

The structures of the Haar-like features, dissociated dipoles, MN-LBP, ICF and ACF are shown above figures

The common AdaBoost based training methods include Real AdaBoost, Gentle AdaBoost, Discrete AdaBoost and other derived Boosting methods. These AdaBoost training methods can select powerful features as weak classifiers, which can form a strong classifier for object detection.Though AdaBoost based detection is very fast, scanning a high-resolution image is still time-consuming.

Finally, here is a comparison of the above 4 kind of method(Figure 9 [20]):

2) The Novel Deep Learning-base Method:

	Category	Paper	Year	Method	Detected colors
Color Based Detection Methods	RGB based thresholding	[2]	2010	Normalized RGB thresholding	Red, blue, yellow
		[30]	2010	Color Enhancement	Red, blue, yellow
		[31]	2015	Color Enhancement	Red, blue, yellow
	Hue and saturation thresholding	[2]	2010	Hue and saturation thresholding	Red, blue, yellow
		[33]	2004	LUTs based HS thresholding	Red, blue, yellow
	Thresholding on other spaces	[2]	2010	Ohta thresholding	Red, blue, yellow
		[34]	2015	Lab thresholding	Red, blue, yellow, green
	Chromatic/Achromatic Decomposition	[2]	2010	RGB, HIS, Ohta decomposition	white
		[34]	2015	RGB based achromatic segment	white
	Pixel classification	[2]	2010	SVM classification	Red, blue, yellow
		[36]	2012	Probabilistic neural networks	Red, blue, yellow

Fig. 4. Color Based Detection Method.

	Category	Paper	Year	Method	Detected shapes
Shape Based Detection Methods	Shape detection	[38]	2015	Hough	Circle and triangle
		[39]	2008	Radial symmetry transform	Circle
		[86]	2004	Radial symmetry transform	Polygons
	Shape analysis and matching	[41]	2003	Complex shape models	Circle, polygons
		[42]	2008	Shape decomposition	Circle, square, triangle
	Fourier transformation	[26]	2011	Fourier descriptors	Circle, square, triangle
		[43]	2008	Fast Fourier Transformation	Circle, square, triangle
	Key points detection	[45]	2014	SIFT	Circle, square, triangle, octagon
		[15]	2014	Harris corner	Circle, triangle
		[46]	2014	Interest points clustering	Different shapes

Fig. 5. Shape Based Detection Method.

The Convolutional Neural Network (CNN) based detection methods learn features through convolutional networks. In recent years, with the development of deep learning, many different deep neural network structures have appeared and made breakthroughs in different detection areas. The use of CNN for the TSD problem started in [21] and [22]. These works use a CNN classifier to classify objects from backgrounds and need ROIs extraction methods to get candidates. Nowadays, there are some networks that have fast performance, such as You Only Look Once (YOLO) net. Zhang et al. [23] utilized YOLOv2 to design their real-time traffic sign detection method. Liu et al. [24] used a YOLO CNN to classify traffic signs and MSRCR image augmentation during pre-processing. In the improvement phase, they used YOLOv5 to automate traffic sign categorization and improved training methods and network architecture. It achieved a 99.8% accuracy rate on the GTSRB dataset and 98.4% precision on the CCTSDB. Wang et al. [25] separated the proposed deep ensemble learning algorithm into two methods. First, after the traffic scene process, the algorithm detects the traffic sign as two categories with the YOLOv5s network. Then, it processes the traffic sign to recognize the traffic sign into seven classes using the MobileNet network. The result of the proposed algorithm showed high performance, with 95.83% accuracy, 87.34% true prediction, and 191.3 milliseconds (ms) of inference time. Sharma and Kumar's study [26] provides YOLOv8 for traffic signal recognition in the advanced version that operates in a real-time environment for road safety improvement. Intensively and widely tested and trained using a complex set of data, the YOLOv8 model gained a notable boost over its predecessors in key performance metrics such as precision, recall, and F1-

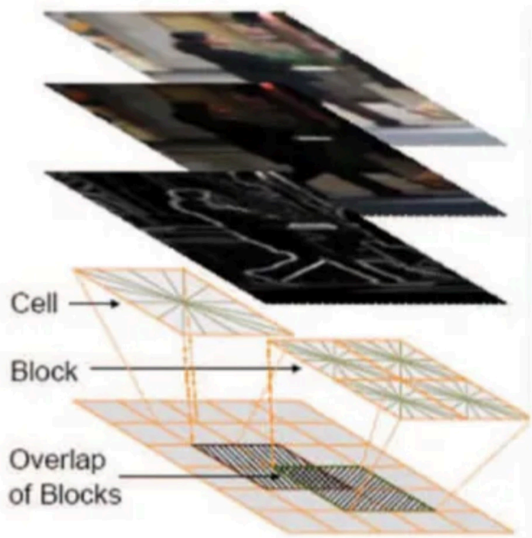

Fig. 6. HOG descriptor.

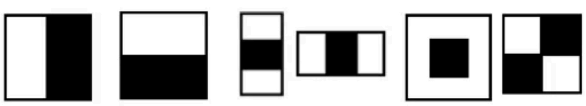

Fig. 7. Haar Like Feather.

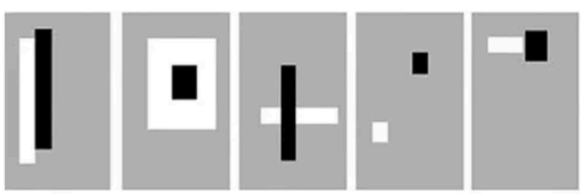

Fig. 8. Dissociated Dipoles.

score. The system performed extremely well, with the F1 score being 90.18%, recall of 89.5%, natural and language translation error rates of 5%, and precision of 91%, with only 2% errors.

3) Comparasion:

Here is a comparison of the performance of the above method, which expressed by different dataset. (Figure 10 [20]):

D. Reproduction of Current Paper

1) Reproduction and Analysis of [27]:

We have selected this paper for reproduction because it introduces a versatile plug-and-play module in the domain of recognition tasks. The underlying model is a CNN, a well-known architecture for image recognition, and our target model, YOLO, is a specialized form of CNN. Our intention is to integrate this module into our final model.

One of the compelling reasons for choosing this paper is its significant experiments on optimizers, particularly the relationship between training speed and the performance of the final model. These insights could potentially help us reduce training time and achieve superior performance.

The core of the paper revolves around STPs (**Spatial Transformer Networks**):

- A modular component designed to dynamically manage geometric transformations.
- Consists of three key components: Localization Network, Grid Generator, and Sampler.
- Enhances robustness by adjusting inputs to facilitate better feature extraction.
- In summary, this module is well-suited for any CNN to mitigate the effects of dynamically distorted figures. STNs are not only beneficial for traffic sign recognition but also applicable to other tasks demanding geometric invariance.

Based on our optimizer experiment outcomes, we concur with the paper's findings:

- SGD necessitates meticulous learning rate tuning and shows slower convergence.
- Adam demonstrated the quickest convergence and delivered the best overall performance.
- RMSprop outperformed SGD in certain situations but was marginally less effective than Adam.
- In conclusion, Adam's adaptability makes it an optimal choice for irregular datasets, while SGD remains a strong contender for large-scale datasets with proper tuning.

2) Reproduction and Analysis of [28]:

We have selected this paper for reproduction because it introduces significant improvements to the YOLOv2 algorithm, specifically tailored for real-time Chinese traffic sign detection. YOLOv2 is a powerful and widely used architecture in object detection, and its modifications in this paper offer practical advancements in handling the diverse and challenging conditions of traffic sign detection. Our goal is to integrate these optimizations into our traffic sign recognition YOLOv5 model to enhance detection performance and real-time capabilities.

A major reason for choosing this paper is its focus on optimizing YOLOv2 for real-time applications, particularly for Chinese traffic signs, which involve unique challenges such as varying sign sizes, lighting conditions, and sign types. The proposed modifications allow the model to maintain high detection accuracy while operating efficiently in real-time scenarios, which is crucial for autonomous driving and intelligent transportation systems.

The core contributions of the paper are the following key improvements to YOLOv2:

- **Network Architecture Modification:**
  - The architecture of YOLOv2 is adapted to improve feature extraction, especially for small or obscured traffic signs. This modification boosts

Paper	Year	Database Size	Feature Extractor	Classifier Used	Recognition Rate (in %)	Remarks
Benallal et al. [5]	2003		RGB colour segmentation			
Berkaya et al. [13]	2015		Proposed algorithm for circular signs	SVM	97.04	Novel idea
Ohara et al. [16]	2002	3 Signs	1NN each for matching shape and colour		>95	No proper database Not applicable in real time
Romadi et al. [20]	2014		Circle by Hough Triangle by Ramer-Douglus-Rucar	SURF & FLAN	85	
Bahlmann et al. [21]	2005		Traffic sign detection	LDA		Also uses motion information.
Ruta et al. [23]	2009	48 signs Polish,Japanese	Equiangular polygons	Gaussian	85.3	Novel idea

Fig. 9. Review of Techniques that Use Shape and Color for Detection.

Dataset	Methods	Prohibitive (AUC)	Danger (AUC)	Mandatory (AUC)	Time (s)
GTSDB	HOG+LDA [6]	70.33%	35.94%	12.01%	N/A
	Hough-like [6]	26.09%	30.41%	12.86%	N/A
	Viola-Jones [6]	90.81%	46.26%	44.87%	N/A
	HOG+LDA+SVM [89]	100%	99.91%	100%	3.533
	ChnFtrs [25]	100%	100%	96.98%	N/A
	HOG+SVM [67]	99.98%	98.72%	95.76%	3.032
	SVM+Shape [68]	100%	98.85%	92.00%	0.4-1
	SVM+CNN [69]	N/A	99.78%	97.62%	12-32
	SFC-tree [88]	100%	99.20%	98.57%	0.192 (3.19 GHz CPU)
	CNN [E-53]	99.89%	99.93%	99.16%	0.162 (Titan X GPU)
	ACF+SPC+LBP+AdaBoost [58]	100%	98.00%	97.57%	N/A
	AdaBoost+SVR [59]	100%	100%	99.87%	N/A
	AdaBoost+CNN+SVM [73]	99.45%	98.33%	96.50%	N/A
BTSD	ChnFtrs [25]	94.44%	97.40%	97.96%	1~3 (Intel Core i7 870 CPU, GTX 470 GPU
	AdaBoost+SVR [59]	93.45%	99.88%	97.78%	0.05~0.5 (Intel Core-i7 4770 CPU)
	AN+FRPN [72]	AP(%): 50.82%(Small), 88.05%(med), 96.82%(large)			0.128 (Tesla K20 GPU)
	Faster-RCNN in [72]	AP(%): 43.93%(Small), 97.8%(medium), 98.31%(large)			0.165 (Tesla K20 GPU)
TT100k	Fast R-CNN in [10]	Recall: 56%; Accuracy: 50% Curves can be found in [10]			N/A
	Multi-class Network [10]	Recall: 91%; Accuracy: 88% Curves can be found in [10]			N/A
	AN+FRPN [72]	AP(%): 49.81%(Small), 86.9%(med), 96.05%(large)			0.128 (Tesla K20 GPU)
	Faster-RCNN in [72]	AP(%): 31.22%(Small), 77.17%(med), 94.05%(large)			0.165 (Tesla K20 GPU)
LISA	ICF in [11]	87.32% (Diamond)	96.03% (Stop)	91.09% (NoTurn)	N/A
	ACF in [11]	98.98% (Diamond)	96.11% (Stop)	96.17% (NoTurn)	N/A

Fig. 10. Review of Techniques that Use Shape and Color for Detection.

the model's ability to detect hard-to-spot signs that might otherwise be missed.

- **Post-Processing Optimization:**
  - Post-processing techniques are improved to better handle bounding box predictions, reducing false positives and enhancing the precision of detected traffic signs.
- **Custom Traffic Sign Dataset:**
  - The paper uses a customized dataset containing a diverse set of Chinese traffic signs, which helps the model generalize better to the variety of signs encountered in the real world.

Based on our reproduction experiments, we observe similar results to the paper's findings:

- **Real-Time Performance:**
  - The model retains YOLOv2's real-time performance while achieving a noticeable improvement in detection accuracy, especially for small and distant traffic signs.
- **Robustness to Different Conditions:**
  - The preprocessing and multi-scale detection enhancements enable the model to perform well in diverse environments, including varying lighting and weather conditions, which is critical for real-time applications.

III. Contributions of This Study

We have made significant progress in traffic sign detection and recognition through the following contributions:

A. Applying Transfer Learning

Utilized transfer learning techniques to fine-tune variations of YOLOv5 models on the TT100K dataset, enabling better adaptation to traffic sign detection tasks with limited training data.

1) YOLOv5 introduction: YOLO is a state-of-the-art deep learning model known for its speed and accuracy in object detection tasks. Unlike other object detection algorithms, YOLO frames the detection problem as a single regression problem, which enables it to process images in real-time.

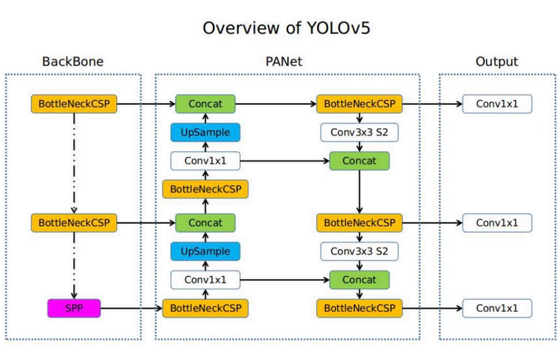

Fig. 11. YOLOv5s architecture

YOLOv5 is highly flexible and can be easily integrated into a variety of deployment environments, whether it be in a real-time application in a vehicle or on a cloud-based monitoring system. This ease of deployment ensures that our model can be tested and used in TT100K TSDR.

YOLOv5 also supports transfer learning, allowing us to fine-tune the model with TT100K.

YOLOv5s is the second smallest and fastest variant of YOLOv5. It is optimized for real-time performance on devices with limited computational resources. While it sacrifices some accuracy compared to larger models, it is highly suitable for use cases requiring high-speed processing, such as live video streaming or embedded systems. So we apply this architecture to transfer training.

2) Training Process: Here is the metrics monitored while training. We trained the different variations of YOLOv5 for 300 500 epoches:

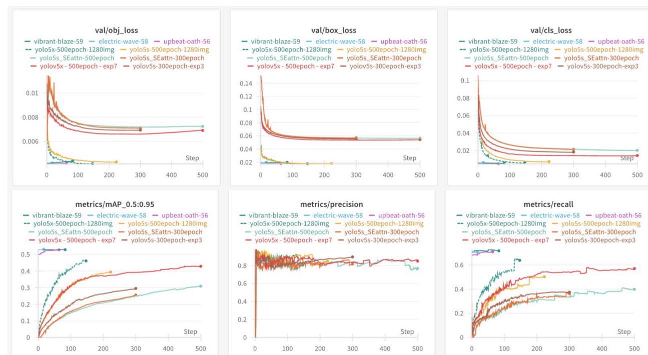

Fig. 12. YOLOv5s training

And the boxes are visualized as the following:

Fig. 13. YOLOv5s boxes

Also, the confusion matrix is nice and clean, well-established yolo model successfully handle this task:

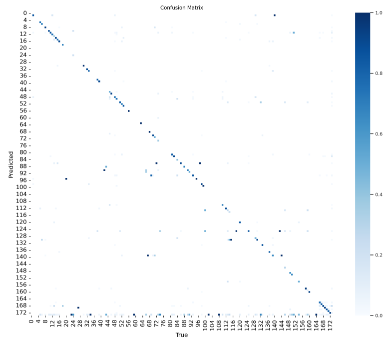

Fig. 14. YOLOv5s confusion matrix

B. Integrating Attention Module

Integrated the SE (Squeeze-and-Excitation) module (Figure 15) into YOLO to enhance feature extraction and recalibration, significantly improving detection accuracy and robustness under varying conditions.

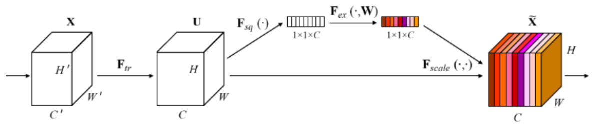

Fig. 15. SE Module Integrated into the YOLO Framework.

C. Analyzing and Comparing Performance

Conducted a comprehensive analysis of different parameter configurations, comparing their impact on model performance to identify optimal settings for accuracy and efficiency.

TABLE II
Metrics of YOLO Variations and Baseline Models (TT100K)

Variations	$mAP_{0.5:0.95}$	$mAP_{0.5}$	Precision	Recall
yolov5s$_{640}$	0.310	0.489	0.870	0.410
yolov5x$_{640}$	0.431	0.656	0.866	0.580
yolov5s-seattn$_{640}$	0.358	0.526	0.897	0.421
yolov5s$_{1280}$	0.462	0.671	0.723	0.510
yolov5x$_{1280}$	0.532	0.781	0.855	0.720
Fast R-CNN	N/A	N/A	0.50	0.56
Multi-class Network	N/A	N/A	0.88	0.91
AN+FRPN	0.4981	N/A	N/A	N/A
Faster-RCNN	0.3122	N/A	N/A	N/A

Table II presents the performance metrics of various YOLO model variations and baseline methods on the TT100K dataset. The metrics include $mAP_{0.5:0.95}$, $mAP_{0.5}$, Precision, and Recall, highlighting the comparative strengths of different approaches. YOLOv5 models demonstrate superior performance across most metrics, particularly with higher resolutions (e.g., YOLOv5x1280 achieving the highest $mAP_{0.5:0.95}$ of 0.532 and $mAP_{0.5}$ of 0.781). The integration of SE attention modules (yolov5s-seattn640) improves the precision to 0.897 while maintaining competitive recall. Among the baseline methods, AN+FRPN achieves a notable $mAP_{0.5:0.95}$ of 0.4981, whereas Multi-class Network demonstrates the highest recall of 0.91. This comparison underscores the effectiveness of modern YOLO variations in traffic sign detection tasks, particularly when enhanced with attention mechanisms and higher resolutions. We have reach SOTA level in TSDR task.

D. Deploying the Model

Successfully deployed the optimized model on both host systems and AI edge devices, supporting real-time traffic sign detection in both online and offline scenarios.

These contributions demonstrate the project's success in advancing traffic sign detection technology, achieving high accuracy, adaptability to real-world scenarios, and practical deployment feasibility.

IV. Research Effect Demonstration

We integrated an attention module into the basic YOLOv5s model. This enhancement allowed us to achieve superior performance with minimal impact on the model's parameter count and inference time compared to the original model. Notably, our modified model demonstrated significantly improved robustness in environments with poor lighting conditions.
The following data will be presented to substantiate these claims, providing a comprehensive comparison between the performance of the original YOLOv5s model and our enhanced version with the attention module.

- Refer to table II, it shows that se attention module does not affect yolov5s's performance.
- Refer to figure 11 and 12, it shows that yolov5s-se can get a higher confidence at the same image.
- Refer to figure 13 and 14, it shows that yolov5s-se can detect traffic sign at the condition of bad light environment that original yolov5s can not detect.

TABLE III
Comparision between yolov5s and yolov5s-se

Model	**Parameters**	**GFLOPs**	**Inference time**
yolov5s	7476706	17.2	1.2ms
yolov5s-se	7613578	17.3	1.2ms

V. Future Work

Building upon the advancements achieved in this project, there are several potential research directions for future work:

1) **Exploration of Advanced Attention Mechanisms:** While the SE module has demonstrated effectiveness in this project, future work could explore the integration of self-attention mechanisms

Fig. 16. yolov5s on daytime

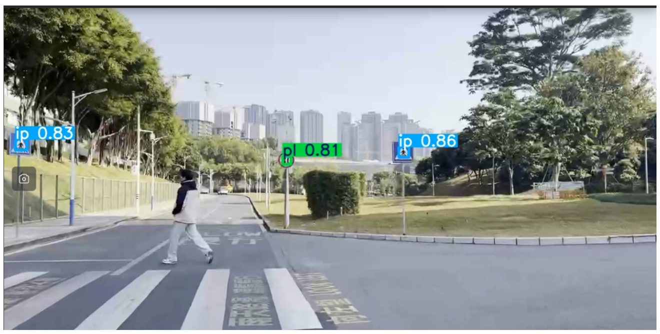

Fig. 17. yolov5s-se on daytime

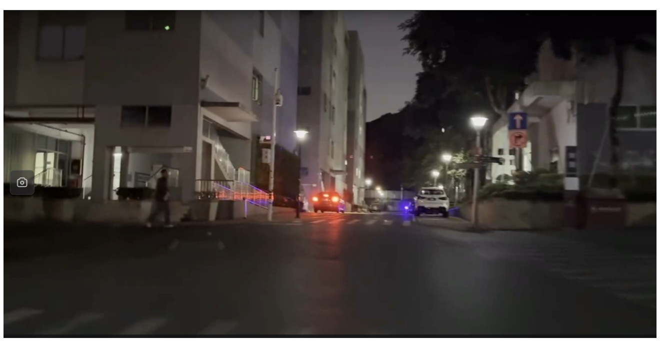

Fig. 18. yolov5s at night

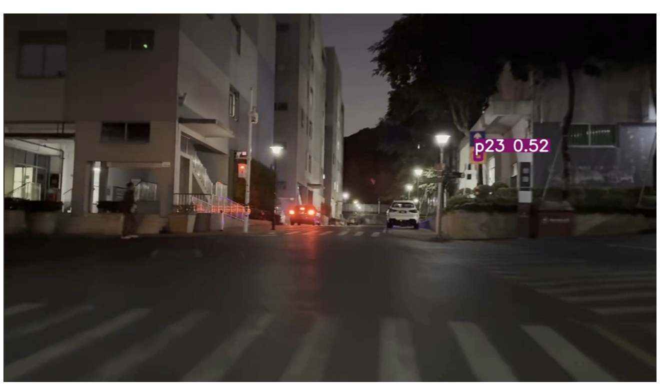

Fig. 19. yolov5s-se at night

into the traffic sign detection framework. Self-attention allows the model to capture nonlocal correlations among features, which could improve the detection accuracy in scenarios with complex spatial relationships or occlusions. Additionally, future research could investigate methods to fuse self-attention with conventional attention mechanisms, as demonstrated in recent advancements in image captioning tasks. [29] This approach may further enhance the model's ability to process diverse traffic conditions.

2) **Optimization for Resource-Constrained Devices:** To improve real-world applicability, future research can focus on optimizing the model for deployment on ultra-low-power devices, such as microcontrollers or IoT edge nodes, without compromising detection accuracy.
3) **Extension to Multi-Language Traffic Sign Recognition:** Expand the scope of the model to handle multilingual traffic signs, enabling broader applicability in international traffic systems.
4) **Real-Time Multi-Task Learning:** Investigate the feasibility of combining traffic sign detection with other real-time tasks, such as lane detection or pedestrian recognition, within a single unified framework.
5) **Data Augmentation and Synthetic Data Utilization:** Develop advanced data augmentation techniques or use synthetic data generation to address the challenges of imbalanced datasets and improve the robustness of the model.

These research directions aim to further enhance the robustness, efficiency, and applicability of traffic sign detection systems, paving the way for more intelligent and adaptive transportation technologies.

VI. Conclusion

In this project, we integrated relatively advanced techniques into the YOLO framework, enabling the model to achieve SOTA-level performance in traffic sign detection and recognition on the TT100K dataset. Key contributions include the application of transfer learning, the integration of the SE attention module, comprehensive parameter analysis, and the successful deployment of our optimized model on various platforms. These advancements have led to significant improvements in detection accuracy, robustness, and real-time performance.

Future research directions, such as exploring advanced attention mechanisms, optimizing models for resource-constrained devices, and extending functionality to multi-task learning, offer numerous opportunities for further innovation. Through these efforts, we envision continued progress in making autonomous and assisted driving systems more reliable and adaptable to diverse environments.

References

[1] World Health Organization, "Global status report on road safety 2015," 2015. [Online]. Available: https://www.who.int/publications-detail/global-status-report-on-road-safety-2015

[2] M. I. Research and Advisory, "Adas market size & share analysis - growth trends & forecasts (2024 - 2029)," June 2024, retrieved December 3, 2024. [Online]. Available: https://www.mordorintelligence.com/industry-reports/advanced-driver-assistance-systems-market

[3] ——, "Autonomous vehicle market size & share analysis - growth trends & forecasts (2024 - 2029)," July 2023, retrieved December 3, 2024. [Online]. Available: https://www.mordorintelligence.com/industry-reports/autonomous-driverless-cars-market-potential-estimation

[4] ——, "Smart transportation market size & share analysis - growth trends & forecasts (2024 - 2029)," June 2024, retrieved December 3, 2024. [Online]. Available: https://www.mordorintelligence.com/industry-reports/smart-transportation-market

[5] "Road signs dataset." [Online]. Available: https://makeml.app/datasets/road-signs

[6] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel, "Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark," in *International Joint Conference on Neural Networks*, no. 1288, 2013.

[7] R. Timofte, M. Mathias, R. Benenson, and L. Van Gool, "Traffic sign recognition - how far are we from the solution?" in *International Joint Conference on Neural Networks (IJCNN 2013)*, August 2013, * equal contributions.

[8] "Lisa traffic sign dataset," 2012, the World's largest dataset containing pictures of US traffic signs. [Online]. Available: <https://vbn.aau.dk/en/datasets/lisa-traffic-sign-dataset>

[9] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu, "Traffic-sign detection and classification in the wild," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[10] B. Sanyal, R. K. Mohapatra, and R. Dash, "Traffic sign recognition: A survey," in *2020 International Conference on Artificial Intelligence and Signal Processing (AISP)*, 2020, pp. 1–6.

[11] J. Zhang, Y. Lv, J. Tao, F. Huang, and J. Zhang, "A robust real-time anchor-free traffic sign detector with one-level feature," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 8, no. 2, pp. 1437–1451, 2024.

[12] M. Swathi and K. V. Suresh, "Automatic traffic sign detection and recognition: A review," in *International Conference on Algorithms*, 2017, pp. 1–6.

[13] S. M and S. K. V, "Automatic traffic sign detection and recognition: A review," in *International Conference on Algorithms*, 2017, pp. 1–6.

[14] C. Liu, S. Li, F. Chang, and Y. Wang, "Machine vision based traffic sign detection methods: Review, analyses and perspectives," *IEEE Access*, vol. 7, pp. 86 578–86 596, 2019.

[15] S. H. S. G. and C. M. Patil, "An approach towards efficient detection and recognition of traffic signs in videos using neural networks," in *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2016, pp. 456–459.

[16] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. CVPR*, San Diego, CA, USA, Jun. 2005, pp. 886–893.

[17] S. Salti, A. Petrelli, F. Tombari, N. Fioraio, and L. D. Stefano, "Traffic sign detection via interest region extraction," *Pattern Recognit.*, vol. 48, no. 4, pp. 1039–1049, Apr. 2015.

[18] F. Zaklouta and B. Stanciulescu, "Real-time traffic sign recognition in three stages," *Robot. Auto. Syst.*, vol. 62, no. 1, pp. 16–24, 2014.

[19] Y.-L. Hou, X. Hao, and H. Chen, "A cognitively motivated method for classification of occluded traffic signs," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 2, pp. 255–262, Feb. 2017.

[20] C. Liu, S. Li, F. Chang, and Y. Wang, "Machine vision based traffic sign detection methods: Review, analyses and perspectives," *IEEE Access*, vol. 7, pp. 86 578–86 596, 2019.

[21] Y. Wu, Y. Liu, J. Li, H. Liu, and X. Hu, "Traffic sign detection based on convolutional neural networks," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, Dallas, TX, USA, Aug. 2013, pp. 1–7.

[22] R. Qian, B. Zhang, Y. Yue, Z. Wang, and F. Coenen, "Robust chinese traffic sign detection and recognition with deep convolutional neural network," in *Proceedings of the International Conference on Natural Computation*, Zhangjiajie, China, Aug. 2015, pp. 791–796.

[23] J. Zhang, M. Huang, X. Li, and X. Jin, "A real-time chinese traffic sign detection algorithm based on modified yolov2," *Algorithms*, vol. 10, no. 4, p. 127, Nov. 2017.

[24] A. Liu, Y. Liu, and S. Kifah, "Deep convolutional neural network for enhancing traffic sign recognition developed on yolo v5," in *2024 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*, Bhubaneswar, India, 2024, pp. 1–6.

[25] L.-J. Wang, T. Suwattanapunkul, J. Thalauy, and P. Jansengrat, "Research on traffic sign detection and recognition system using deep ensemble learning," in *2024 6th International Conference on Computer Communication and the Internet (ICCCI)*, Tokyo, Japan, 2024, pp. 61–66.

[26] J. Sharma, D. Kumar, and A. Malhotra, "Advanced traffic signal recognition using real-time detection using yolo v8," in *2024 2nd World Conference on Communication & Computing (WCONF)*, Raipur, India, 2024, pp. 1–4.

[27] . Arcos-García, J. A. Alvarez-Garcia, and L. M. Soria-Morillo, "Deep neural network for traffic sign recognition systems: An analysis of spatial transformers and stochastic optimisation methods," *Neural Networks*, vol. 99, pp. 158–165, 2018.

[28] J. Zhang, M. Huang, X. Jin, and X. Li, "A real-time chinese traffic sign detection algorithm based on modified yolov2," *Algorithms*, vol. 10, no. 4, p. 127, 2017.

[29] Z. Li, Y. Li, and H. Lu, "Improve image captioning by self-attention," in *Neural Information Processing*, T. Gedeon, K. W. Wong, and M. Lee, Eds. Cham: Springer International Publishing, 2019, pp. 91–98.