Available *online* at : http://jnte.ft.unand.ac.id/

Jurnal Nasional Teknik Elektro

| ISSN (Print) 2302-2949 | ISSN (Online) 2407-7267 |

Integrating YOLOv7 with FixMatch for Enhancing Vehicle Detection Performance in Mixed Traffic Environments

Yandri Zaita[1], *Khairun Saddami*[1,2,3*], *Nasaruddin Nasaruddin*[1]

[1]*Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia*
[2]*Telematic Research Centers, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia*
[3]*AI Research Group, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia*

ARTICLE INFORMATION

Received: October 4, 2024
Revised: November 25, 2024
Accepted: November 25, 2024
Available online: November 30, 2024

KEYWORDS

Object Detection, YOLOv7, FixMatch, Semi-supervised, Mixed Traffic.

CORRESPONDENCE

E-mail: khairun.saddami@usk.ac.id

A B S T R A C T

A major challenge in the development of object detection technology is the significant reliance on large labeled datasets, which requires substantial time and memory for manual annotation—especially in complex, mixed traffic environments with varied vehicle types, congestion levels, and unpredictable motion patterns. This study addresses this issue by integrating the semi-supervised learning technique, FixMatch, into the YOLOv7 object detection model, utilizing 4000 transportation-related datasets. The FixMatch technique enables the model to detect unlabeled objects effectively through strong and weak augmentation methods. In this study, the detected objects in the mixed traffic environment include public transportation, pedicabs, cars, motorcycles, and trucks. This study achieved an impressive 97.5% detection accuracy by leveraging unlabeled data, demonstrating the model's efficiency and effectiveness in identifying vehicles under diverse traffic conditions. Consequently, integrating the FixMatch method into YOLOv7 provides a practical and efficient solution for object detection in situations where collecting labeled data is challenging, such as in dynamic and highly variable traffic environments.

INTRODUCTION

Object detection is identifying and locating a specific object in an image or video. The main focus of object detection is to recognize the presence and position of the object and provide a bounding box to delimit the object [1]. As technology evolves, object detection has made significant progress with various methods and algorithms, such as Region-based Convolutional Neural Network (R-CNN) [3], Fast Region-based Convolutional Neural Network (Fast R-CNN) [4], Faster Region-based Convolutional Neural Network (Faster R-CNN) [5], Region-based Fully Convolutional Network (RFCN) [6], and You Only Look Once (YOLO) [7]. These models have shown improved accuracy in various case studies. However, traditional object detection models still face significant limitations in complex mixed-traffic environments. Conditions such as variations in vehicle type, size, speed, and the potential for occlusion between objects are challenges that need to be overcome to keep detection accuracy high.

The reliability of the detection model is also highly dependent on the availability of a large amount of labeled image data that must be accurately labeled, a time- and resource-consuming process [8]. We have to annotate each image before it is fed into the detection model for training data. One of the objects detected using this model is vehicles, such as cars, motorcycles, public transportation, tricycles, and trucks. Unlabeled data must go through a manual labeling stage before it is used as training data. In previous studies, difficulties have also arisen in detecting overlapping objects and the large memory requirements for processing such large-scale data [9]-[13].

To solve the challenge, this research aims to reduce the dependency on manual labeling by adopting a semi-supervised learning approach that utilizes pseudo-labeling and consistency regularization techniques applied to the FixMatch method [14], [15]. FixMatch offers an efficient solution by using unlabeled data through the concept of pseudo-labeling combined with the principle of consistency, thus enabling optimal use of unlabeled data in detection model training. As such, FixMatch can be an appropriate approach to object detection in mixed traffic environments where manual labeling is complex, offering a balance between detection efficiency and accuracy.

This research significantly contributes to vehicle detection in mixed-traffic environments using the Yolov7 method alongside FixMatch. One of the key advancements is combining the Yolov7 approach with unlabeled data to detect vehicles in these complex settings effectively. Additionally, we apply the FixMatch method, which enhances model performance by predicting labels on the

https://doi.org/10.25077/jnte.v13n3.1250.2024

vehicles in mixed-traffic situations, again leveraging the potential of unlabeled data. Together, these strategies represent a noteworthy enhancement in vehicle detection.

METHODS

The hardware used for this research is a laptop and camera with Windows 11 Home Single Language 64-bit specifications (10.0, Build 22621) (22621.ni_release.220506-1250) and 1080p resolution, frame rate 30 fps, night vision feature. Visual Studio code, Anaconda Navigator, and app. label box are supporting software for building models.

Research Methodology

This research was carried out in steps that included several stages to maximize the results, as shown in Figure 1. The first stage is collecting data from previous research [16][17]. We also did a literature study by studying and deepening the research topic from books and previous research references [18]. Then, the dataset was preprocessed by resizing to match the model requirement. Furthermore, the dataset is trained on the vehicle detection and FixMatch methods. Next is the testing stage. The program that has been designed is then made for the testing stage using the vehicle detection method and FixMatch. Suppose the test results are still far from the expected results. In that case, the algorithm needs to be improved so that the results of the vehicle detection program match the expected accuracy, the analysis stage is performed, and finally, the results of testing the vehicle detection method are explained by combining the FixMatch method [19].

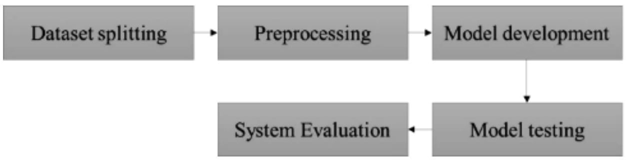


Figure 1. Research Flow

Split Dataset

The dataset used in this study was obtained from previous research called MXT-Dataset [18, 19]. The dataset consists of 4000 frames that were acquired in a mixed-traffic environment. The total number of object in the dataset are 28348 objects. The dataset-splitting process is an essential step in this research to ensure the quality and representation of the data used in training the vehicle detection model, as shown in Figure 2. After the data is collected and annotated, the dataset is divided into training, validation, and testing sets. The training set covers about 80% of the total dataset. It trains the YOLOv7 and FixMatch models, with random sorting ensuring variation in vehicle types, traffic conditions, and image capture times.

The validation set, which includes about 10% of the dataset, is used to monitor model performance during training and perform hyperparameter tuning, helping to detect issues such as overfitting and underfitting. The test set, also covering about 10% of the dataset, was used to evaluate the final performance of the model on new data that was utterly unseen during training and validation. Disaggregation of the dataset was done using stratified sampling techniques to maintain a balanced proportion of different vehicle types and traffic conditions, reduce bias, and ensure consistent representation within each set.

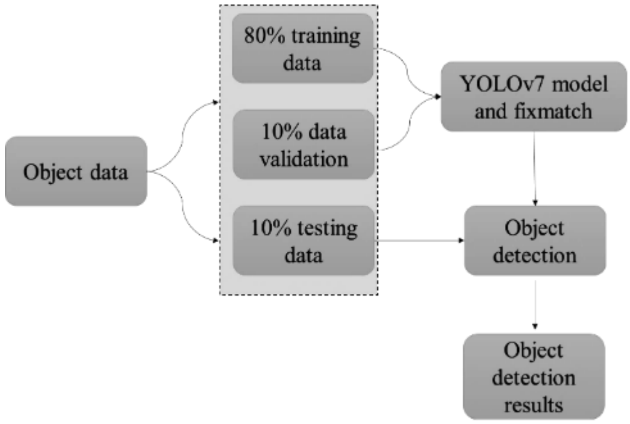


Figure 2. Data split flow

The data was collected via high-resolution cameras placed at several locations with mixed traffic characteristics at various times of day and night to capture multiple traffic conditions, as shown in Figure 3, including bright daytime conditions, nighttime with low lighting, and severe environmental conditions. Manual annotation was performed on each captured video frame, accurately labeling each vehicle to train the model with the correct data. Once labeled, the dataset was randomly but proportionally split to ensure each set represents the same traffic conditions. When applied to data that has never been seen, the resulting model has good generalization capabilities.

This dataset is divided into three main parts, as shown in Table 1. This dataset is truncated to ensure that the focal area is analyzed. The vehicle is in an optimal position within each frame. The cropping is done by separating the main object from the background or irrelevant elements so that the data used in model training is cleaner and of higher quality. This step also aims to reduce noise from areas outside the relevant objects.

The dataset used in this study consists of several categories after object separation in the image. Then, the dataset is labeled as 50% of the 28,348 objects. Thus, the total dataset used in this study is 28,348 objects. The division of this dataset is done to ensure that the model developed can be appropriately tested and produce accurate and reliable results. Proper division between training data and testing data is essential to avoid overfitting and ensure that the model can be generalized well to data that has never been seen before.

Table 1. Dataset Split

Dataset Section	**Number of Datasets**
Train	3.200
Test	400
Validasi	400
Total	**4.000**

https://doi.org/10.25077/jnte.v13n3.1250.2024

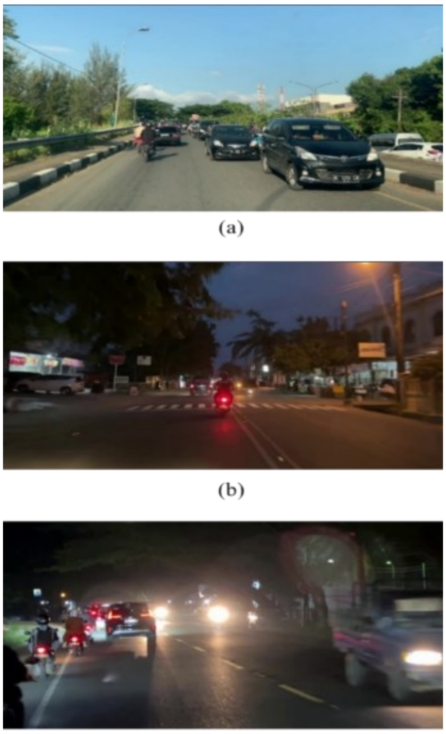

(a)
(b)
(c)

Figure 3. Dataset Image Samples from Vehicle (a) daytime (b) limited lighting (c) nighttime

Preprocessing

The dataset is preprocessed to improve the quality and fit of the model. Figure 4 illustrates the process of preprocessing object data using augmentation techniques as part of an object detection system that utilizes the YOLOv7 model coupled with FixMatch. The process starts with unlabeled object data, such as images of vehicles or other objects that must be detected. The initial stage of this research is augmentation preprocessing.

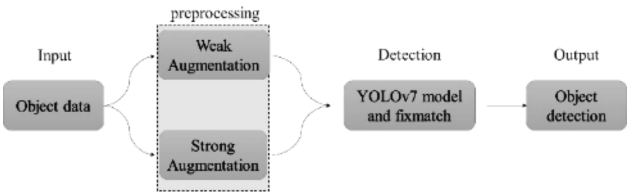


Figure 4. The Preprocessing Flow

As in Table 2, the data undergoes weak augmentation, where minor modifications such as small rotations, scale changes, or light lighting adjustments are applied. The purpose of weak augmentation is to create a slightly different image without changing the primary information of the object, thus assisting the model in generating accurate pseudo-labels for unlabeled data. Once the pseudo-label is generated, the same data is subjected to substantial augmentation, where the image is given more significant changes, such as large rotations or more extreme color distortions. The resulting data from weak and strong augmentation preprocessing is shown in Table 3.

Table 2. Object Data

Object	Number of Objects
Car	10.625
Motor	15.306
Public Transportation	838
Pedicab	779
Truck	800
Total	**28.348**

Table 3. Object data after augmentation

Object	Number of Objects
Label	14.174
Unlabel	14.174
Validasi	3.831
Testing	3.806
Total	**35.985**

This step tests the model's ability to produce consistent predictions despite drastic changes to the image. The augmented data was then fed into the YOLOv7 model combined with FixMatch, where the model was trained to produce consistent and accurate predictions based on the pseudo-labels and validation from the substantial augmentation. This process, ultimately, enables more precise and robust object detection in various environmental conditions, especially when unlabeled data is limited.

Model Development

In the YOLOv7 and FixMatch-based vehicle detection systems, the process starts by utilizing a pre-processed dataset, as described in Table 3. This dataset includes various images or video frames containing vehicles, where the data is prepared to train the YOLOv7 model with the steps shown in Figure 5. The YOLOv7, an advanced object detection model, is tasked with detecting vehicles in images or videos by generating bounding boxes and labels for each detected vehicle object. The bounding box indicates the vehicle's position in the image, while the label provides information about the type of vehicle, such as a car, truck, or motorcycle.

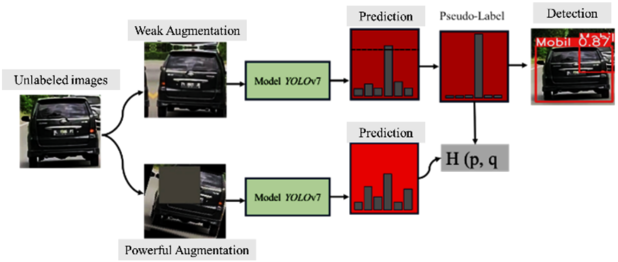


Figure 5. The proposed model

After YOLOv7 performs the initial detection, the resulting output, including the bounding box coordinates and the vehicle type label, becomes the input for the FixMatch model. The FixMatch model is a semi-supervised learning model designed to improve detection accuracy, especially when labeled data is limited. In this context, FixMatch uses two types of data: labeled data resulting from YOLOv7 and unlabeled data, which may come from other images or video frames that have not been fully processed.

FixMatch works by combining these two types of data. First, it generates predictions from the labeled data provided by YOLOv7 and then reinforces the training process using the unlabeled data.

In this stage, FixMatch performs weak and strong augmentation on the unlabeled data to generate pseudo-labels used in the training process. The model relies heavily on high-confidence predictions from the unlabeled data, and these pseudo-labels help improve the overall detection results.

The predictions obtained through FixMatch are then further processed to produce the final detection results. These detection results include more detailed and accurate information about the detected vehicle, including more precise positioning and more precise identification of the vehicle type. By combining the strengths of YOLOv7 in fast object detection and FixMatch in accuracy improvement through semi-supervised learning, the system can provide reliable and efficient detection results, even in conditions where the availability of labeled data is minimal.

This model is carried out with hyperparameter settings, as shown in Table 4. The hyperparameter selection in this study was done to achieve a balance between the accuracy and efficiency of the model in detecting objects in mixed-traffic environments. Computational limitations are a constraint on models such as YOLOv7 that require high resources, and the risk of overfitting may occur if the dataset is less diverse, reducing the reliability of the model in different actual conditions.

Table 4. Hyperparameter

No.	Hyperparameter	Specification
1	Learning Rate	$3x10^{-1}$, 10^{-4}
2	Batch Size	64
3	Epochs	170, 200
4	Optimizer	AdamW
5	Input image size	96x96
6	Unlabeled Batch Ratio	1:5

Model Testing

At this stage, test data will be used to evaluate the model's performance in unlabeled vehicle detection. The evaluation is done by measuring precision, recall, and accuracy. Precision Recall indicates how many relevant images are retrieved by the system. Accuracy reflects how close the model's prediction is to the actual value of the pictures not used in the previous stages.

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (2)$$

$$F1 - score = 2 \times \frac{precision\ x\ recall}{precision+recall} \quad (3)$$

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (4)$$

The interpretation of a confusion matrix is essential for evaluating model performance. Key components include True Positives (TP), which represent the number of samples accurately classified within the actual class. True Negatives (TN) denote the samples that are correctly identified as not belonging to the class in question. Conversely, False Positives (FP) refer to instances where samples that do not belong to the class are incorrectly classified as part of it. Lastly, False Negatives (FN) indicate samples that are part of the class but have been mistakenly classified as not belonging to it. Collectively, these elements contribute to a thorough assessment of the model's accuracy and effectiveness in image recognition tasks.

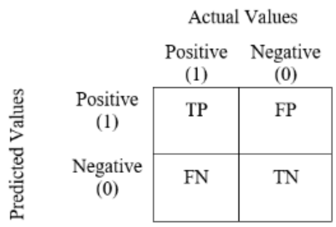


Figure 7. Confusion Matrix

RESULTS AND DISCUSSION

Augmentation result of unlabeled data

This research results in vehicle object detection using YOLOv7 and FixMatch models in a mixed-traffic environment. The detection process starts with the processing of vehicle image data involving augmentation, both weak and strong augmentation, before being trained using the YOLOv7 model. This augmentation aims to increase the diversity of the data and assist the model in recognizing vehicle objects under various conditions. YOLOv7 is used to detect vehicles in images or video frames, generating bounding boxes and labels for each detected vehicle.

The results from YOLOv7 are then optimized with the FixMatch model, which applies semi-supervised learning techniques. FixMatch utilizes the predictions from YOLOv7 as pseudo-labels for unlabeled data, improving detection accuracy with prediction consistency through more muscular augmentation. The model considers the probability distribution to compare the prediction on the original and augmented images, then uses the high similarity value to improve the training process.

In addition, some hyperparameters used in augmentation preprocessing include scaling to resize the image to detect objects of different sizes easily—rotation in rotating the image to accommodate different viewing angles in object detection. Then, translation moves the position of objects in the image to introduce spatial variations. Shifting parts of the image create mild distortions that can improve the model's robustness to object shape changes—finally, adjusting the image's brightness to enable object detection under diverse lighting conditions.

These augmentation techniques, combined with FixMatch's strength in utilizing unlabeled data, allow the YOLOv7 model to achieve more accurate detection results, especially in complex traffic environments. The final detection result provides more detailed information about the position and type of vehicle, as shown in the final result image with a bounding box and confidence score.

Training Results

A training stage is carried out so the model can recognize image data. The model training stage is carried out at two learning rate values, namely 0.3 and 0.0001. The two learning rates are

https://doi.org/10.25077/jnte.v13n3.1250.2024

compared to the best learning rate value following the built model. The results of the YOLOv7 and Fixmatch model accuracy train at a learning rate of 0.3 show a decrease from epoch 60 to epoch 180. The results of the YOLOv7 and Fixmatch model accuracy train at a learning rate of 0.0001 show a reduction from epoch 40 to epoch 60.

Based on the model accuracy train results on the learning shown in Figure 8, there are several advantages of training the YOLOv7 model combined with FixMatch. One of the main advantages is the model's ability to achieve high initial accuracy, demonstrating its effectiveness in recognizing basic patterns in the data from the beginning of training. Although the accuracy dropped in the middle of the training process, the model showed signs of recovery, as seen by the significant increase in accuracy in the final stage of training. This improvement indicates that the model has the adaptive ability to improve its performance after facing a challenging training phase. In addition, these results also show that FixMatch, as a semi-supervised learning method, successfully helps the model to make better use of unlabeled data, especially in the latter stages of training.
The model was then validated with learning rates of 0.3 and

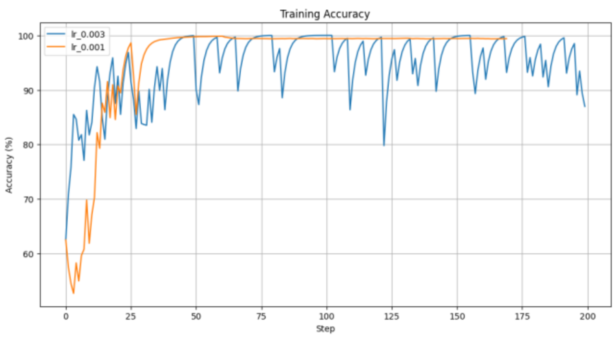


Figure 8. Accuracy of Train Model

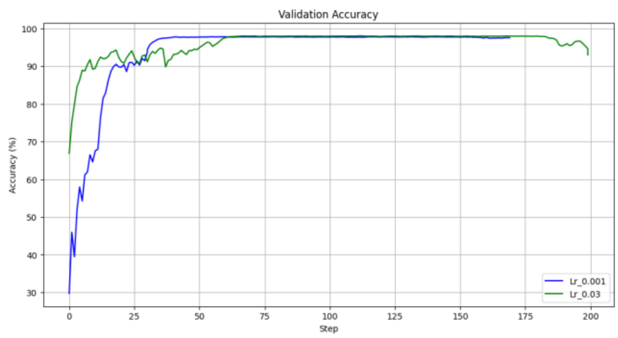


Figure 9. YOLOv7 and FixMatch Model Validation Accuracy

0.0001, and the results are shown in Figure 9. It can be seen that this validation accuracy graph highlights several advantages of the YOLOv7 model combined with FixMatch. One of the main advantages is its ability to achieve and maintain high accuracy throughout most of the training process. As shown in Figure 9, the validation accuracy value of the model with a learning rate of 0.3 reaches 98% after about 60 epochs. It demonstrates the model's effectiveness in consistently understanding and learning the data. Despite a dip at the end of the training, the accuracy value remains high at 94.39%. Meanwhile, the model accuracy value results with a learning rate 0.0001 reached 97.5% at epoch 40.

These results show that the model managed to provide performance stability after the initial training phase, indicating that the model is not only dependent on the training data but also able to generalize well to the validation data. This reflects the effectiveness of FixMatch in supporting the model in making better use of unlabeled data, thereby strengthening the model's learning ability during training.

Furthermore, a train loss is performed to determine the model's performance on new data and ensure that the model not only works well on training data but can also work well on data that has never been seen before. This training also uses a learning rate of 0.3 and a learning rate of 0.0001, as shown in Figure 10. Figure 10 shows the train loss result at a learning rate of 0.3, which is 0.4, and the train loss result at a learning rate of 0.0001, which is 0.1.

These results show the advantages of the YOLOv7 model when combined with FixMatch. One of the main advantages is the significant and stable decrease in loss during the early stages of training, which indicates that the model effectively and quickly understands the patterns in the data. The consistency of the loss rate suggests the model's ability to maintain strong performance with minimal fluctuations, which indicates efficient data utilization and reduced risk of overfitting throughout most of the training process.

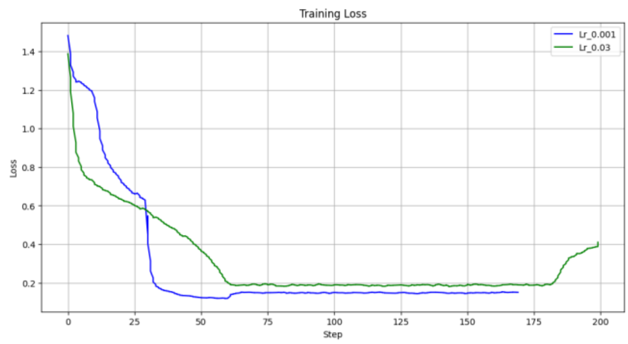


Figure 10. Train Loss Model YOLOv7 and FixMatch

Then, the loss model validation results using a learning rate of 0.3, as shown in Figure 11, show the results reaching 0.1. While the loss model validation results using a learning rate of 0.3, as shown in Figure 11, show results reaching 0.08. These results show that the advantages of the YOLOv7 model with FixMatch, especially during the early stages of training, are significant and stable loss reduction, reflecting the model's ability to understand the data pattern quickly. In addition, between the 60 and 180 epochs of training using a learning rate of 0.3, the loss remains consistently low and stable, indicating that the model can maintain optimal performance with slight variation.

This consistency indicates efficient data utilization and a low risk of overfitting during most of the training process. Table 5 shows the results of the training model comparison on the model training of YOLOv7 and FixMatch models. Based on the training, it shows that this model works well in recognizing the vehicle dataset at a learning rate of 0.0001 with 97.5% accuracy.

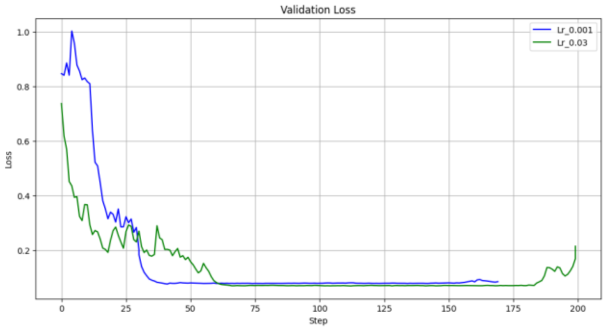


Figure 11. Validasi Loss Model YOLOv7 and FixMatch

Table 5. Training result

Learning Rate	Train Accuracy	Validation Accuracy	Train Loss	Validation Loss
0.03	0.31	93.47 %	0.41	0.21
0.0001	0.04	97.5%	0.1	0.08

Final Test

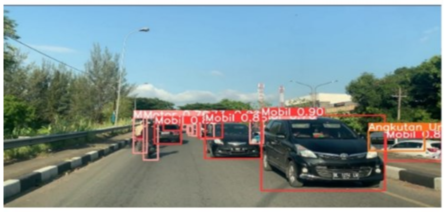
The YOLOv7 and FixMatch models used in this study successfully detect different types of vehicles in mixed-traffic environments. Based on the test results, the YOLOv7 Model coupled with FixMatch showed excellent performance detecting vehicles under various lighting conditions. During the daytime (i.e., Figure 12), the model detected almost all vehicles on the road with high accuracy, such as cars, motorcycles, and public transportation. The resulting detection boxes are clear and non-overlapping, indicating the model's ability to distinguish objects in a well-lit environment effectively.

At night, as shown in Figure 13, despite lower illumination and potential interference from vehicle lights, the model could still detect various vehicles on the road with high confidence. The detection results show little noise or false positives, indicating that the model can adapt well to lighting conditions. This adaptability is crucial for real-world applications, especially in traffic environments with diverse conditions, as shown in Figure 14.

System Evaluation

The last stage of this research is to evaluate the system that has been built using a confusion matrix. Figure 15 shows a confusion matrix of the detection model generated from this research. This confusion matrix includes several main classes from the dataset: Car, Motorcycle, Public Transportation, Rickshaw, and Truck. This matrix provides information on how the model classifies each sample in the test dataset. Figure 15 shows the parameter values of the confusion matrix for each data as shown in Table 6. Each value is calculated using equation 1 to equation 4 and shows the results as in Table 7.

(a)



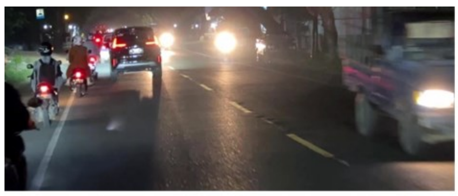
(b)

Figure 12. YOLOv7 Detection and FixMatch Daytime State (a) original (b) detection result

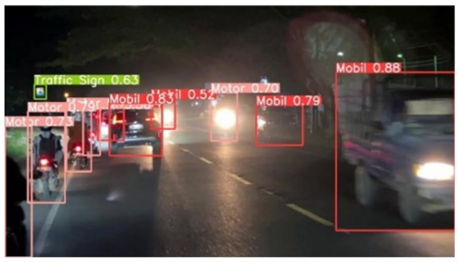
(a)

(b)

Figure 13. YOLOv7 Detection and FixMatch Night State (a) original (b) detection result



(a)



(b)

Figure 14. YOLOv7 Detection Opaque State (a) original (b) detection result

https://doi.org/10.25077/jnte.v13n3.1250.2024

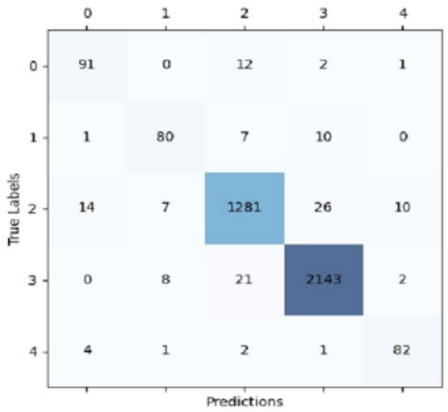


Figure 15. the confusion metrics

Table 6. Evaluation of classification models

No.	Object	TP	FN	TN	FP
1	Public Transportation	91	33	3.484	15
2	Pedicab	80	15	3.466	18
3	Car	1.281	39	3.201	57
4	Motor	2.143	96	3.088	31
5	Truck	82	16	3.967	8

Table 7. Performance results

No.	Object	Accuracy	Presisi	Recall	F1 Score
1	Public Transportation	98,68%	98,68%	98,68%	98,68%
2	Pedicab	99,08%	99,08%	99,08%	99,08%
3	Car	97,90%	97,90%	97,90%	97,90%
4	Motor	97,63%	97,63%	97,63%	97,63%
5	Truck	99,41%	99,41%	99,41%	99,41%

Table 8. Comparison of Previous Research

Metode	Amount of Data	Accuracy
YOLOv7 [13]	100% data label	mAP accuracy is about 93.7%
YOLOv8 [20]	100% data label	The accuracy of the mAP50 is about 99.1% and the mAP50-95- 83.3%
YOLOv7 + FixMatch	**50% data label, 50% Unlabeled data**	**Accuracy 97,5 %**

Based on Table 8, YOLOv7 with 100% label data achieved an accuracy of about 93.7% [13], while YOLOv8 with complete label data had an mAP50 of about 99.1% and mAP50-95 of 83.3% [20]. However, what is most interesting is that using YOLOv7 with FixMatch, which uses 50% labeled data and 50% unlabeled data, still maintains a high accuracy of 97.5%. This shows that this approach is very effective in reducing the reliance on labeled data while still maintaining good performance and saving time regarding annotation.

CONCLUSIONS

This study proposed a combination of YOLOv7 and FixMatch to detect vehicles in a mixed-traffic environment. Based on the results, the YOLOv7 and FixMatch Models proved effective in detecting different types of vehicles in mixed-traffic environments with high accuracy despite using limited labeled datasets. The models maintained consistent performance under varying lighting conditions and viewing angles, showing potential as a reliable solution for vehicle detection applications. For future research, several aspects need attention. Exploration of other data augmentation techniques, such as MixUp or CutMix, can help improve the robustness of the model. In addition, using more diverse and complex datasets from various locations and weather conditions will strengthen the robustness testing of the model. Given the high computational requirements of the YOLOv7 model, future research could focus on developing a more lightweight and efficient model without sacrificing accuracy, which is essential for real-time applications on resource-constrained devices.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Education, Culture, Research, and Technology, Indonesia, under Applied Research Grant: 683/UN11.2.1/PG.01.03/SPK/DRTPM/2024. The authors thank the research partner of the Department of Transportation, Aceh Province.

REFERENCES

[1] Amit, P. Felzenszwalb, and R. Girshick, “Object Detection,” Comput. Vis., no. January, pp. 1–9, 2020, doi: 10.1007/978-3-030-03243-2_660-1.

[2] R. Padilla, S. L. Netto, and E. A. B. Da Silva, “A Survey on Performance Metrics for Object-Detection Algorithms,” Int. Conf. Syst. Signals, Image Process., vol. 2020-July, no. July, pp. 237–242, 2020, doi: 10.1109/IWSSIP48289.2020.9145130.

[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 580–587, 2014, doi: 10.1109/CVPR.2014.81.

[4] R. Girshick, “Fast R-CNN,” Proc. IEEE Int. Conf. Comput. Vis., vol. 2015 Inter, pp. 1440–1448, 2015, doi: 10.1109/ICCV.2015.169.

[5] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[6] P. K. Mandal, C. Leo, and C. Hurley, “Horizontal Federated Computer Vision,” pp. 1–11, 2023, [Online]. Available: http://arxiv.org/abs/2401.00390

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.

[8] D. Mahajan et al., “Exploring the Limits of Weakly Supervised Pretraining,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11206 LNCS, pp. 185–201, 2018, doi: 10.1007/978-3-030-01216-8_12.

[9] J. Redmon and A. Farhadi, “YOLO9000: Better, faster, stronger,” Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 6517–6525, 2017, doi: 10.1109/CVPR.2017.690.

[10] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” 2018, [Online]. Available: http://arxiv.org/abs/1804.02767

[11] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” 2020, [Online]. Available: http://arxiv.org/abs/2004.10934

[12] C. Li et al., “YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications”.

[13] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors,” pp. 7464–7475, 2023, doi: 10.1109/cvpr52729.2023.00721.

[14] D.-H. Lee, “Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks,” ICML 2013 Work. Challenges Represent. Learn., pp. 1–6, 2013, [Online]. Available: https://www.kaggle.com/blobs/download/forum-message-attachment-files/746/pseudo_label_final.pdf

[15] P. Chen, L. Li, X. Zhang, S. Wang, and A. Tan, “Blind quality index for tone-mapped images based on luminance partition,” Pattern Recognit, vol. 89, pp. 108–118, May 2019, doi: 10.1016/j.patcog.2019.01.010.

[16] P. Cai, Y. Lee, Y. Luo, D. Hsu, and R. O. Mar, “SUMMIT : A Simulator for Urban Driving in Massive Mixed Traffic”.

[17] A. Afdhal, K. Saddami, S. Sugiarto, Z. Fuadi, and N. Nasaruddin, “Real-time object detection performance of yolov8 models for self-driving cars in a mixed traffic environment”. In 2023 2nd International conference on computer system, information technology, and electrical engineering (COSITE) (pp. 260-265). IEEE. 2023.

[18] A. Afdhal, K. Saddami, M. Arief, S. Sugiarto, Z. Fuadi, and N. Nasaruddin, “MXT-YOLOv7t: An Efficient Real-Time Object Detection for Autonomous Driving in Mixed Traffic Environments”. IEEE Access, in press, IEEE. 2024.

[19] M. Devgan, G. Malik, and D. K. Sharma, Semi-Supervised Learning. 2020. doi: 10.1002/9781119654834.ch10.

[20] D. Reis, J. Kupec, J. Hong, and A. Daoudi, “Real-Time Flying Object Detection with YOLOv8,” 2023, [Online]. Available: http://arxiv.org/abs/2305.09972.

 https://doi.org/10.25077/jnte.v13n3.1250.2024