Journal of Applied Intelligent System (e-ISSN : 2502-9401 | p-ISSN : 2503-0493)
Vol. 8 No. 3, November 2023, pp. 400 – 412
DOI:

Enhancing Augmentation-Based Resnet50 for Car Brand Classification

Triga Agus Sugiarto[1], **Moch Arief Soeleman**[2]
[1,2]Informatics Engineering, Computer Science Faculty, Dian Nuswantoro University
E-mail : triga3390@gmail.com[1], arief22208@gmail.com[2]
**Corresponding author*

Pujiono[3]
[3]Informatics Engineering, Computer Science Faculty, Dian Nuswantoro University
E-mail: pujiono@dsn.dinus.ac.id[3]

Abstract – This research focuses on car classification and the use of the ResNet-50 neural network architecture to improve the accuracy and reliability of car detection systems. Indonesia, as one of the countries with high daily mobility, has a majority of the population using cars as the main mode of transportation. Along with the increasing use of cars in Indonesia, many automotive industries have built factories in this country, so the cars used are either local or imported. The importance of car classification in traffic management is a major concern, and vehicle make and model recognition plays an important role in traffic monitoring. This study uses the Vehicle images dataset which contains high-resolution images of cars taken from the highway with varying viewing angles and frame rates. This data is used to analyze the best-selling car brands and build car classifications based on output or categories that consumers are interested in. Digital image processing methods, machine learning, and artificial neural networks are used in the development of automatic and real-time car detection systems. The ResNet-50 architecture was chosen because of its ability to overcome performance degradation problems and study complex and abstract features from car images. Residual blocks in the ResNet architecture allow a direct flow of information from the input layer to the output layer, overcoming the performance degradation problem common in neural networks. In this paper, we explain the basic concepts of ResNet-50 in car detection and popular techniques such as optimization, augmentation, and learning rate to improve performance and accuracy. in this study, it is proved that ResNet has a fairly high accuracy of 95%, 92% precision, 93% recall, and 92% F1-Score..

Keywords - CNN, ResNet-50, Augmentation, LR Scheduler, Car Brands, Vehicles, Classification.

1. INTRODUCTION

Indonesia stands out as a populous nation characterized by significant daily mobility patterns. Among the prevalent modes of transportation, cars have emerged as a primary choice for the public. The escalating number of car enthusiasts in Indonesia has attracted numerous automotive industries to establish manufacturing factories within the country. Consequently, the vehicles traversing Indonesian roads represent a fusion of domestically manufactured cars and imported ones originating from diverse nations. This phenomenon has led to the growth of industries that not only cater to a substantial labor force but also drive economic expansion [1], [8]. Vehicle classification stands as a pivotal aspect within modern traffic supervision and regulation systems. The recognition of vehicle brands and models assumes a vital role in traffic monitoring [2], [9]. Based on their specifications, cars come in various variants or types. However, the "Vehicle images dataset for brand and model recognition," as presented in Data in Brief, does not provide a distinguishing ranking on its website regarding which car brands are most favored by consumers. This dataset predominantly consists of high-quality images. Compiled data from the year 2021, totaling 3847 instances across 48 image attributes, was extracted from high-resolution videos (1920x1080) captured by cameras stationed along roadways, capturing varying perspectives and frame rates. Such classification is indispensable to enable consumers to analyze the most popular car brands in their highly desired categories. The abundance of different car brands necessitates the creation of classes that are categorized as either best-selling or not, thereby enabling manufacturers and consumers to analyze the car brands that are most successful in production based on their outputs or categories [2]–[9]. In recent years, the advancement of computer technology and image processing has led to the development of highly sophisticated and precise car detection systems, utilizing methods like digital image processing, machine learning, and artificial neural networks. This progress holds immense significance, particularly in the realms of traffic surveillance and road safety system enhancement. Notably, the ResNet50 architecture has gained prominence as a widely adopted neural network structure in image processing, renowned for its capacity to surmount performance degradation issues often associated with deepening networks [10]–[18]. Within the domain of car classification, Reset serves as a valuable model for identifying car types and categories within images, adept at capturing intricate visual features that enable the differentiation of diverse car types [16]–[18]. A pivotal aspect of Reset is its incorporation of residual blocks, a groundbreaking innovation enabling direct information flow from input to output layers, effectively addressing the challenge of performance degradation during network depth expansion. This attribute empowers Reset to achieve exceptional performance in car classification, establishing its efficacy as a neural network architecture for recognizing varied car types from images. This exposition extensively elucidates the concept of the Reset 50 architecture in car detection, delving into its applications to heighten the precision and reliability of car detection systems. The narrative further encompasses an exploration of popular techniques inherent to Reset 50, including optimization, augmentation, and learning rate adjustments [19]–[21]. By comprehending the foundational principles and advantages inherent to the Reset 50 role in car classification, the overarching goal is to optimize car classification systems, leading to a substantial enhancement in knowledge and efficiency pertaining to car brand recognition.

2. RESEARCH METHOD

2.1. Proposed Method

The method used to conduct research in the project to design vehicle detection applications is the classification method. The following are the stages carried out in the study using the categorization method: The following is a brief overview of the workflow of the classification method for enhancing the performance of the ResNet-50 model based on augmentation and the learning rate scheduler with the Adam optimizer used to detect car brands. To apply the method used, at this stage, the car image data in this scenario were divided into two groups, namely training and testing. Training data is a collection of existing datasets, and later we will train them to make brand predictions in advance of accuracy, so the more datasets that are similar, the higher the accuracy that will be produced. Data testing is a way for us to test programs or applications that have been made using data from outside, and later the data will be compared with the data set. That is why the data set must have a large number, because the larger the number of data sets, the more accurate the data we get. Each step in Figure 1 is explained in detail in the next section.

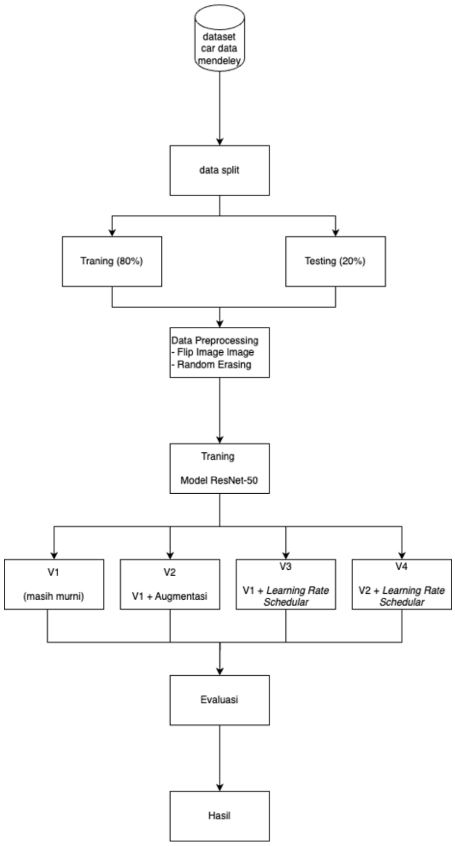

Figure 1 flowchart of the preprosed model

2.2. Dataset

The dataset consists of 3847 images of different vehicles with different makes and models. The dataset has 48 different vehicle model classes annotated in 48 different folders; each folder is named with the respective vehicle make and model name. The data collection is taken from the following link: https://data.mendeley.com/datasets/hj3vvx5946/1. An example of the dataset used can be seen in Figure 2 and 3 to show how many images are in the form of graphs.

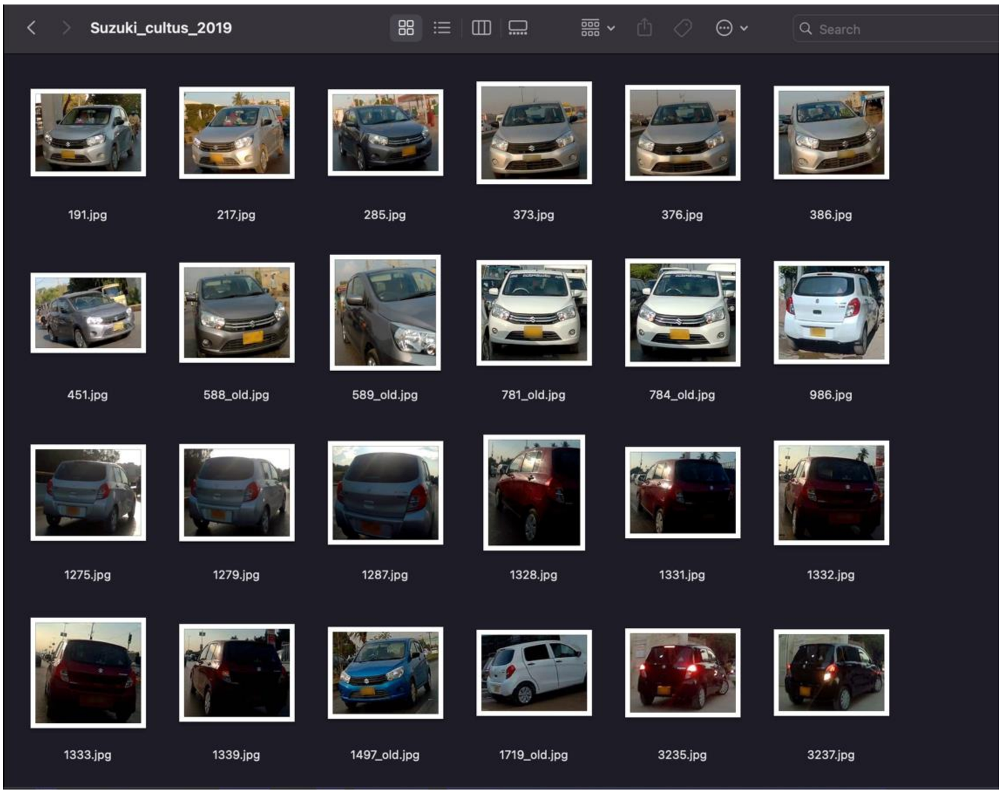

Figure 2 Dataset sample

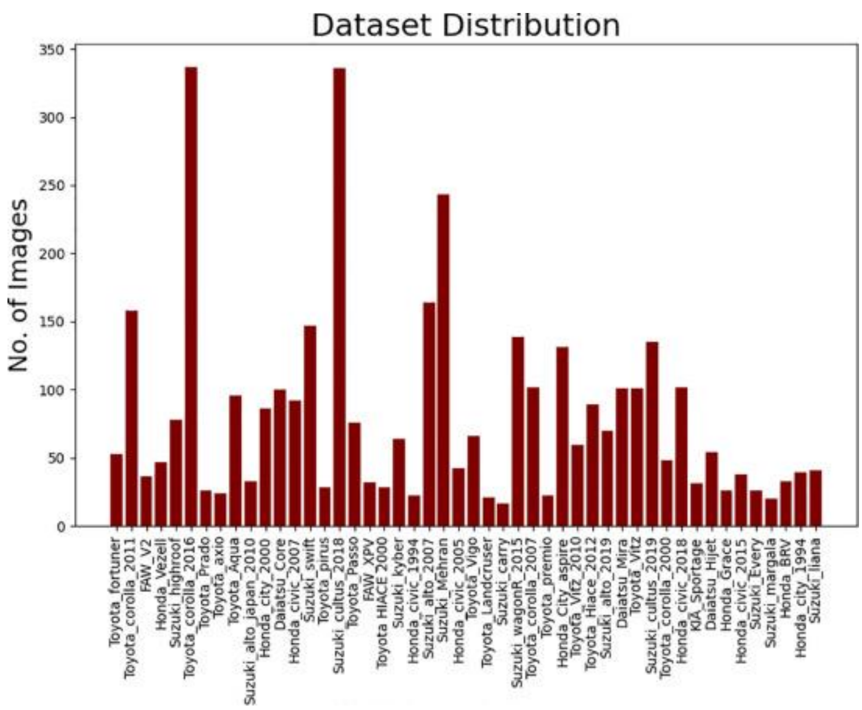

Figure 3 Graph vehicle

2.3. *Data Split*

Separation of data carried out in this study using ratio splitting of 80%–20% will divide the dataset into two parts, including 80% as training data and 20% as testing data. This approach aims to optimize model performance by using enough data to train the model effectively and ensuring that the model has a good ability to generalize to data that has never been seen before. The training data (80%) is used to train the model, estimate parameters, and learn patterns from the dataset. After training the model, testing data (20%) is used to objectively test the model and measure its performance with appropriate evaluation metrics. By separating the data into training and validation sections, we can get better estimates of model performance and avoid overfitting or underfitting. This 80%–20% data split approach strikes a good balance between training a model with enough data and testing a model with independent data to verify its effectiveness and reliability.

2.4. Data Preprocessing

In data pre-processing, augmentation stands out as a crucial technique employed to enhance the diversity of datasets and improve the efficacy of machine learning models. This method involves introducing controlled variations in the original image, thereby enhancing the robustness of the model and its generalization capabilities. One commonly employed approach to augmentation is random horizontal flips, wherein the image is mirrored along the horizontal axis, introducing variations in orientation and perspective. This technique is particularly valuable in training the model to recognize objects from different angles. Moreover, the integration of random erasing further enhances the diversity of the dataset by selectively covering parts of the image with random patterns, simulating occlusion events or missing data. By combining random horizontal flips and random erasing, the completeness of the dataset is significantly expanded, enabling the model to learn and adapt to a broader range of real-world scenarios during data pre-processing. This ultimately leads to improved model performance and accuracy in various data processing tasks. For clarification, Figure 4 illustrates examples of datasets that have undergone the random horizontal flip stage, while Figure 5 depicts datasets that have undergone the random erasing stage.

Figure 4 Random Flip

Figure 5 Random Erasing

2.5. RESNET-50

ResNet-50 is a convolutional neural network architecture known for its depth and excellent performance in image recognition tasks. It is also frequently used in car brand detection. Consisting of 50 layers, ResNet-50 belongs to the ResNet family of architectures developed to address the problem of vanishing gradients when training very deep neural networks. ResNet-50's unique residual block design enables direct information flow from the input to the output of each block, overcoming problems that arise in very deep networks. This allows ResNet-50 to extract complex hierarchical features from images, including visual features that are important for detecting car brands in images. Due to its natural expertise in learning complex features, ResNet-50 is often used as a basis in a variety of computer vision applications, including object detection and image recognition, including car brand detection in a broader domain. for the ResNet-50 architecture can be seen in Figure 6.

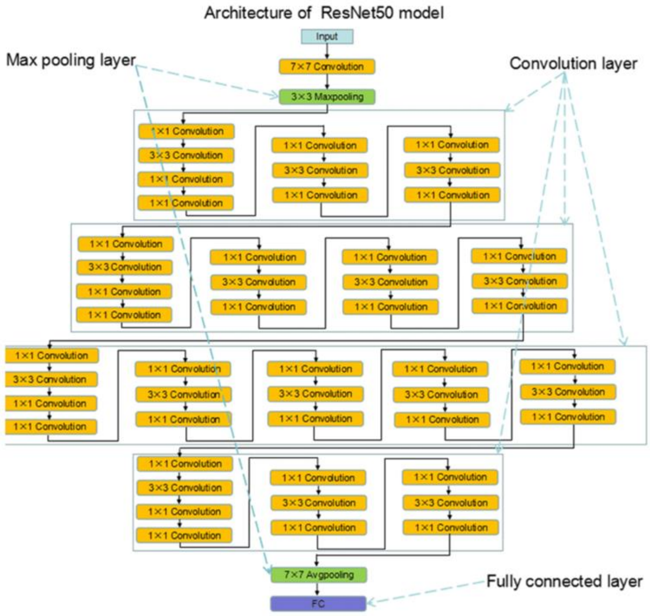

Figure 6 RESNET-50 Architecture

2.6. Model Training

Model training will be carried out using the architectural model ResNet50. ResNet50 model training will be conducted with four different scenarios to illustrate the variations in the training process. First, the model will be trained using the original dataset without augmentation and without the use of a scheduler for learning rate. This will provide a basic overview of model performance without additional manipulation of data or adjustments to learning rates. Then, in the second scenario, data augmentation will be applied to the dataset, such as rotation, cropping, or rotating, but without using the scheduler for learning rate. It aims to see the effect of augmentation on model performance without automatic adjustment for learning rate. In the third scenario, the dataset will be used without augmentation, but with the use of a scheduler for learning rate. The use of this scheduler will help in dynamically adjusting the learning rate during training, which can improve the stability and efficiency of model training. Finally, in the fourth scenario, the dataset will be augmented, and the scheduler will be used for the learning rate. By combining these two techniques, it is hoped that the model can learn better features and have better performance in terms of accuracy and generalization to data that has never been seen before. Using these four scenarios, training a ResNet50 model will provide a more complete understanding of how factors such as data augmentation and the use of the scheduler learning rate can affect the performance and generalizability of a model in an image recognition task.

2.7. Model Testing

At the model testing stage, using various evaluation metrics for classification, including accuracy, precision, recall, and F1-Score. The following is the formula used to calculate these various evaluation metrics:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\ \% \quad (1)$$

$$Precision = \frac{TP + TN}{FP + TP} * 100\ \% \quad (2)$$

$$Recall = \frac{TP + TN}{FP + TP} * 100\ \% \tag{3}$$

$$FI\ Score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} * 100\ \% \tag{4}$$

3. RESULTS AND DISCUSSION

In this discussion, it can be concluded that ResNet-50 is better when compared to CNN (4-layer convolution), which will be shown in Table 1.

Table 1 Result accuracy and loss

Model	Version	Performance Results		Best Epoch	time
		Val Accuracy	Min Val Loss		
ResNet-50	V1: Pure - Original	0.9161	0.0218	5	352
	V2: V1 + Augmentation	0.9268	0.0182	4	284
	V3: V1 + LR Scheduler	0.9467	0.0136	7	415
	V4: V2 + LR Scheduler	0.9534	0.116	10	751
CNN (4-Layer Conv)	V1: Pure - Original	0.4634	0.2191	1	55
	V2: V1 + Augmentation	0.5406	0.1355	1	55
	V3: V1 + LR Scheduler	0.5473	0.4308	1	59
	V4: V2 + LR Scheduler	0.5659	0.2219	2	116

The table above shows a comparison of ResNet-50 with CNN. CNN has an accuracy of 0.4634-0.5659 and a loss of 0.2191-0.2219, which is relatively low when compared to ResNet-50, which has an accuracy of 0.9161-09534 and a loss of 0.0218-0.116.

3.1. Model Testing Result

The following are the results of model testing on ResNet-50 and CNN Base, which will be displayed in Table 2.

Table 2 Model Test Result

Model	Version	Evaluation Metrics (%)			
		Acc	Prec	Recall	F1-Score
ResNet-50	V1: Pure - Original	0.92	0.86	0.89	0.86
	V2: V1 + Augmentation	0.93	0.88	0.91	0.88
	V3: V1 + LR Scheduler	0.95	0.91	0.92	0.91
	V4: V2 + LR Scheduler	0.95	0.92	0.93	0.92
CNN (4-Conv Layer)	V5: Pure - Original	0.46	0.34	0.42	0.45
	V6: V5 + Augmentation	0.54	0.40	0.51	0.41
	V7: V5 + LR Scheduler	0.55	0.43	0.54	0.45
	V8: V6 + LR Scheduler	0.57	0.45	0.57	0.47

In Table 2, it can be found that the ResNet-50 model is superior in all aspects to CNN, which uses 4 convolutional layers starting from accuracy, precision, recall, and F1-Score.

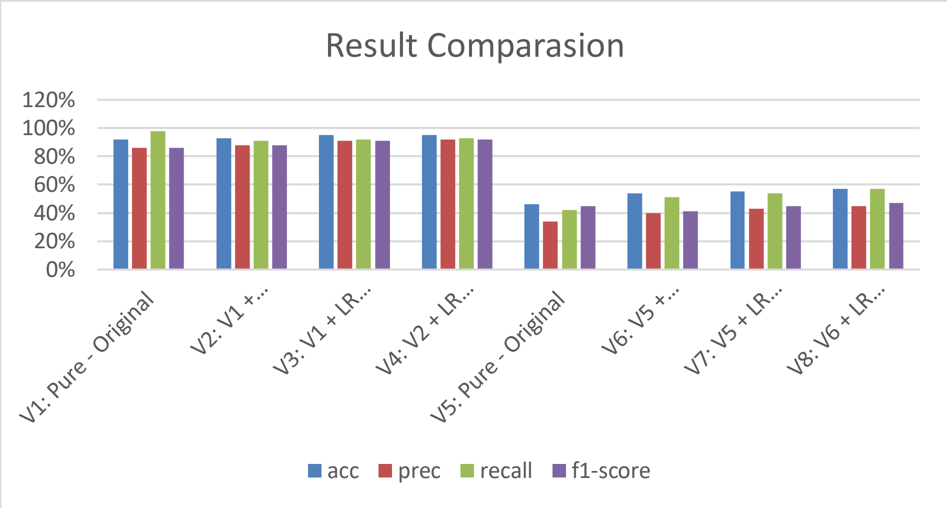

Figure 7 Result Comparison of each Trial

Figure 7 provides an outline of the performance metrics for different model variations (V1–V8) in terms of accuracy, precision, recall, and F1 score. The original model (V1) achieved high scores overall with 92% accuracy, 86% precision, 98% recall, and an 86% F1 score. Augmentation recognition (V2) maintains high performance, achieving 93% accuracy, 88% precision, 91% recall, and an 88% F1 score. Integration of LR Scheduler into the original model (V3) further improved the results, resulting in 95% accuracy, 91% precision, 92% recall, and an F1 score of 91%. The combination of augmentation and LR Scheduler (V4) maintains these high standards, achieving 95% accuracy, 92% precision, 93% recall, and a 92% F1 score. In contrast, the baseline native model (V5) performs lower with 46% accuracy, 34% precision, 42% recall, and a 45% F1 score. Augmentation (V6) improves these metrics to 54% accuracy, 40% precision, 51% recall, and 41% F1 score. Similarly, introducing LR Scheduler to the baseline (V7) resulted in 55% accuracy, 43% precision, 54% recall, and an F1 score of 45%. The combination of augmentation and LR Scheduler (V8) resulted in further improvements, achieving 57% accuracy, 45% precision, 57% recall, and an F1 score of 47%. These metrics collectively demonstrate the significant impact of various techniques on overall model performance.

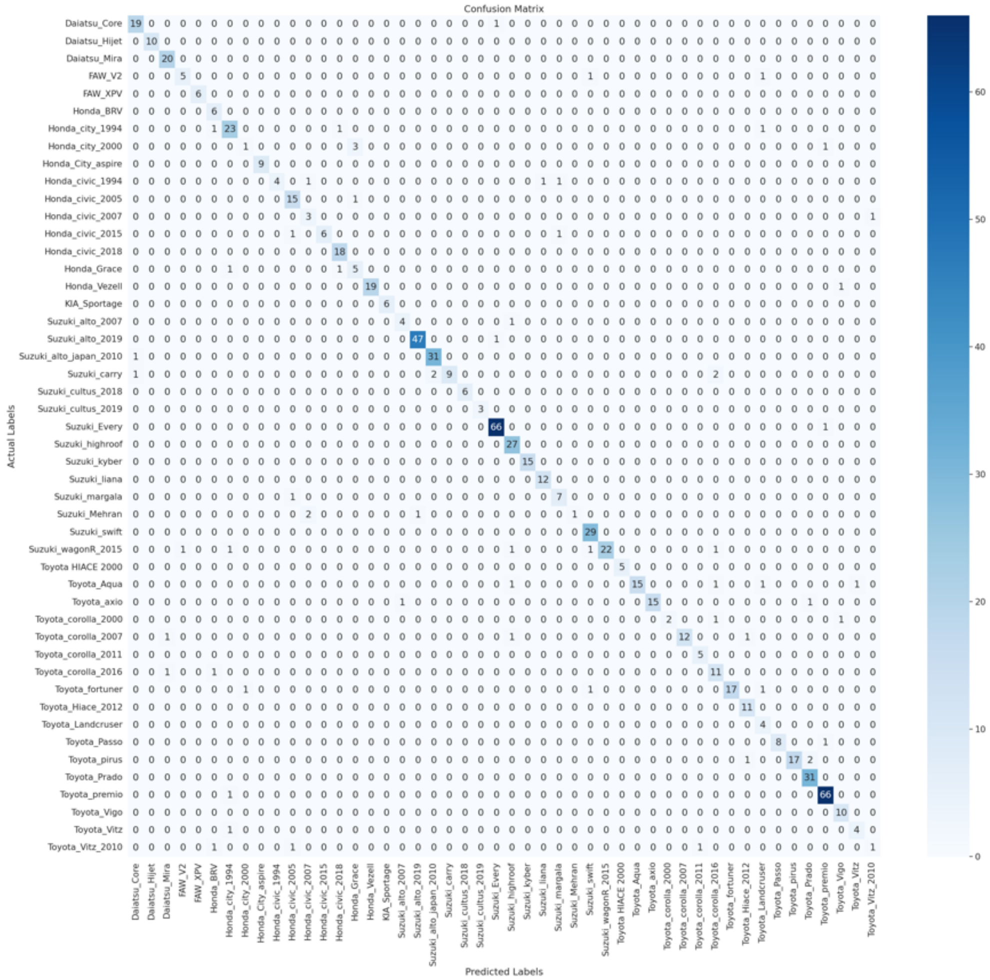

Figure 8 Confusion Matrix

Figure 8 is a confusion matrix image in the form of numbers. There are 48 types of car brands in it, as well as predicted labels and actual labels.

3.2. Model Training

The following is a comparison of the method I used with previously existing methods, which will be displayed in Table 3.

Table 3 Comparasion of methods/models

No.	Model	Accuracy
1	CNN + augmentation + LR scheduler	56.59%
2	ResNet 152 [2]	69.24%
3	MobileNet [2]	73.54%
4	VGG16 [2]	74.32%
5	ResNet-50 + Augmentation + LR scheduler	95.34%

In Table 3, it is shown that there are five algorithms that have been used to classify cars. It can be concluded that CNN + augmentation + LR scheduler is the weakest model with an accuracy of 56.59%, which proves that CNN + augmentation + LR scheduler is still the weakest rating, and ResNet50 + augmentation + LR scheduler has the highest accuracy, reaching 95.34%, where the model is the best. Used for the current research.

3.3. Implementation

The application of ResNet-50 involves predictive results obtained from test and training data for car brand recognition. In particular, for the car brand "Daihatsu_Core", the predictions from the test data show an accuracy rate of 100% with a prediction time of 0.11612 seconds, as can be found in Figure 9, while the predictions from the training data also show an accuracy of 100% with a faster prediction time of 0.04793 seconds, as can be seen in Figure 10. In addition, the predictions obtained from training data containing darker images show a slightly reduced accuracy of 97% with a prediction time of 0.08934 seconds, which can be seen in Figure 11.

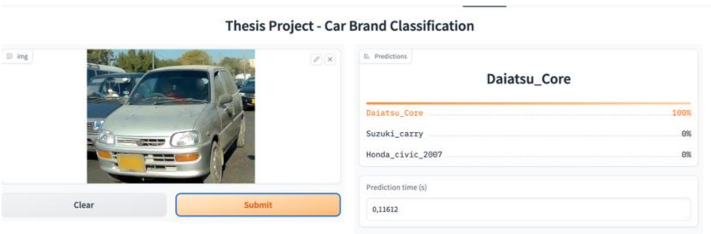

Figure 9 Car From test data

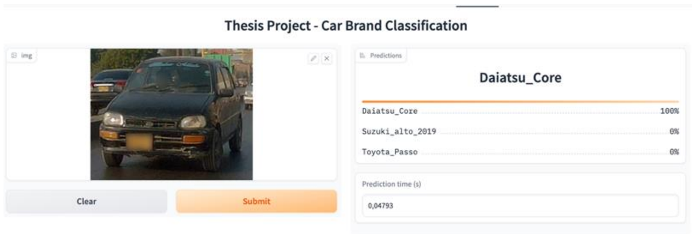

Figure 10 Car From train data

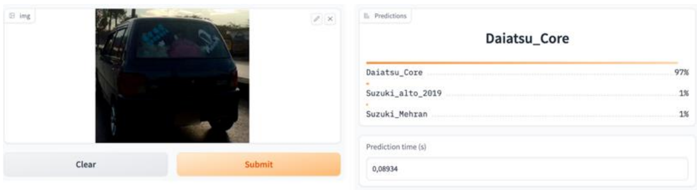

Figure 11 Picture dark car

4. CONCLUSION

To improve the accuracy of the ResNet-50 model, this research has implemented augmentation and LR scheduling. According to the results of the experiments that have been carried out, it is proven that the ResNet-50 architecture is able to provide significant performance improvements compared to conventional CNN architectures. In testing, the results showed that ResNet-50 was able to achieve an accuracy of 92%, which is a huge improvement from the 46% accuracy obtained from the previous conventional CNN. Additionally, there were considerable improvements in precision, recall, and f1-score, all of which rose from around 30–40% to the 80–90% range. By combining ResNet-50 with other methods such as data augmentation and the use of an LR scheduler, model performance can be further improved. In these tests, the results show a significant performance increase compared to using pure ResNet-50. Model accuracy increased to 95%, precision reached 92%, recall reached 93%, and f1-score reached 92%. This shows that the use of additional methods can help significantly improve the quality of model predictions. Thus, this evaluation shows that the use of the ResNet-50 architecture, either with or without the incorporation of additional methods, can bring significant improvements in the accuracy and performance of ResNet-50 models. This gives hope for further development in the use of this architecture in various applications that require better prediction capabilities.

REFERENCES

[1] H. Paul, A. Sartika Wiguna, and H. Santoso, "Penerapan Algoritma Support Vector Machine Dan Naive Bayes Untuk Klasifikasi Jenis Mobil Terlaris Berdasarkan Produksi Di Indonesia," 2023.

[2] M. Ali, M. A. Tahir, and M. N. Durrani, "Vehicle images dataset for make and model recognition," Data Brief, vol. 42, Jun. 2022, doi: 10.1016/j.dib.2022.108107.

[3] R. Walambe, A. Marathe, and K. Kotecha, "Multiscale object detection from drone imagery using ensemble transfer learning," Drones, vol. 5, no. 3, Sep. 2021, doi: 10.3390/drones5030066.

[4] U. Mittal, P. Chawla, and R. Tiwari, "EnsembleNet: a hybrid approach for vehicle detection and estimation of traffic density based on faster R-CNN and YOLO models," Neural Comput Appl, vol. 35, no. 6, pp. 4755–4774, Feb. 2023, doi: 10.1007/s00521-022-07940-9.

[5] D. Komolovaite, A. Krisciunas, I. Lagzdinyte-Budnike, A. Budnikas, and D. Rentelis, "Vehicle Make Detection Using the Transfer Learning Approach," Elektronika ir Elektrotechnika, vol. 28, no. 4, pp. 55–64, 2022, doi: 10.5755/j02.eie.31046.

[6] O. Bourja et al., "End-to-End Car Make and Model Classification using Compound Scaling and Transfer Learning," 2022. [Online]. Available: www.ijacsa.thesai.org

[7] A. Ammar, A. Koubaa, M. Ahmed, A. Saad, and B. Benjdira, "Vehicle detection from aerial images using deep learning: A comparative study," Electronics (Switzerland), vol. 10, no. 7, Apr. 2021, doi: 10.3390/electronics10070820.

[8] Z. Liu, T. Lian, J. Farrell, and B. Wandell, "Soft Prototyping Camera Designs for Car Detection Based on a Convolutional Neural Network."

[9] C. Shorten, T. M. Khoshgoftaar, and B. Furht, "Deep Learning applications for COVID-19," J Big Data, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-020-00392-9.

[10] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," IEEE Trans Neural Netw Learn Syst, vol. 33, no. 12, pp. 6999–7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827.

[11] C. Garbin, X. Zhu, and O. Marques, "Dropout vs. batch normalization: an empirical study of their impact to deep learning," Multimed Tools Appl, vol. 79, no. 19–20, pp. 12777–12815, May 2020, doi: 10.1007/s11042-019-08453-9.

[12] S. Technická univerzita (Košice, IEEE Hungary Section, M. IEEE Systems, and Institute of Electrical and Electronics Engineers, SAMI 2020 : IEEE 18th World Symposium on Applied Machine Intelligence and Informatics : proceedings : January 23-25, 2020, Herl'any, Slovakia.

[13] J. Kim and D. Lee, "Facial Expression Recognition Robust to Occlusion and to Intra-Similarity Problem Using Relevant Subsampling," Sensors, vol. 23, no. 5, Mar. 2023, doi: 10.3390/s23052619.

[14] S. H. S. Basha, S. R. Dubey, V. Pulabaigari, and S. Mukherjee, "Impact of fully connected layers on performance of convolutional neural networks for image classification," Neurocomputing, vol. 378, pp. 112–119, Feb. 2020, doi: 10.1016/j.neucom.2019.10.008.

[15] Aristoteleio Panepistēmio Thessalonikēs. Research Dissemination Center and Institute of Electrical and Electronics Engineers, 2019 8th International Conference on Modern

Circuits and Systems Technologies (MOCAST) : May 13-15, 2019, Aristotle University Research Dissemination Center (KEDEA), Thessaloniki, Greece.
[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2016. [Online]. Available: http://image-net.org/challenges/LSVRC/2015/
[17] D. Sarwinda, R. H. Paradisa, A. Bustamam, and P. Anggia, "Deep Learning in Image Classification using Residual Network (ResNet) Variants for Detection of Colorectal Cancer," in Procedia Computer Science, Elsevier B.V., 2021, pp. 423–431. doi: 10.1016/j.procs.2021.01.025.
[18] L. Wen, X. Li, and L. Gao, "A transfer convolutional neural network for fault diagnosis based on ResNet-50," Neural Comput Appl, vol. 32, no. 10, pp. 6111–6124, May 2020, doi: 10.1007/s00521-019-04097-w.
[19] E. D. Cubuk, B. Zoph, V. Vasudevan, and Q. V Le Google Brain, "AutoAugment: Learning Augmentation Strategies from Data," 2019. [Online]. Available: https://pillow.readthedocs.io/en/5.1.x/
[20] H. Inoue, "Data Augmentation by Pairing Samples for Images Classification," Jan. 2018, [Online]. Available: http://arxiv.org/abs/1801.02929
[21] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random Erasing Data Augmentation." [Online]. Available: https://github.com/zhunzhong07/Random-Erasing.