Grenze International Journal of Engineering and Technology, June Issue

Deep Residual Learning and U-Net Segmentation for Automated Traffic Sign Detection and Classification

Naragjun A[2] and Manju N[2]
[1-2]Department of Information Science and Engineering, JSS Science and Technology University, Mysuru, India
Email: manjun007@jssstuniv.in

***Abstract*— The manual inventory management of the traffic signs is proposed to be replaced by an automated traffic sign detection and recognition framework as per the study. The system comprises U-Net for the precise segmentation of traffic signs from complicated road surroundings and employs CNN and ResNet152V2 models for multi-class classification. To ensure robust training and generalization under different conditions, a dataset of ,000 augmented images in 52 traffic sign categories was used. The CNN model resulted in a validation accuracy of 82.1%, whereas the ResNet152V2 network, which was made deeper with the residual connections for feature extraction, achieved a higher 98.3% accuracy. The main reason for the better performance of ResNet152V2 over traditional CNN architectures in the experiments is the improvement of stability during convergence and the reduction of overfitting. Deep residual networks are one of the most effective solutions for the traffic sign recognition problem in real-time scenarios, as this work demonstrates. Furthermore, it provides a basis for their use in Advanced Driver-Assistance Systems (ADAS) and autonomous vehicle applications, thus, increasing safety and the efficient management of traffic.**

***Keywords*— Traffic sign, CNN, Resnet152V2, Segmentation.**

I. Introduction

Effectively managing the inventory of street signs that display the directives of the traffic is a task of a great magnitude and of a high importance that, at the same time, ensures the safety of the drivers and guarantees the smooth flow of the traffic. The operation is mostly done manually [1]. A vehicle-mounted camera takes images of directions, and a human operator performs manual localization and recognition offline to check the database for consistency. Unfortunately, the very rate at which such manual work goes on increases exponentially due to road stretches of thousands of kilometres. By automating this task, the reduction in the amount of manual work would be massive. It will also enable development in safety since damaged or missing traffic signs will be detected in a very short time. In the computer vision field, the most recognized problem is that of traffic sign recognition. Hence, numerous efficient detection and recognition algorithms have been proposed. Nevertheless, very few of them are specifically designed for the categories that are most related to the advanced driver-assistance systems (ADAS) and self-driving cars [2] [3]. The problem of how to detect and recognize many different kinds of traffic signs remains largely unexplored. Various previous benchmarks have addressed the issue of traffic-sign recognition and detection tasks. [4] However, several of them have solely focused on the recognition of traffic signs (TSR) while leaving out the detection of traffic signs (TSD) which is the problem of determining the exact location of a traffic sign.

Grenze ID: 01.GIJET.12.2.42_25

There are some benchmarks that not only address the issue of TSD but most of them concentrate only on a few categories of traffic signs that is those that are relevant for ADAS and autonomous vehicles applications [5]. Most of the categories, which are present in such benchmarks, look very different from each other with very little inter-category variance and, thus, they can be identified by the use of handcrafted detectors and classifiers.

II. Related Work

Machine learning (ML) and deep learning (DL) techniques have been the main factors that have brought to the attention the necessity of the automatic recognition and classification of traffic signs (TSR) in the context of ADAS and self-driving cars, which are the most recent applications of this field [6]. The first methods for traffic sign detection were based on color segmentation and shape analysis. However, the state of the art has moved towards more accurate and efficient models due to the use of deep learning in general and Convolutional Neural Networks (CNNs) in particular [7]. These models are very powerful as they automatically learn and extract key features from the images which allows them to correctly identify and label traffic signs in real-time. Occasionally, methods like CNN-SVM and CNN-MLP have been used to enhance classification robustness, while ensemble learning techniques have been employed for better generalization in different environments. They use CNN, SVM, and Random Forest to create a standard detection system for recognizing road marking that is broken. In. [8] implemented RCNN for the segmentation of road markings and then used Otsu's method for the identification of the road marking sign that is damaged. By reconstructing the entire road marking area, they removed the road marking section and calculated the decay rate by dividing the damaged part by the fully reconstructed road marking area. Some researchers have also employed YOLO for the detection of damaged road marking signs. In [9], the authors deployed YOLOv5 to localize the different types of road surface damage, which also included blurred road markings. [10] conducted an experiment to compare the detection performance of YOLOv5, YOLOv7, and Faster RCNN in detecting damaged road marking signs

III. Methodology

The investigation explores the possibility of utilizing cutting-edge Deep Learning and Machine Learning methods for building a traffic sign recognition system that is highly efficient. As part of the research, the ResNet-152V2model, and CNN techniques were used and the schematic of the method implemented is given in Figure 1.

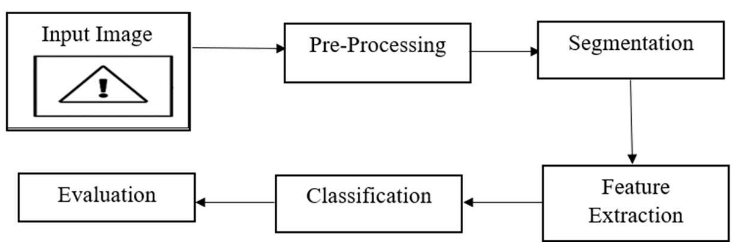


Fig. 1. Proposed Methodology

A. Dataset

This dataset originates from a traffic sign repository and has been categorically divided into 52 different classes. It contains around 12,000 labeled images that represent a variety of traffic sign categories. The proper and detailed classification facilitates a deep learning model for efficient training and testing for strong traffic sign identification. Table 1 shows dataset composition summary

TABLE I. Dataset Composition Summary

Parameter	Description
Dataset Type	Traffic Sign Image Dataset
Total Number of Images	12,000
Number of Classes	52 Traffic Sign Categories
Image Type	Labeled RGB Traffic Sign Images
Segmentation Method	U-Net Architecture
Classification Models	CNN and ResNet152V2
Input Image Size	150 × 150 Pixels

B. Data pre-processing

By employing visual components such as graphs, charts, and maps, data visualization tools facilitate the identification of patterns and anomalies in data. The dataset of traffic sign images is divided into 52 classes is shown in Figure 2.

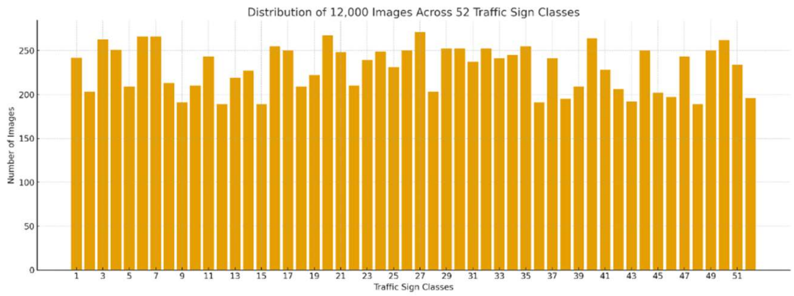


Fig. 2. Data Visualization

To enhance the ability of the classification model to generalize, the data augmentation was utilized with the traffic sign dataset to increase its size and variety. New samples were created artificially from the original images by using the mentioned techniques. This operation has, in fact, increased the dataset to about 24,000 images of 52 classes, as shown in figure 3 distribution graph. These changes in the actual environment of the signs allow the model to become resistant to overfitting and thus be capable of correctly identifying the signs regardless of the different weather conditions

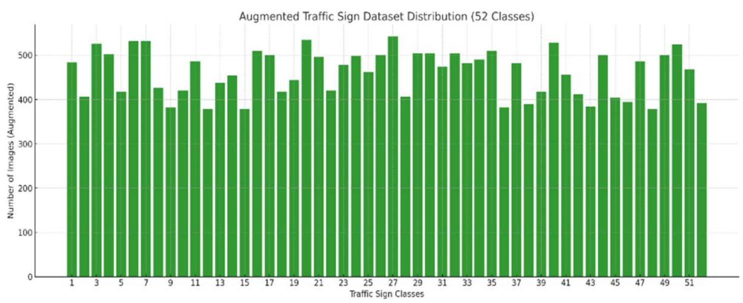


Fig. 3. Data Augmentation

C. U-Net Architecture for traffic Sign Segmentation

U-Net is one of the fully convolutional neural networks that has been developed primarily for image segmentation tasks. Such a symmetric encoder–decoder structure allows it to have great localization abilities of small objects like traffic signs even in very complex road scenes. In the case of traffic sign classification problems, first, with the help of U-Net, the traffic signs are separated from the background, after which the classifier performs the classification or recognition can be done straight away by using the segmented masks. The encoder slowly reduces the spatial resolution of the original image but simultaneously it captures high-level semantic features. Any block in the encoder consists of two consecutive convolutional layers both of which use small 3×3 kernels, then there is a ReLU (Rectified Linear Unit) activation and a max pooling operation with 2×2 window and a stride of 2 for down sampling. This path obtains the entire contextual information of the traffic signs, that is their geometric shapes and also the colour patterns. There is a bottleneck in the center of U-net which just has convolutional layers without pooling layers. Here, the model extracts features in the most abstract way, thus it gets the traffic sign in a very tightly compressed latent space. Because of this bottleneck, the features that are most important for differentiating traffic signs from background elements, such as trees, vehicles, road textures, etc., are kept. The decoder part can increase the spatial resolution by upsampling feature maps with transposed convolutions (up-convolutions). The position of each upsampling step is linked with the corresponding encoder feature maps through skip connections. These connections provide the network with the areas between the detailed spatial features where the edges and outlines of traffic signs are located. The co-function enables the network to have both the global context and precise localization which is very essential for locating small and distant traffic signs.

Figure 4 depicts the U-Net architecture for traffic sign classification while figure 5 illustrates the output of the segmented image.

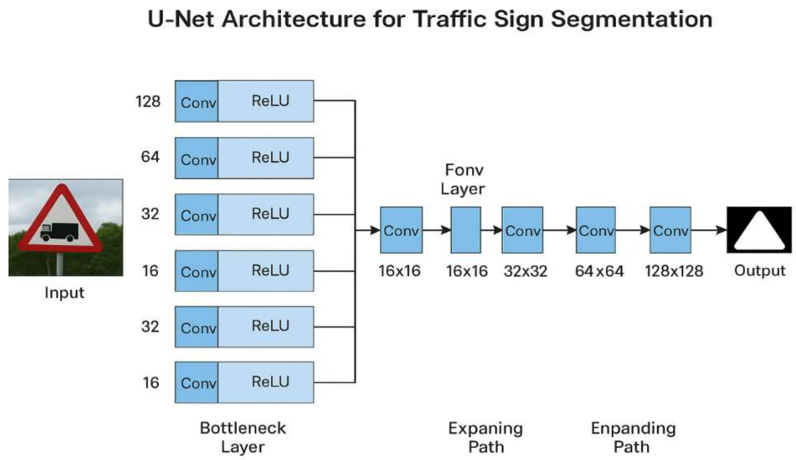


Fig. 4. U Net Architecture for Traffic Sign Segmentation

Fig. 5. Segmented Image using U Net Architecture

IV. System Implementation

The CNN takes in images of 150×150 pixels and operates a sequential compositional structure. Initially, it uses Conv2D layers (128 and 256 filters) for the extraction of the basic and advanced features of the images, such as shapes and textures. Each convolutional layer employs the ReLU activation as its non-linearity, and after that comes the MaxPooling2D layers to cut down on the spatial dimensions, lower the computational cost, and solve the problem of overfitting. In SoftMax layer, the probabilities of each category are presented as the output of the network. The stages of the model end with a categorical cross-entropy loss function and the Adamax optimizer which is quite fitting for multi class classification. The Adamax optimizer, a variation of the Adam method that utilizes the infinity norm, is the one that guarantees a smooth convergence with the presence of the sparse gradients. Figure 6. shows the CNN architecture for traffic sign classification.

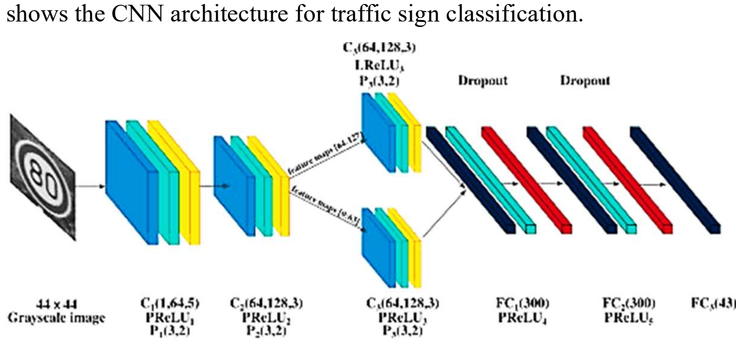
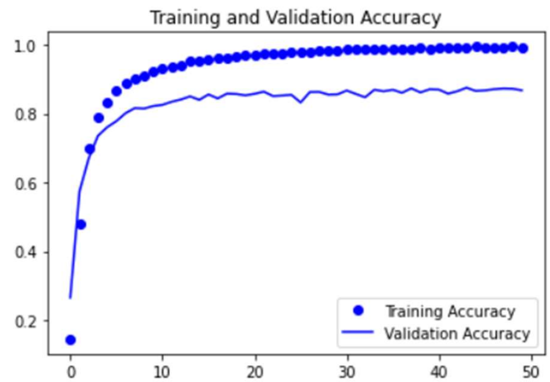


Fig. 6. CNN architecture for Traffic Sign Classification

The model was trained for 50 epochs with a learning rate of 0.0001. Figure 7 shows the training and validation accuracy over 50 epochs, while Figure 8 depicts the loss. Whenever the validation loss is more visible, the training accuracy that is also showcased in Figure 7, decreases, but with every iteration, it goes up. The model obtained training accuracy is 84.3% and validation accuracy is 82.1% using CNN with 80:20 training and validation split.



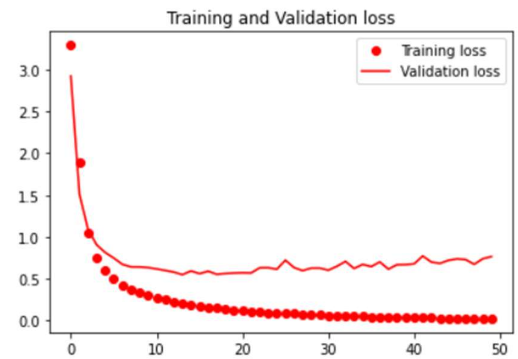
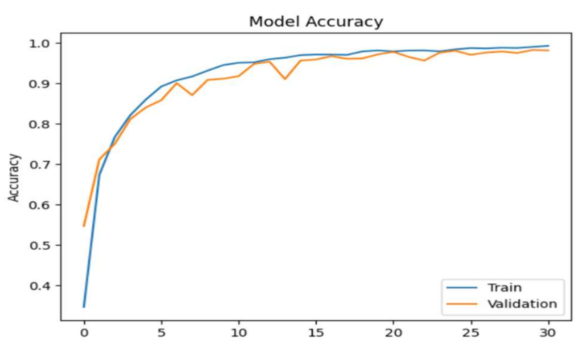


Fig. 7. CNN Training and Validation Accuracy

Fig. 8. CNN Training and Validation Loss

The ResNet152V2 model consists of a 7×7 convolution and max pooling at the beginning, followed by single, double, triple, and quadruple stages of bottleneck residual layers respectively. In every one of the bottlenecks, the 1×1 → 3×3 → 1×1 convolution operations are carried out where Batch Normalization and ReLU activation are combined before convolution (pre-activation) thus enabling better gradient propagation in deep networks. The input from the skip connections is directly combined with the output of the block, thus, the effective residual learning is guaranteed. Once feature extraction is done, the network applies global average pooling to get a concise feature vector, which is subsequently passed to a fully connected softmax layer. This final layer for traffic sign recognition has been modified to match the number of classes in the dataset, thereby facilitating correct classification of different traffic sign categories.The model was trained for 50 epochs with a learning rate of 0.0001. Figure 10 shows the training and validation accuracy over 50 epochs, while Figure 11 depicts the loss. Whenever the validation loss is more visible, the training accuracy that is also showcased in Figure 10, decreases, but with every iteration, it goes up. The model obtained training accuracy is 99.2% and validation accuracy is 98.3% using resnet152V2 with 80:20 training and validation split.



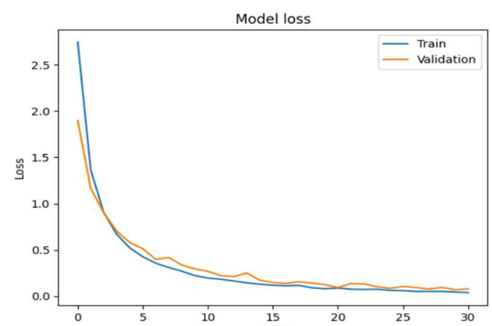


Fig. 10. Resnet152V2 Training and Validation Accuracy

Fig. 11. Resnet152V2 Training and Validation Loss

Table 2 illustrates the comparison between the CNN and ResNet152V2 models' performance based on the main four evaluation metrics. The CNN model was able to achieve an accuracy of 84.30% respectively, thus the performance level was reasonable. On the other hand, the ResNet152V2 model had a remarkable enhancement of performance with an accuracy of 99.20%. The findings indicate that the ResNet152V2 model is superior to the conventional CNN model in all the evaluation metrics and, therefore, this is an indication of the ability of the ResNet152V2 model to delve deeper for the discriminative and most separable features which leads to the increase of the overall classification accuracy.

TABLE II. CLASSIFICATION REPORT

Model	Accuracy	Precision	Recall	F1-Score
CNN	84.30	83.30	84.20	84.20
Renet152V2	99.20	98.90	98.40	98.90

V. CONCLUSION

The efficient management of traffic sign inventory is a very significant measure that, on the one hand, ensures the safety of the drivers, and on the other hand, makes the traffic flow smoothly. Generally, this operation is done manually. Present study results effectively convey the superior power of sophisticated deep learning architectures for image classification tasks. The comparison of standard CNN and ResNet152V2 models performance indicates that ResNet152V2 obtains significantly higher accuracy, precision, recall, and F1-score. This better result is mainly due to the implementation of residual connections that allow deeper network architectures and more effective feature extraction. The findings highlight the importance of deep residual networks in tackling complex visual recognition problems, thus, they become more robust, have better generalization, and are more efficient. Hence, this research positions ResNet152V2 as a very potent model in achieving state-of-the-art results in image classification, thus, it opens the door to the next research and applications in advanced computer vision systems.

REFERENCES

[1] Megalingam, R. K., Thanigundala, K., Musani, S. R., Nidamanuru, H., & Gadde, L. (2023). Indian traffic sign detection and recognition using deep learning. International journal of transportation science and technology, 12(3), 683-699.

[2] Nuthakki, R., Aameen, A., Kumar, N., & Mishra, S. K. (2023, September). Traffic Signal Recognition System Using Deep Learning. In 2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET) (pp. 636-639). IEEE.

[3] Baruah, A., Kumar, R., & Gupta, M. (2023, April). Analysis of traffic sign recognition for automated transportation systems using neural networks. In 2023 IEEE 8th International Conference for Convergence in Technology (I2CT) (pp. 1-5). IEEE.

[4] Utane, A., Mohod, S., Thakare, Y., & Ismail, F. S. M. (2025). An ensemble learning with deep feature extraction approach for recognition of traffic signs in advanced driving assistance systems. Journal of Integrated Science and Technology, 13(6), 1133-1133.

[5] Manongga, W. E., & Chen, R. C. (2025). Road Marking Sign Damage Detection and Classification Using Deep Learning and Fuzzy Method. IEEE Access.

[6] Cai, W., Yu, X., & Rong, X. (2025). CRS-NET: A CNN Network for Traffic Sign Detection. IEEE Access.

[7] Qiao, X. (2023). Research on Traffic sign recognition based on CNN Deep Learning Network. Procedia Computer Science, 228, 826-837.

[8] Khan, M. A., Park, H., & Chae, J. (2023). A lightweight convolutional neural network (CNN) architecture for traffic sign recognition in urban road networks. Electronics, 12(8), 1802.

[9] Kishor, K., Sangal, S., & Shahroz. (2023, March). Real-Time Traffic Signs and Lane Line Detection. In Doctoral Symposium on Computational Intelligence (pp. 897-906). Singapore: Springer Nature Singapore.

[10] Postovan, A., & Eraşcu, M. (2023, September). Architecturing binarized neural networks for traffic sign recognition. In International Conference on Artificial Neural Networks (pp. 87-98). Cham: Springer Nature Switzerland.