Performance Enhancement Technique for Traffic Sign Recognition using Combining Label Smoothing and Focal Loss

Hyunseo Jeong and Eunkyung Kim

Department of Artificial Intelligence Software
Hanbat National University
Daejeon, Korea
hyunseo@edu.hanbat.ac.kr, ekim@hanbat.ac.kr

***Abstract*—In this paper, we propose a novel loss function to enhance the performance of a YOLOv8 based object detection model for traffic sign recognition under diverse road conditions. The proposed method aims to improve detection accuracy in challenging environments such as nighttime and adverse weather. To optimize performance, we experimentally evaluate three loss functions, i.e., Binary Cross Entropy, Focal Loss, and Label Smoothing Cross Entropy under identical training conditions. The results show that Focal Loss effectively improves Recall by focusing on hard samples and minority classes, while Label Smoothing reduces overfitting and improves Precision. By combing the strengths of both, the proposed loss function achieves robust overall performance with optimal hyperparameters. Visual inspection using the test dataset further confirms that the model reliably detects and classifies traffic signs across a variety of environmental conditions.**

Keywords—YOLOv8, Traffic Sign Recognition, Object Detection, Focal Loss, Label Smoothing Cross Entropy, Combined Loss Function, Autonomous Driving

I. Introduction

Traffic signs play a crucial role in ensuring road safety by providing drivers with essential information such as speed limits, pedestrian crossings, and warning zones. With the advancement of autonomous driving technologies and Advanced Driver Assistance Systems (ADAS), the demand for Traffic Sign Recognition (TSR) systems that offer both accuracy and real-time performance has been steadily increasing [1].

Early Traffic Sign Recognition approaches primarily relied on traditional image processing techniques and handcrafted feature extraction. However, these methods showed limitations in real-world environments due to their vulnerability to changes in lighting, partial occlusions of signs, and complex backgrounds [2]. To address these issues, recent research [3] has actively explored object detection technologies based on deep learning, particularly Convolutional Neural Networks (CNN).

Among the many object detection frameworks [4, 5], You Only Look Once (YOLO) has emerged as a leading algorithm in real-time applications. YOLO performs detection in a single forward pass of the network, allowing it to balance both speed and accuracy effectively [6].

In this study, we examine the architectural characteristics and practical applicability of YOLO based models YOLOv5, YOLOv7, YOLOv6, and YOLOv8. Based on this analysis, we select YOLOv8 as the most suitable model in terms of accuracy and ease of use, and subsequently design and implement a traffic sign recognition system using it. Furthermore, considering real-world driving environments, we perform image augmentation using Albumentations, a Python based open source library, to simulate nighttime and adverse weather conditions, which contributes to enhancing the model's generalization capability [7].

To further enhance the performance of the model, we conduct experiments using various loss functions, including Binary Cross Entropy [8], Focal Loss [9], and Label Smoothing Cross Entropy Loss [10], and analyze how each loss function affects key performance metrics such as precision and recall. The experimental results show that while each loss function demonstrates strengths in specific metrics, it is difficult to achieve optimal performance across all metrics simultaneously.

Therefore, we propose a novel approach that combines the high precision of Label Smoothing with the high recall of Focal Loss to improve the overall performance of the traffic sign recognition model.

Through this series of experiments, we validate the effectiveness of the YOLOv8 based traffic sign recognition model and demonstrate its potential for building a robust object detection system capable of maintaining high recognition performance under diverse road conditions.

The rest of this paper is organized as follows: We begin with the research methods in Section II. Experiment results are presented in Section III followed by the conclusion in Section IV.

II. Research Methods

In this section, we explain the research methodology for developing a robust traffic sign recognition model, including i) the class imbalance present in the dataset, ii) image augmentation techniques simulating real-world road conditions, which are applied exclusively to the training data to improve generalization performance, iii) comparisons of various YOLO based models with YOLOv8 selected as the final model due to its structural efficiency and training stability, and iv) Binary Cross Entropy, Focal Loss, and Label Smoothing Cross Entropy, which are the loss functions to enhance the performance of traffic sign recognition.

2.1 Dataset

In this study, we utilize a publicly available traffic sign dataset from Kaggle [11]. The dataset consists of 877 original images and includes four classes: Traffic Light, Stop, Speed Limit, and Crosswalk. The number of objects per class is distributed (see Fig.1) as follows: Speed Limit (783), Traffic Light (170), Crosswalk (200), and Stop (91). Note that class distribution shows a significant imbalance, with the Speed Limit class accounting for a disproportionately large portion of the dataset. Such class imbalance may increase the risk of overfitting to certain classes during training, highlighting the need for appropriate strategies to mitigate this issue.

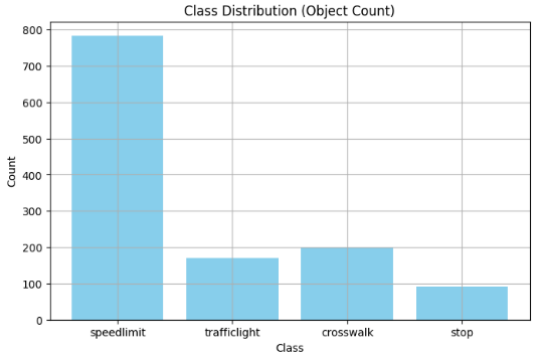


Fig. 1. Visualization of Class Imbalance in Dataset.

Each image is accompanied by an annotation file in XML format, which provides precise object localization information for object detection tasks. To prepare the data for training with the YOLOv8 model, we convert the annotation files from XML to the TXT format required by YOLO. We then randomly shuffle the entire dataset and split it into three subsets training (80%), validation (10%), and testing (10%).

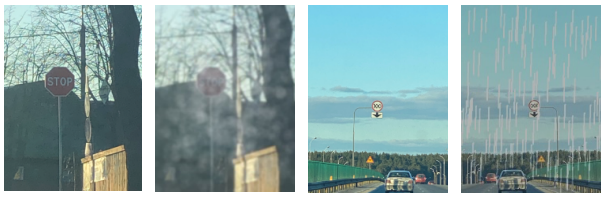
The training dataset consists of a total of 1,406 images, including both original and augmented images. For the validation and test datasets, we use only original images to ensure a fair evaluation of the model's generalization performance. This approach aims to train a model capable of responding robustly to various real-world road conditions.

2.2 Image Augmentation

To enhance traffic sign recognition performance under diverse environmental conditions, we apply image augmentation exclusively to the training dataset. These augmentation techniques increase the variability of the training data, thereby improving the model's generalization capability and mitigating overfitting.

Using the Albumentations library, we employ various augmentation methods, including random brightness and contrast adjustment (*RandomBrightnessContrast*), motion blur (*MotionBlur*) to simulate blurring effects, fog (*RandomFog*), rain (*RandomRain*), and grayscale conversion (*ToGray*) to mimic nighttime conditions [7]. These augmentations simulate a range of real-world driving scenarios such as daytime, nighttime, rainy, and foggy weather conditions.

Fig. 2 presents examples of original images from the training dataset alongside their corresponding augmented versions. For validation and testing, we use only original images without augmentation to evaluate the actual performance of the model. This set up enables a comparative analysis of the impact of augmented training data on the model's generalization ability.



(a) (b) (c) (d)

Fig. 2. Examples of original images with (a), (c) and respective augmented images with (b), (d).

2.3 Model Selection

The primary goal of this study is to enhance recognition accuracy and generalization capability in diverse road environments by combining an anchor-free object detection architecture with an optimized loss function. To achieve this, we review and compare four YOLO based models YOLOv5, YOLOv7, YOLOv6, and YOLOv8 in terms of architectural characteristics, experimental feasibility, and training stability.

YOLOv5 is one of the most widely used object detection models, known for its fast inference speed and stable accuracy. However, it retains the conventional anchor-based structure, which limits its ability to flexibly detect objects of varying sizes. In real-world traffic scenarios involving irregular lighting and complex backgrounds, the model's sensitivity to anchor box configurations can hinder its performance, particularly for small traffic signs. Therefore, we exclude YOLOv5 from selection.

YOLOv7 integrates several high performance modules and achieves impressive accuracy. Nonetheless, its complex architecture and high computational cost make it unsuitable for our experimental environment. Given the relatively small dataset used in this study, there is also a risk of overfitting, which further reduces its applicability.

YOLOv6 partially adopts anchor-free characteristics and is initially considered a viable candidate. However, during experimentation, it frequently encounters training instabilities and errors. Despite various hyperparameter adjustments, the model fails to demonstrate consistent performance and reliable convergence, which restricts its use in this study.

YOLOv8, on the other hand, adopts a fully anchor-free architecture that directly predicts object centers, resulting in more compact and accurate detections. This structure enables the model to better adapt to objects of various sizes and positions. Additionally, YOLOv8 offers high usability and integrates seamlessly into our experimental environment. It also demonstrates stable convergence and strong detection performance during training. For these reasons, we select YOLOv8 as the most suitable model for our objectives and use it as the foundation for developing the proposed traffic sign recognition system.

2.4 Loss Functions

To improve the performance of the traffic sign recognition model, we explore and apply various loss functions. In real-world traffic scenarios, class imbalance frequently occurs, and certain samples may be difficult to predict due to varying lighting or weather conditions. To address these challenges and enhance the generalization performance of the model, we

experiment with three primary loss functions: Binary Cross Entropy (BCE) [8], Focal Loss [9], and Label Smoothing Cross Entropy Loss [10].

A. Binary Cross Entropy

Binary Cross Entropy is one of the most widely used loss functions for binary classification tasks and serves as the default loss function in YOLOv8. It computes the loss based on the logarithmic difference between the predicted probability and the ground truth label. The loss decreases as the predicted probability approaches the true label. The BCE loss is defined as follows [8]:

$$\mathcal{L}_{BCE} = -[y \cdot \log(\hat{y}) + (1-y) \cdot log(1-\hat{y})], \quad (1)$$

where $y \in \{0,1\}$ and $\hat{y} \in \{0,1\}$ denote the ground truth label and the predicted probability after applying the sigmoid function, respectively. While BCE encourages the model to make confident predictions, it may not perform well in cases involving class imbalance or hard to classify samples as it lacks mechanisms to address sample difficult or frequency.

B. Focal Loss

Focal Loss is designed to address class imbalance and the dominance of easily classified samples by assigning greater weight to hard to classify examples. It is particularly effective when certain classes are significantly underrepresented, which can lead to their being overlooked during training. The Focal Loss is defined as follows [9]:

$$\mathcal{L}_{Focal} = -\alpha_t(1-p_t)^{\gamma} log(p_t), \quad (2)$$

where α_t, p_t, and γ denote a class-specific weighting factor, the predicted probability for the true class, and the focusing parameter, respectively. By reducing the loss contribution from easy examples, Focal Loss encourages the model to focus on more challenging samples. This is particularly beneficial under difficult conditions such as nighttime or adverse weather, where traffic signs may be harder to detect.

C. Label Smoothing Cross Entropy Loss

Label Smoothing is a regularization technique that prevents the model from becoming overconfident in its predictions, thereby improving generalization performance. Instead of assigning a probability of 1 to the correct class and 0 to all others (as in traditional one-hot encoding), label smoothing softens the target distribution as follows:

$$\tilde{y}_k = y_k(1-\alpha) + \frac{\alpha}{K}, \quad (3)$$

where $\tilde{y}_k$, $\alpha \in \{0,1\}$, and K denote the smoothed target probability, the smoothing factor, and K is the number of classes, respectively. The corresponding loss function is then defined as [10]:

$$\mathcal{L}_{LS} = -\sum_{i=1}^{K} \tilde{y}_i \log(p_i). \quad (4)$$

Here, $\tilde{y}_i$ and p_i denote the smoothed ground truth distribution and the predicted probability after applying the softmax function, respectively. This approach encourages the model to allocate some probability mass to non-target classes, which reduces overfitting and enhances generalization, especially in noisy or imbalanced datasets.

III. EXPERIMENTS

3.1 EXPERIMENTAL SETUP

In this study, we train a traffic sign recognition model based on the pretrained YOLOv8s model. The training is conducted in a GPU environment using the Ultralytics YOLOv8 library built on the PyTorch framework. The main training configurations are as follows: 50 epochs, a batch size of 8, an input image size of 416×416, and an early stopping patience of 30.

To compare model performance, we apply three different loss functions Binary Cross Entropy, Focal Loss, and Label Smoothing Cross Entropy under identical training conditions. This setup enables a quantitative analysis of the impact of each loss function on model performance.

To ensure balanced recognition performance across diverse conditions and enhance generalization under real-world environments, we use both original and augmented images in the training dataset. For validation and testing, we use only original images without augmentation to ensure a fair evaluation of the model's generalization capability.

Model performance is evaluated using widely adopted object detection metrics: Precision, Recall, mAP@50, and mAP@50-95.

3.2 COMPARISON OF LOSS FUNCTIONS

To evaluate the impact of different loss functions on the performance of the traffic sign recognition model, we conduct a comparative experiment using three loss functions: Binary Cross Entropy (BCE)**,** Focal Loss, and Label Smoothing Cross Entropy under identical training conditions. Table I summarizes the results of this experiment.

TABLE I. PERFORMANCE COMPARISON BY LOSS FUNCTION

Loss Function	*Precision*	*Recall*	*mAP@50*	*mAP@50-95*
Binary Cross Entropy	0.874	0.854	0.883	0.778
Focal Loss ($\boldsymbol{\alpha}$=0.25, $\boldsymbol{\gamma}$=2)	0.864	**0.870**	0.852	0.744
Label Smoothing ($\tilde{\boldsymbol{y}}_{\boldsymbol{i}}$=0.1)	**0.916**	0.785	0.839	0.729

Binary Cross Entropy demonstrates overall balanced performance, with moderate Precision (0.874) and Recall (0.854), serving as a stable baseline across all metrics. Focal Loss achieves the highest Recall (0.870), indicating superior detection performance for hard to learn samples or underrepresented classes, which is particularly effective in complex road scenarios such as nighttime or adverse weather conditions. However, this comes with a slight decrease in Precision (0.864) compared to Binary Cross Entropy. Label Smoothing, on the other hand, achieves the highest Precision (0.916), effectively reducing overfitting by mitigating overconfidence and enhancing generalization. This, however,

results in a notably lower Recall (0.785), suggesting some weak signals might be missed.

No single loss function outperforms the others across all evaluation metrics. While Label Smoothing produces the highest Precision, its lower Recall compared to Focal Loss and Binary Cross Entropy highlights a tradeoff between Precision and Recall. Conversely, Focal Loss's highest Recall but relatively lower Precision implies an increase in false positives.

Based on these findings, we propose a method that combines the high Precision of Label Smoothing with the high Recall of Focal Loss. In the next experiment, we design a loss function to incorporate the strengths of both approaches and empirically validate its performance.

3.3 Combined Loss: Label Smoothing + Focal Loss

In the experiments in Section 3.2, Label Smoothing Cross Entropy and Focal Loss each demonstrate strengths in terms of Precision and Recall, respectively. Unfortunately, any loss function dose not achieve superior performance across all evaluation metrics simultaneously. Therefore, we propose a novel loss function that combines the advantages of both Label Smoothing and Focal Loss.

The proposed loss function integrates Label Smoothing Cross Entropy into the Focal Loss framework, aiming to emphasize difficult samples while mitigating overconfidence and improving the model's generalization capability. Given the class imbalance present in the dataset, which often leads to reduced prediction accuracy for minority classes we combine the strengths of both approaches. The combined loss is defined as follows:

$$L_{combined} = -\sum_{i=1}^{K} \tilde{y}_i (1 - p_i)^{\gamma} \log(p_i), \qquad (5)$$

where $\tilde{y}_i$, p_i, and γ denote the smoothed label, the predicted probability after applying the softmax function, and the focusing parameter inherited from Focal Loss, respectively.

To identify the optimal combination of the two core hyperparameters the smoothing factor $\tilde{y}_i$ and the focusing parameter γ we conduct a series of experiments. The results are summarized in TABLE II.

TABLE II. Performance by Smoothing and Gamma Parameters

Smoothing	*gamma*	*Precision*	*Recall*	*mAP@50*	*mAP@50-95*
0.05	1.8	0.880	0.852	0.845	0.724
0.03	2.0	0.942	0.871	0.939	0.775
0.04	2.0	0.929	0.896	0.927	0.78
0.05	**2.0**	**0.966**	**0.885**	**0.92**	**0.789**
0.06	2.0	0.898	0.886	0.877	0.759
0.1	2.0	0.920	0.874	0.887	0.769
0.1	2.1	0.944	0.894	0.914	0.773
0.07	2.1	0.882	0.841	0.832	0.756

The combination of $\tilde{y}_i = 0.04$ and $\gamma = 2.0$ yields high Recall and mAP@50, but relatively low Precision. In contrast, the combination of $\tilde{y}_i = 0.05$ and $\gamma = 2.0$ achieves slightly better performance in Precision and mAP@50-95, offering a more balanced overall performance. Therefore, we select this configuration as the baseline for subsequent experiments.

Based on this baseline setting, we perform further fine-tuning of the smoothing factor $\tilde{y}_i$. The results are presented in TABLE III.

TABLE III. Fine-Tuning Around Best Smoothing Value

Smoothing	*gamma*	*Precision*	*Recall*	*mAP@50*	*mAP@50-95*
0.047	2.0	0.94	0.874	0.901	0.769
0.05	**2.0**	**0.966**	**0.885**	**0.92**	**0.789**
0.053	2.0	0.949	0.895	0.913	0.777
0.055	2.0	0.927	0.883	0.92	0.782

This fine-tuning experiment confirms that the configuration with $\tilde{y}_i = 0.05$ and $\gamma = 2.0$ achieves the best overall results in terms of Precision, mAP@50, and mAP@50-95, while also maintaining a high Recall score. Thus, we validate this setting as the most suitable in terms of balanced performance.

Following the loss function optimization, we perform additional post processing to further enhance detection performance. Specifically, we adjust the confidence threshold and Intersection over Union (IoU) threshold. The results are summarized in TABLE IV.

TABLE IV. Confidence and IoU Thresholds Tuning

Smoothing	*gamma*	*Precision*	*Recall*	*mAP@50*	*mAP@50-95*	*Conf*	*IoU*
0.05	2.0	0.966	0.885	0.92	0.789		
0.05	2.0	0.966	0.885	0.921	0.79		0.6
0.05	**2.0**	**0.969**	**0.885**	**0.939**	**0.822**	**0.25**	

Note: In this table, blank entries under the 'Conf' and 'IoU' columns indicate that the corresponding thresholds are not applied at all.

The final configuration using $\tilde{y}_i = 0.05$, $\gamma = 2.0$, and confidence threshold of 0.25 achieves the highest performance across all evaluation metrics: Precision (0.969), Recall (0.885), mAP@50 (0.939), and mAP@50-95 (0.822). These results indicate that the model is effectively optimized to filter out uncertain predictions while maintaining sensitivity to critical objects.

3.4 Vizualization of Detection Results

In addition to quantitative metrics, we visually assess the detection performance of the model using the test dataset. Fig. 3 presents examples in which the model successfully recognizes and classifies traffic signs from randomly selected test images.

Despite various challenging conditions such as lighting changes, background complexity, and object size variations, the model accurately recognizes signs including Speed Limit, Traffic Light, Crosswalk, and Stop.

These results visually confirm that the proposed loss function and optimization strategies are effective, even under real-world road conditions.

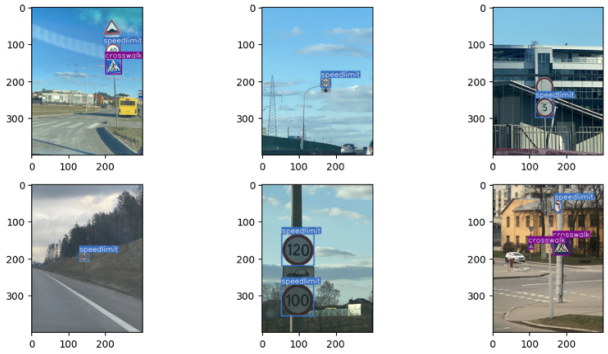


Fig. 3. Visualization of Traffic Sign Recognition.

IV. Conclusion

In this paper, we propose a YOLOv8 based traffic sign detection system designed to robustly recognize traffic signs under diverse road conditions. Through a comparative analysis of the architectural characteristics and training stability of YOLO based models, we select YOLOv8 an anchor-free architecture as the final model as it best aligns with the objectives of this study.

We experimentally apply various loss functions, including Binary Cross Entropy, Focal Loss, and Label Smoothing Cross Entropy, to address challenges such as illumination changes, adverse weather conditions, and class imbalance frequently encountered in real-world environments. While each loss function demonstrates strengths in specific metrics such as Precision or Recall, none achieves consistently balanced performance across all evaluation indicators.

To overcome this limitation, we design a novel combined loss function that integrates the overfitting mitigation effect of Label Smoothing with the focus on hard samples emphasized by Focal Loss, making it particularly robust against class imbalance. Experimental results show that with a smoothing factor of $\tilde{y}_i = 0.05$ and focusing parameter $\gamma = 2.0$, the model achieves the best overall performance: Precision of 0.969, Recall of 0.885, mAP@50 of 0.939, and mAP@50-95 of 0.822. Additionally, adjusting the confidence threshold to 0.25 further improves the accuracy of post processing.

Visualization of detection results using randomly selected images from the test dataset confirms that the model accurately recognizes and classifies traffic signs under various conditions including daytime, nighttime, rain, and fog. These findings verify that the proposed combined loss function enables effective and robust detection in real-world scenarios.

This study presents a loss function combination strategy to enhance the performance of YOLOv8 based traffic sign detection and demonstrates that high recognition accuracy can be achieved even under class imbalance and complex environmental conditions.

For future work, we plan to improve model robustness through additional training and evaluation on large-scale real-world road datasets and to explore the application of real-time, lightweight detection models for deployment in autonomous driving systems and smart city infrastructure.

Acknowledgment

This work was partly supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) - ITRC (Information Technology Research Center) grant funded by the Korea government (MSIT) (IITP-2025-RS-2024-00437886, 40%), by Basic Science Research Program through the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (RS-2025-25408839, 40%), and by the National Program for Excellence in SW), supervised by the IITP in 2025 grant funded by the Korea government (MSIT) (2022-0-01068, 20%).

References

[1] Markus Mathias, Radu Timofte, Rodrigo Benenson and Luc Van Gool, “Traffic sign recognition – How far are we from the solution?,” The 2013 international joint conference on Neural networks (IJCNN), 2013.

[2] Yanzhao Zhu, and Qi Yan Wei, “Traffic sign recognition based on deep learning,” Multimedia Tools and Applications 81.13 (2022): 17779-17791.

[3] Jinfeng Cao, Bo Peng, Mingzhong Gao, Haichun Hao, Xinfang Li, and Hongwei Mou, "Object detection based on CNN and Vision-Transformer: A survey," IET Computer Vision 19.1 (2025): e70028.

[4] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” Advances in neural information processing systems 28 (2015).

[5] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg “SSD: Single shot multibox detector,” European conference on computer vision, Dec. 2016.

[6] Peiyuan Jiang, Daji Ergu, Fangyao Liu, Ying Cai, and Bo Ma, “A Review of Yolo Algorithm Developments,” Procedia Computer Science, Volume 199, 2022.

[7] Alexander Buslaev, Alex Parinov, Eugene Khvedchenya, Vladimir I. Iglovikov, and Alexandr A. Kalinin, “Albumentations: fast and flexible image augmentations,” Information 11.2 (2020): 125.

[8] Juan Terven, Diana M. Cordova-Esparza, Alfonso Ramirez-Pedraza, Edgar A. Chavez-Urbiola, and Julio A. Romero-Gonzalez, “Loss functions and metrics in deep learning,” arXiv preprint arXiv:2307.02694 (2023).

[9] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár, “Focal loss for dense object detection” Proceedings of the IEEE international conference on computer vision. 2017.

[10] Müller, Rafael, Simon Kornblith, and Geoffrey E. Hinton, “When does label smoothing help?,” Advances in neural information processing systems 32 (2019).

[11] AndrewMvd, "Road sign detection," Kaggle, [Online]. Available: https://www.kaggle.com/datasets/andrewmvd/road-sign-detection.