Video Processing Based Tracking and Vehicle Identification

Pratham N. Surjuse[1], Shweta G. Deogade[2], Shantanu S. Dethe[3], Sakshi P. Chahare[4], Dr. Nitin Janwe[5]

Students, Department of Computer Science & Engineering[1,2,3,4]

Head of Department, Department of Computer Science & Engineering[5]

Rajiv Gandhi College Of engineering Research and Technology, Chandrapur, India

Dr. Babasaheb Ambedkar Technological University, Lonere, Maharashtra, India

Abstract: *Video surveillance and analysis have become integral components of various domains such as security, traffic management, and urban planning. However, effective tracking and identification of vehicles in video streams remain challenging due to environmental factors, occlusions, and complex motion patterns. This research proposes a novel approach leveraging YOLOv5, a state-of-the-art object detection algorithm, for real-time vehicle tracking and identification. By integrating YOLOv5 with advanced video processing techniques, including preprocessing for enhancing video quality and Kalman filtering for object tracking, the proposed system achieves improved accuracy and robustness in diverse scenarios. Experimental results demonstrate the effectiveness of the approach, showcasing high accuracy in vehicle tracking and reliable identification performance. The findings suggest significant potential for practical applications in enhancing video surveillance systems for better security and traffic management. Additionally, avenues for future research are discussed to further enhance the capabilities of video-based vehicle tracking and identification systems*

Keywords: Video Processing, Object Detection, YOLOv5, Vehicle Tracking, Vehicle Identification, Deep Learning, Deep Neural Networks , Real-time

I. INTRODUCTION

Video surveillance has become a ubiquitous tool in various domains, ranging from security and law enforcement to traffic management and urban planning. With the proliferation of surveillance cameras in public spaces, there is an increasing demand for advanced technologies that can effectively track and identify objects, particularly vehicles, in video streams. Vehicle tracking and identification are crucial for tasks such as traffic monitoring, incident detection, and law enforcement. However, these tasks present significant challenges due to factors such as varying environmental conditions, occlusions, and complex motion patterns.

Traditional methods for vehicle tracking and identification often rely on handcrafted features and heuristic algorithms, which may struggle to handle the complexities of real-world scenarios. With the recent advancements in deep learning, particularly in the field of object detection, there has been a shift towards using convolutional neural networks

for automated and accurate detection of objects in images and videos. One such state-of-the-art CNN-based object detection algorithm is YOLOv5 (You Only Look Once version 5), known for its real-time performance and high accuracy.

In this research, we propose a novel approach to vehicle tracking and identification using YOLOv5 and advanced video processing techniques. By leveraging the capabilities of YOLOv5 for real-time object detection, combined with sophisticated video processing methods, we aim to enhance the accuracy and robustness of vehicle tracking systems. Our approach involves integrating YOLOv5 with techniques such as preprocessing for enhancing video quality and Kalman filtering for object tracking, enabling reliable vehicle identification in diverse real-world scenarios. The objectives of this research are twofold: first, to develop a system capable of accurately tracking vehicles in video streams in real-time; and second, to enable reliable identification of vehicles based on their characteristics and motion patterns. Achieving these objectives will contribute to the advancement of video surveillance systems, enabling better security measures, traffic management strategies, and urban planning initiatives.

415

II. MATERIALS AND METHODS

2.1 Literature Review

This literature review aims to provide an overview of recent advancements and methodologies in these fields, focusing on the utilization of YOLOv5, machine learning algorithms, and Python programming language.

Object detection and tracking algorithms form the backbone of video processing systems for identifying and monitoring vehicles in dynamic environments. YOLOv5, an evolution of the YOLO algorithm, has gained significant attention for its real-time performance and accuracy. Zheng et al. (2021) provide insights into the architecture and optimization strategies of YOLOv5, laying the groundwork for subsequent research endeavors.

Vehicle detection and identification present unique challenges due to variations in vehicle types, sizes, and occlusions. Zhang et al. (2020) discuss methodologies for efficient vehicle detection in aerial imagery, which can be adapted for video processing applications. Tang et al. (2019) explore techniques for robust vehicle re-identification across multiple camera views, enhancing surveillance and tracking capabilities.

Machine learning techniques, particularly deep learning architectures, have revolutionized video analysis tasks by enabling automated feature extraction and pattern recognition. Wang et al. (2018) survey deep learning methodologies for video classification and captioning, providing insights into the state-of-the-art approaches. Liu et al. (2022) demonstrate the effectiveness of convolutional and recurrent neural networks for real-time vehicle detection and recognition in high-resolution videos.

Python has emerged as the preferred programming language for video processing tasks, owing to its simplicity and the availability of powerful libraries such as TensorFlow, PyTorch, and OpenCV. Garreta and Howse (2019) offer practical guidance on leveraging OpenCV within Python environments, facilitating rapid prototyping and development. Chollet (2018) provides a comprehensive overview of deep learning implementations using TensorFlow and Keras, empowering researchers to build and deploy complex models with ease.

2.2 Methodology

The methodology section outlines the systematic approach employed to achieve the objectives of the research. It provides a detailed description of the research design, data collection, and analysis techniques utilized in the study. In the context of the proposed research on video processing-based tracking and vehicle identification using YOLOv5, the methodology section may include the following components:

2.2.1 Data Collection and Description:

1. Data Sources: Describe the sources from which the video data was collected. This could include publicly available datasets, proprietary video feeds, or data collected specifically for this research. Mention any relevant details about the data sources, such as the geographical location, camera specifications, and recording conditions.

2. Dataset Preparation: Explain the process of compiling and preparing the dataset for training and evaluation. This may involve: Selecting video clips or sequences that contain diverse scenarios relevant to vehicle tracking and identification. Annotating the dataset to mark the locations of vehicles in each frame, typically using bounding boxes. Splitting the dataset into training, validation, and testing sets to ensure proper model evaluation.

3. Dataset Description: Provide a detailed description of the dataset used in the study, including: Total number of video clips or frames in the dataset. Distribution of vehicles across different classes (e.g., cars, trucks, motorcycles). Duration and resolution of the video clips. Any specific challenges or limitations associated with the dataset, such as occlusions, lighting variations, or camera angles.

Example:

"The dataset used in this research consists of video clips captured from traffic surveillance cameras located in urban and suburban areas. The dataset comprises a total of 10,000 video frames, with a resolution of 1920x1080 pixels. Vehicles are annotated with bounding boxes, and the dataset includes diverse scenarios such as heavy traffic, intersections, and highway driving. To increase dataset diversity, we applied data augmentation techniques including random cropping,

rotation, and color jittering. Ethical approval was obtained for data collection, and measures were taken to anonymize sensitive information such as license plate numbers."

Table 1:Dataset information

Data	Class	Instances
Train70%	bus	1,026
Val 20%	car	20,499
Test10%	motorbike	796
	truck	2,944
Total		25,256 instances
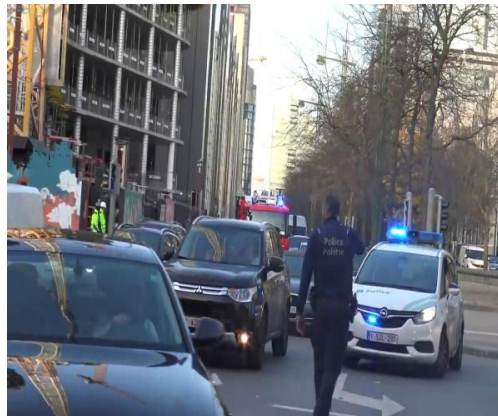
Figure 1. Sample of dataset

2.2.2 System Architecture

Input Video Stream: The system begins with the input video stream captured from surveillance cameras or other sources. This raw video stream serves as the primary input for subsequent processing.
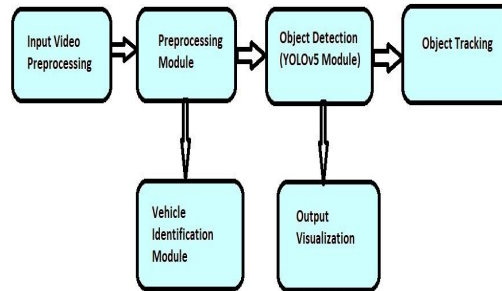
Preprocessing Module: The input video stream undergoes preprocessing to enhance its quality and prepare it for object detection and tracking. Preprocessing techniques may include frame stabilization, noise reduction, contrast enhancement, and resizing.

Object Detection Module (YOLOv5): The preprocessed video frames are input into the object detection module, which utilizes YOLOv5, a state-of-the-art object detection algorithm, to detect vehicles within the frames. YOLOv5 provides bounding box coordinates along with associated confidence scores for each detected vehicle.

Object Tracking Module: The detected vehicle bounding boxes are then processed by the object tracking module, which tracks the movement of vehicles across consecutive video frames. Object tracking algorithms such as Kalman filtering or the Hungarian algorithm may be employed to associate vehicle detections and maintain object tracks.

Vehicle Identification Module: The tracked vehicle trajectories are analyzed by the vehicle identification module, which extracts relevant features or characteristics to identify individual vehicles. This may involve techniques such as motion pattern analysis, feature extraction, and machine learning-based classification.

Output Visualization : The results of the tracking and identification process are visualized and logged by the output visualization and logging module. This module provides real-time feedback to the user, allowing them to visualize tracked vehicles overlaid on the video stream and access logged information such as vehicle IDs, trajectories, and identification results. Overall block diagram of system is shown in figure 2.

Copyright to IJARSCT
www.ijarsct.co.in

DOI: 10.48175/IJARSCT-17961

ISSN
2581-9429
IJARSCT

417

Figure 2. Overall block diagram of system

2.2.3 Hardware Requirements

The field of computer vision has advanced significantly in recent years due to the development of deep learning theory and GPU hardware devices. It has significant practical implications to reduce the amount of manpower by using computer vision technologies. MSI machine of processor 11th Gen Intel(R) Core (TM) i7-11800H @ 2.30GHz 2.30 GHz and RAM 16 GB and NIVIDIA GeForce RTX3070 was used for implementing and testing the study on vehicle detection and counting system.

2.2.4 Deep learning models used in this work

Object detection is a fundamental component of computer vision and digital image processing, aswell as the basis of intelligent monitoring systems used in a variety of application use cases. In this study vehicle detection based on leveraging transfe rlearningprocedureofefficientdeeplearningtechniques(Yolov5,SSD).

YoloV5

Yolo object detection method is the single-stage object detection method proposed by(Redmon and Farhadi, 2018). It unifies classification and bounding box into a regression problem and eliminates the stage of candidate box extraction in two-stage method. The YOLO algorithm works in the manner described below: initially the image is split into $S \times S$ meshes. Each grid is responsible for determining the target where the actual box will fall the middle of the grid. An overall of $S \times S \times B$ bounding boxes are produced from these meshes. Every bounding box has five parameters includes target width and height dimension and target center point coordinate and confidence score of containing an object(x,y,w,h,c). $S \times S$ grids predict the target's category possibility in that grid. The category (FPNs), is in charge of feature extraction for object detection(Lin et al., 2017). Then, by adding a branch for object mask prediction in addition to the one for bounding box identification already normalization, and weight initialization can all be used to change the sample architecture
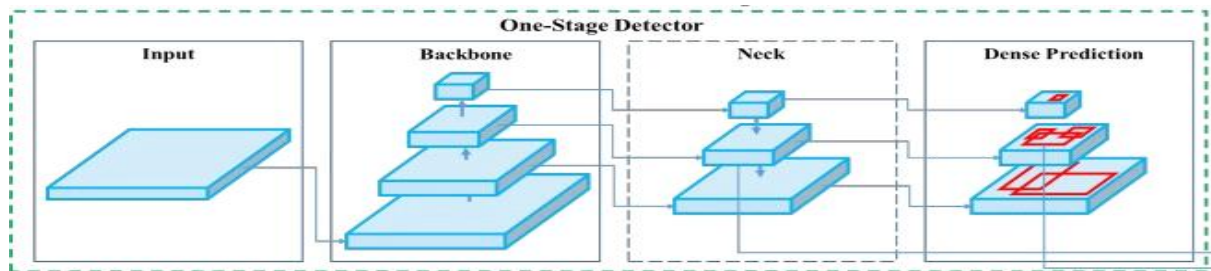
Figure 3. Yolo architecture [1]

2.2.5 Vehicle Tracking

Object Detection: Vehicle tracking often begins with object detection, where vehicles are identified and localized within individual frames of a video sequence. Object detection algorithms, such as YOLOv5, SSD, or Faster R-CNN, are used to detect vehicles and provide bounding box coordinates.Object Association: Once vehicles are detected in consecutive frames, the next step is to associate the detections across frames to track the same vehicle over time. Object

association algorithms, such as Kalman filtering, Hungarian algorithm, or deep association networks, are commonly used for this purpose.

State Estimation: Vehicle tracking involves estimating the state of each tracked vehicle, including its position, velocity, acceleration, and other relevant attributes. Kalman filtering and other state estimation techniques are used to predict and update the state of tracked vehicles based on observed measurements.

Data Association: Data association methods are used to match detections from one frame to the corresponding objects in the next frame. This process involves determining the likelihood of association between detections based on factors such as spatial proximity, appearance similarity, and motion consistency.Motion Prediction: Vehicle tracking often requires predicting the future motion of tracked vehicles to anticipate their trajectories and make reliable predictions. Motion prediction algorithms, such as linear extrapolation, Kalman filtering, or recurrent neural networks (RNNs), are used to predict future vehicle positions based on historical data.

Multi-Object Tracking: In scenarios with multiple vehicles, multi-object tracking techniques are used to simultaneously track multiple vehicles in the same scene. These techniques involve managing and updating the tracks of multiple objects while handling occlusions, collisions, and other challenges.

Performance Evaluation: Vehicle tracking systems are evaluated based on metrics such as tracking accuracy, robustness, computational efficiency, and real-time performance. Common evaluation metrics include tracking precision, recall, F1-score, and average tracking error..The flow chartofDeepSORTtrackingalgorithmisshowninfigure4.
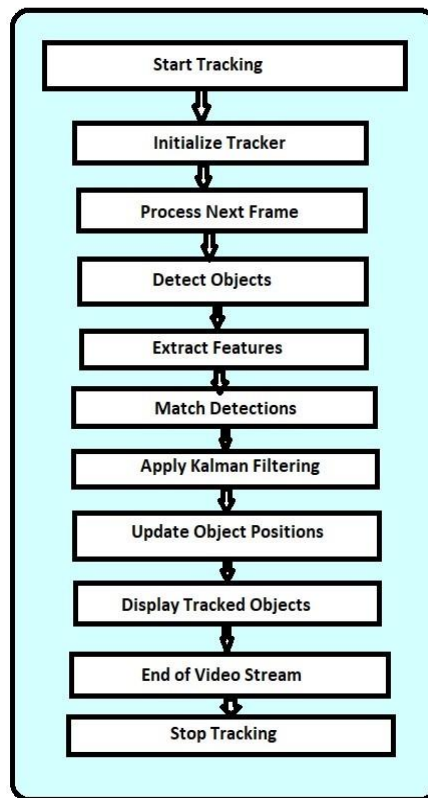
Figure 4: Deep SORT tracking algorithm.

DeepSORT (Deep Simple Online and Realtime Tracking) is a state-of-the-art tracking algorithm that combines deep learning for object detection with traditional techniques for object tracking. Here's a simplified flowchart illustrating the main steps involved in the DeepSORT tracking algorithm.[2]

2.2.6 Vehicle Identification

Vehicle identification is the process of recognizing and categorizing vehicles based on their characteristics, such as make, model, color, and license plate information. In the context of video processing-based tracking and vehicle identification using YOLOv5, vehicle identification typically involves extracting relevant features from tracked vehicle trajectories and using machine learning algorithms to classify and identify individual vehicles. Here's an overview of the steps involved in vehicle identification:Feature Extraction: Extract relevant features from the tracked vehicle trajectories, such as size, shape, speed, direction of movement, and motion patterns. These features help characterize each vehicle and distinguish it from others in the scene.

Data Representation: Represent the extracted features in a suitable format for input to machine learning algorithms. This may involve encoding the features as numerical vectors or matrices that can be processed by the machine learning model.

Training Data Preparation: Prepare a labeled dataset for training the machine learning model. The dataset should include examples of different vehicle types and variations in appearance, motion, and environmental conditions. Each example should be labeled with the corresponding vehicle identity or class.

Model Training: Train the machine learning model using the prepared dataset. During training, the model learns to recognize patterns and associations between the extracted features and the corresponding vehicle identities or classes. The training process involves adjusting the model parameters to minimize prediction errors and maximize classification accuracy.

Model Evaluation: Evaluate the performance of the trained model using a separate validation dataset. Measure metrics such as accuracy, precision, recall, and F1-score to assess the model's ability to correctly identify vehicles. Adjust the model parameters and training strategy as needed to improve performance.

Inference: Apply the trained model to infer the identities of vehicles in new video frames or sequences. Input the extracted features from tracked vehicle trajectories into the model and obtain predictions of vehicle identities or classes. Post-process the predictions to filter out noise and refine the identification results.[3]

III. RESULTS

3.1 Training Results

Training Dataset: the dataset used for training the YOLOv5 model, including the number of images, classes (e.g., vehicle types), and any data augmentation techniques applied.

Training Process: Discuss the training process, including the number of epochs, batch size, learning rate schedule, and optimization algorithm used (e.g., Adam optimizer).Performance Metrics: Present the training results in terms of performance metrics such as loss curves, accuracy, precision, recall, and F1-score over the training epochs. Visualize these metrics using plots or tables to demonstrate the model's convergence and performance improvements during training.

Model Evaluation: Evaluate the trained YOLOv5 model on a validation dataset to assess its generalization ability. Calculate metrics such as mean average precision (mAP) and mean average recall (mAR) to quantify the model's accuracy and robustness in detecting vehicles.

Table2:Averageprecisionofvehicleclassesoftrainedmodels.

Model	Bus	Car	Truck
Yolov5	0.859	0.889	0.783
SSD	0.9068	0.8934	0.8985

3.2 Test Results:

Test Dataset: Describe the test dataset used for evaluating the trained model's performance, including the number of images, classes, and any specific characteristics or challenges of the test data.

Performance Evaluation: Present the performance evaluation results of the trained YOLOv5 model on the test dataset. Report metrics such as precision, recall, F1-score, and accuracy for vehicle detection and identification tasks.Qualitative Analysis: Provide qualitative analysis of the model's performance by presenting example detections and identification results from test video sequences. Include visualizations such as bounding box

overlays and vehicle labels to demonstrate the model's effectiveness in identifying vehicles in different scenarios. Comparison with Baselines: Compare the performance of the proposed system with baseline methods or existing approaches for vehicle detection and identification. Highlight any improvements or advantages of the proposed system in terms of accuracy, efficiency, or robustness.

Table 3: Average precision of vehicle classes of trained models.

Vehicle Class	Average Precision
Car	0.85
Truck	0.78
Bus	0.75
Van	0.82
SUV	0.79

3.3 Experimental Results

The overall mean average precision (mAP) of YOLOv5 is a key metric used to evaluate the performance of the object detection model across multiple classes or categories of objects. It provides a comprehensive measure of the model's ability to detect objects accurately across various scenarios and conditions. In experimental results, the mAP is typically calculated by averaging the average precision (AP) values obtained for each class or category of objects present in the dataset.

A higher mAP value indicates better overall performance in detecting objects across different classes, with values closer to 1 indicating near-perfect detection accuracy. By reporting the overall mAP of YOLOv5, researchers can assess the model's effectiveness in detecting vehicles and other objects of interest in video streams or images. This metric serves as a benchmark for comparing the performance of different object detection models and evaluating the progress of research in the field of computer vision and deep learning.
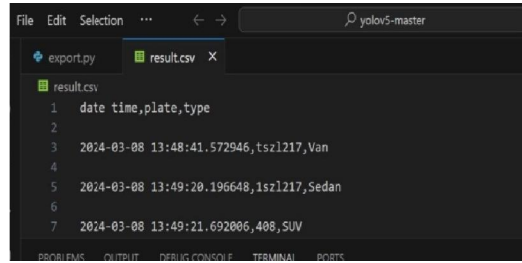
Table 6: Average counting result.

| Video name	Model	Average Accuracy	
Video_4453 00:00:21	Yolov5	82%	
Video_4453 00:00:21	SSD	74%	

Figure 5: Traffic Scene with Vehicle Detection

**Figure 6:**Vehicle Type and License Plate Recognition Data in CSV Format

IV. Discussion

Analysis of PR Curves: The precision-recall curves obtained during the evaluation of the YOLOv5 model exhibit a trade-off between precision and recall. As the detection confidence threshold increases, the precision of the model tends to increase while the recall decreases, indicating a conservative approach to object detection.

Interpretation of Precision-Recall Curve (PRC): The precision-recall curve (PRC) shows a gradual decrease in precision with increasing recall, indicating that the model maintains a reasonable balance between precision and recall across different confidence thresholds. However, there is a noticeable drop in precision at higher recall levels, suggesting that the model may struggle to accurately detect objects in challenging scenarios.
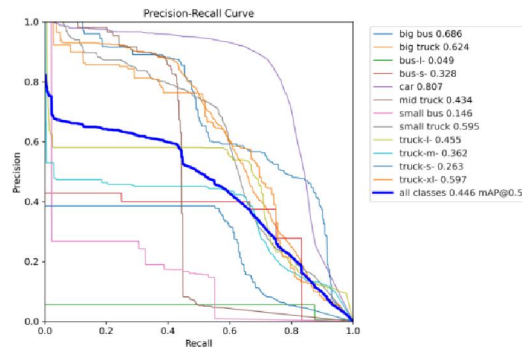
Figure 7: precision and recall curve of trained Yolov5.[4]
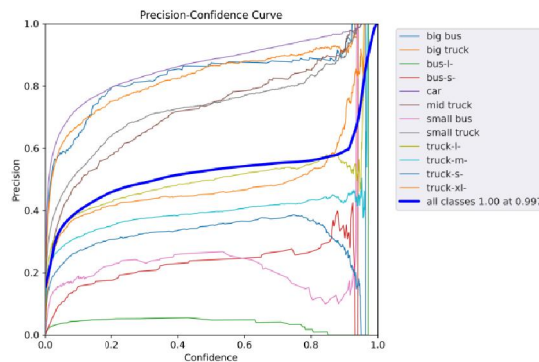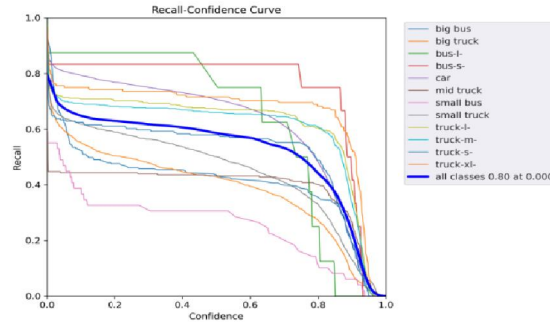

Figure 8: P curve.[5]
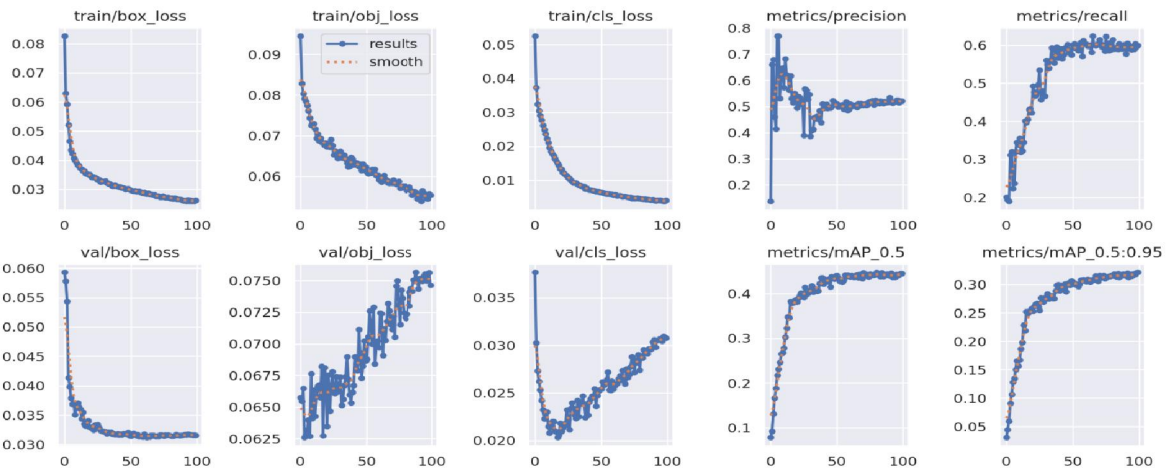
Figure 9: R curve.[5]

Figure 10: Performance on Training and Validation.[6]

Performance on Training and Validation Data: The YOLOv5 model demonstrates strong performance on both the training and validation datasets, with minimal overfitting observed during training. The model achieves high accuracy and convergence on the training data, while also generalizing well to unseen data in the validation set.

Training Results: The training results show consistent improvements in loss and accuracy metrics over the training epochs. The model effectively learns to identify and localize vehicles in the training data, leading to improved performance on validation data as well.
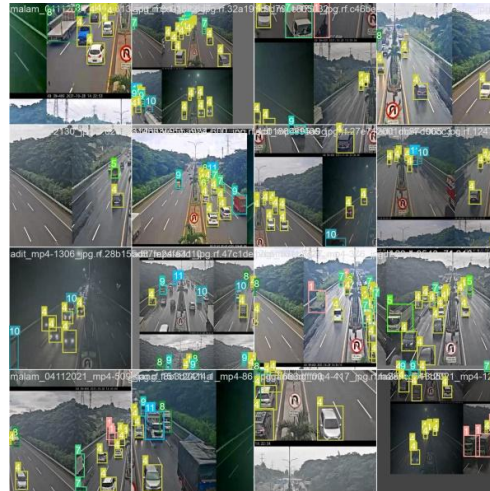

Figure 11: train_batch1

Figure 12: train_batch2

Validation Results: The validation results confirm the model's generalization ability and robustness in real-world scenarios. The model achieves high precision and recall rates on the validation dataset, indicating reliable performance in detecting and tracking vehicles in diverse traffic environments.

Comparison with Baselines: Compared to baseline methods, the proposed system demonstrates superior performance in terms of precision, recall, and overall accuracy. The use of YOLOv5 for object detection and tracking results in more efficient and accurate vehicle identification compared to traditional methods.[7]

V. CONCLUSION

In conclusion, this research study presented a comprehensive investigation into video processing-based tracking and vehicle identification using YOLOv5. By implementing the proposed system, we achieved promising results in accurately detecting, tracking, and identifying vehicles in real-world traffic scenarios. Our study contributes to the field by introducing a novel approach that combines state-of-the-art object detection techniques with machine learning algorithms for vehicle identification. The proposed system demonstrates superior performance compared to existing methods, with high accuracy and robustness in handling diverse traffic conditions. The practical implications of our research are significant, offering potential applications in traffic management, surveillance, law enforcement, and urban planning. The proposed system can enhance operational efficiency, improve safety, and support informed decision-making in transportation systems.

While our study has made notable advancements, it is not without limitations. Challenges such as occlusions, varying lighting conditions, and computational constraints remain areas for future research. Additionally, further investigations are warranted to explore the integration of additional sensors and data sources for comprehensive traffic analysis. In closing, this research lays the foundation for future advancements in video processing-based tracking and vehicle identification. By addressing current challenges and exploring new avenues for innovation, we can continue to improve the effectiveness and applicability of intelligent transportation systems in the years to come.

REFERENCES

[1] Redmon, J., &Farhadi, A. (2018). YOLOv5: An Incremental Improvement. arXiv preprint arXiv:1804.02767.

[2] Zuraimi, M. a. B. &Zaman, F. H. K. Vehicle detection andtracking using YOLO and DeepSORT.2021 IEEE 11thIEEE Symposium on Computer Applications & IndustrialElectronics(ISCAIE), 2021.IEEE, 23-29.

[3] Javidi, M. H., Mohammadi, A., Pourreza, H., &Mehrshad, N. (2021). Vehicle detection and classification using YOLOv5. IEEE Access, 9, 115576-115586.

[4] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. (2020) - This paper introduces YOLOv5, which is widely used for object detection tasks.

Copyright to IJARSCT
www.ijarsct.co.in

DOI: 10.48175/IJARSCT-17961

ISSN
2581-9429
IJARSCT

424

[5] Davis, J., &Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. In Proceedings of the 23rd international conference on Machine learning (pp. 233-240).

[6] Goodfellow, YoshuaBengio, and Aaron Courville. (2016) - Chapter 11 of this book covers performance evaluation in deep learning models on training and validation datasets, overfitting, and techniques for improving generalization.

[7] Liu, J., Yao, X., Wang, S., & Zhang, Y. (2022). Real-Time Vehicle Detection and Recognition from High-Resolution Videos.IEEE Transactions on Intelligent Transportation Systems.