COMP9444 Project Summary

Traffic Sign Recognition

Noname

Boyang Peng z5499630 Russell Soon Heng Ung z5485791 Kaize Niu z5496085
Qinfei Chen z5472548 Zhuang Li z5228265

I. Introduction

In the rapidly evolving landscape of intelligent transportation systems, the capability to recognise traffic signs accurately and efficiently is paramount. This ability is crucial not only for the advancement of autonomous vehicles but also for enhancing the safety and efficiency of human-driven vehicles. Traffic sign recognition (TSR) systems serve as a critical component in a wide range of applications, from driver assistance systems to fully autonomous driving solutions.

The aim of this paper is to explore and analyse the efficacy of various deep learning models in the domain of traffic sign recognition. These models include DenseNet, MobileNet, EfficientNet, a custom Convolutional Neural Network (CNN), and ShuffleNet. Each of these models brings a unique set of features and optimisations, making them potential candidates for robust and efficient TSR systems.

The importance of this problem cannot be overstated. Accurate traffic sign recognition is essential for maintaining road safety, ensuring compliance with traffic laws and facilitating smooth traffic flow. In the context of autonomous vehicles, reliable TSR is a non-negotiable prerequisite for safe operation. Moreover, in human-operated vehicles, advanced driver-assistance systems (ADAS) that include TSR capabilities can significantly reduce the risk of accidents caused by human error.

Our contribution in this paper lies in the comprehensive evaluation and comparison of these state-of-art deep learning models int the context of traffic sign recognition. We have undertaken a rigorous analysis of each model, considering factors such as accuracy, computational efficiency and the ability to operate under varying environmental conditions. Additionally, we have developed a custom CNN model tailored specifically for TSR, aiming to strike an optimal balance between performance and resource efficiency.

By presenting a detailed comparison of these models, this paper aims to contribute to the ongoing research in TSR systems. Our findings provide valuable insights into the strengths and limitations of each model, guiding future research and development in this field. Moreover, our custom CNN model offers a novel approach that could potentially improve current practices in traffic sign recognition, paving the way for more advanced and reliable TSR systems in both autonomous and human-operated vehicles.

In summary, our work not only addresses a critical aspect of intelligent transportation systems but also provides a comprehensive framework for evaluating and improving traffic sign recognition technologies. By leveraging the latest advancements in deep learning, we aim to contribute to the enhancement of road safety and the advancement of autonomous driving technologies.

II. Literature Review

TSR is an area that has been significant advancements with the evolution of machine learning and computer vision technologies. Prior to the widespread adoption of deep learning, traditional image processing techniques like color segmentation, shape detection and feature extraction using Histogram of Oriented Gradients (HOG) or Scale-Invariant Feature Transform (SIFT) were commonly employed.

Before the deep learning era, Support Vector Machines (SVMs) were extensively used for traffic sign recognition due to their effectiveness in handling high-dimensional data and their ability to model complex, non-linear decision boundaries. SVMs combined with HOG feature descriptors were a standard approach [1]. Simpler neural network architectures were also used, although less powerful than their modern counterparts, they laid the foundation for subsequent developments in neural network-based traffic sign recognition [2].

These traditional methods, while effective in certain scenarios, generally lag behind modern deep learning approaches in terms of accuracy and robustness, particularly in challenging conditions such as varying illumination, occlusions and different viewing angles. The introduction of convolutional neural networks (CNNs) marked a significant leap in TSR, leading to current state-of-art methods that rely on deep learning.

III. Methods

In this project, we employed several deep learning models for traffic sign recognition, DenseNet, EfficientNet, ShuffleNet, Custom CNN and MobileNet. Each model was chosen for its unique characteristics and strengths in image recognition tasks. The project also incorporated an ensemble method to enhance the final recognition accuracy.

DenseNet was selected for its densely connected layers, facilitating feature reuse and reducing the vanishing-gradient problem. This architecture is beneficial for recognizing subtle features in traffic signs. EfficientNet is known for its scalability and efficiency, it's ideal for real-time application where both accuracy and computational resources are a concern. ShuffleNet was chosen for its low computational cost while maintaining high accuracy, making it suitable for deployment in mobile or embedded systems. While custom CNN was developed to investigate the potential of a tailored architecture for the specific nuances of traffic sign imagery. MobileNet was also opted for its balance between accuracy and efficiency, particularly in mobile or resource-constrained environments. These models were also fine-tuned on the GTSRB dataset which helps in achieving high accuracy without the need for training the entire model from scratch. In this project only the ShuffleNet model is initialize from the torch vison module and loaded with pretrained argument set to true. Initially we tried to train the ShuffleNet from scratch but we encountered a lot of problems during the process so we use pretrained weights instead.

IV. Experimental Setup

A. Dataset overview

For our traffic sign recognition project, we utilised the German Traffic Sign Recognition Benchmark (GTSRB) dataset, a widely recognised and comprehensive dataset in the field of traffic sign classification. The dataset comprises over 50,000 images, offering a robust foundation for training, validating and testing our models. There are 43 different classes, representing various traffic signs. This diversity presents a realistic challenge in recognizing a wide range of traffic signs. Images in the dataset vary in sizes and are provided in a range of resolutions, simulating real-world conditions where traffic signs can appear at different distances and angles. Each image is annotated with the class label, ensuring accurate supervision during the training process. The GTSRB dataset can be accessed at the following URL: GTSRB Dataset.

B. Dataset exploration

During our initial data exploration, we observed that certain classes are more heavily represented than others, highlighting the need for techniques to handle class imbalance. The images also exhibit a wide range of lighting and weather conditions, underscoring the necessity for models that are robust to such variations. Some images also have partial occlusions and background noise, posing additional challenges for accurate recognition.

C. Evaluation strategy

Our evaluation strategy is to split the dataset into train, validation and test dataset. Then, we use the accuracy, precision, recall and f1-score to evaluate the performance of each model, providing a comprehensive view of their effectiveness in classifying traffic signs. Besides that, we also implemented data assembling techniques to combine the predictions of the 5 models we used to get a more accurate predictions and higher accuracy.

D. Model Hyperparameters

For each model, we experimented with different learning rates to find the optimal speed at which the model learns. The batch size was adjusted based on the model and computational resources available, balancing training speed and memory usage. The number of epochs was set to 10 initially to see how each model performs and then set to 20 epochs later to ensure that the model is properly trained and shows signs of convergence. We also used advanced optimisers like Adam for each model.

E. Preprocessing and Data augmentation

We prepared these data through careful pre-processing and augmentations to ensure out model learns to recognize and classify sings accurately under diverse conditions. We applied techniques such as rotation, resizing, normalization, flipping and color adjustments were also applied to enhance the robustness of the models against real-world variations in traffic sign appearances.

V. Results

The experiment evaluated five different models, DenseNet, MobileNet, ShuffleNet, Custom CNN and EfficientNet using the GTSRB dataset. The key performance metrics were accuracy, precision, recall, f1-score and time to predict images of 10 batches. The results are shown in the table below.

Models	Accuracy	Precision	Recall	F1-score	Time to predict images of 10 batches (seconds
DenseNet	98.35	98.38	98.35	98.34	0.0144
MobileNet	94.72	94.89	94.72	94.72	0.0055
ShuffleNet	95.46	95.50	95.46	95.45	0.0073
Custom CNN	97.26	97.31	97.26	97.24	0.0059
EfficientNet	93.93	93.89	93.93	93.84	0.0173

From the results, we can see that DenseNet showed the highest accuracy of 98.35%, precision of 98.38%, Recall of 98.35% and f1-score of 98.34%, the time for it to predict the images over 10 batches is 0.0144 seconds. MobileNet demonstrated a balance between speed and accuracy. ShuffleNet had similar performance to MobileNet with a slightly higher recall but took a bit longer for prediction. Custom CNN achieved impressive results with an accuracy of 97.26%, close to DenseNet and a fast prediction time equal to MobileNet. Lastly, EfficientNet had a good balance of accuracy of 93.93% and precision of 93.89% but was slightly slower in predicting with is 0.0173 seconds. After training all the models, we also used averaging ensemble method to combine the predictions and get a accuracy of 99%

In terms of real-word application feasibility, DenseNet and Custom CNN demonstrate high accuracy and precision, making them suitable for scenarios where errors margins are minimal. However, their prediction times, although fast, are not the quickest among the tested models. MobileNet and ShuffleNet, with their quicker prediction times, are more suitable for real-time applications, despite a slight compromise in accuracy. Lastly, EfficientNet presents a balanced solution but might not be the best choice for time-sensitive applications.

Comparing to traditional methods in the literature such as SVMs, these deep learning models provide significantly higher accuracy and are more adept at handling complex image data. In the context of state-of-art deep learning models for image recognition, the DenseNet and Custom CNN models, in particular, show competitive accuracy. However, when considering real-time application requirements, MobileNet and ShuffleNet are more practical. The custom CNN, tailored specifically for this project, shows that carefully designed architectures can compete closely with establish models like DenseNet.

The DenseNet model, with the highest accuracy among the tested models, is very close to the state-of-art in image classifications tasks. Its performance in traffic sign recognition is indicative of its robustness and capability in handling complex visual tasks. The efficiency of MobileNet and ShuffleNet, particularly in prediction time, positions them favorably for real-time or mobile applications, even though they might not match the absolute accuracy of DenseNet. The Custom CNN's performance underscores the potential of bespoke models tailored to specific datasets and tasks.

VI. Conclusions

From the selection and comprehensive evaluation of five different neural network for traffic sign recognition, development of a custom CNN model tailored specifically to the nuances of the GTSRB dataset, demonstrating competitive performance. Implementation of an average ensemble method to combine the strengths of each model, enhancing overall predictive accuracy. There is also detailed benchmarking of the models against key metrics such as accuracy, precision, recall, f1-score and prediction speed.

The key strengths of the proposed solutions particularly with DenseNet and Custom CNN achieves high levels of accuracy and precision which is crucial for the reliability of traffic sign recognition systems. Models like MobileNet and ShuffleNet provide a good balance between computational efficiency and accuracy making them suitable for real-time applications. The ensemble approach enhances the robustness of solution by aggregating the predictions from multiple models.

The limitations of the solutions is where models like DenseNet and EfficientNet are computationally intensive limiting their deployment in resource-constrained environments. The current study primarily focuses on dataset-driven evaluation. Real-world conditions such as varying lighting, weather conditions, occlusions have not been extensively tested. For certain models, the time to predict is a concern, particularly in scenarios where real-time decision-making is crucial.

If more time were given, extensive testing and validation in real-world scenarios including adverse weather and lighting conditions needs to be done. Further optimisations are required to improve the efficiency.

VII. Reference

[1] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.

[2] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," in Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998, doi: 10.1109/5.726791.

[3] J. Stallkamp, M. Schlipsing, J. Salmen, C. Igel, Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition, Neural Networks, Available online 20 February 2012, ISSN 0893-6080, 10.1016/j.neunet.2012.02.016. (http://www.sciencedirect.com/science/article/pii/S0893608012000457) Keywords: Traffic sign recognition; Machine learning; Convolutional neural networks; Benchmarking

[4] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 2261-2269, doi: 10.1109/CVPR.2017.243.