Benchmarking Reidentification in Multi-Camera Tracking Systems with YOLOv8 and ResNet-50

Shaurya Pal
School of Engineering
London South Bank University
London, United Kingdom
pals3@lsbu.ac.uk

Tasos Dagiuklas
School of Engineering
London South Bank University
London, United Kingdom
tdagiuklas@lsbu.ac.uk

***Abstract*—The aim of this paper is to benchmark reidentification within a multi-camera tracking system. This benchmark has been carried out by leveraging transfer learning, utilizing YOLOv8 for real-time object detection and ResNet-50 for feature extraction. The objective is to evaluate the system's performance in accurately reidentifying vehicles across multiple cameras in real-world traffic surveillance scenarios. This benchmarking endeavor aims to provide an evaluation framework for assessing the capabilities and limitations of vehicle reidentification techniques, with a focus on their applicability in challenging conditions such as low-light environments, image compression, and object occlusions.**

***Keywords*—Vehicle Reidentification, Multi-Camera Tracking, Object Detection, Feature Extraction, Transfer Learning, Intelligent Transportation Systems.**

I. INTRODUCTION

The modern world is characterized by ever-increasing mobility, with vehicles serving as the lifeblood of our interconnected societies. Effective traffic management, enhanced security, and optimized urban planning have become imperative to ensure the safety, convenience, and productivity of daily life [1]. In this context, vehicle reidentification has emerged as a critical task within the domain of Intelligent Transportation Systems (ITS) and traffic surveillance, holding immense promise for addressing these multifaceted challenges.

Vehicle reidentification, the process of associating the same vehicle across multiple cameras and timeframes, plays a pivotal role in diverse applications. From mitigating traffic congestion and enhancing security measures to informing urban planning and optimizing public transportation systems, its impact is far-reaching [2][3][4]. The ability to track vehicles consistently across different camera viewpoints and timeframes offers a holistic understanding of vehicle movements and enables responsive actions when necessary [5].

However, real-world traffic surveillance scenarios present formidable challenges, including adverse lighting conditions, image compression, and abrupt changes in vehicle direction. Addressing these challenges necessitates the development of a sophisticated, adaptable, and robust vehicle tracking and reidentification system. Such a system must combine state-of-the-art computer vision techniques with innovative object association methodologies.

As urbanization accelerates and traffic-related concerns intensify, the outcomes of this research work hold substantial potential for traffic management, security enhancement, and urban planning. The fusion of cutting-edge technologies and innovative methodologies signifies a significant step towards realizing smarter, safer, and more efficient cities [9].

In addition to its crucial role in vehicle tracking, reidentification techniques have found extensive applications in the field of human tracking and surveillance. With the ever-growing importance of security and public safety, the ability to accurately identify and track individuals across different cameras has become paramount [10]. Reidentification algorithms designed for human subjects have been employed in various scenarios, from monitoring crowded urban areas to enhancing security at transportation hubs and large-scale events [11]. These applications demonstrate the versatility and adaptability of reidentification techniques beyond vehicles, showcasing their potential to address broader surveillance and tracking challenges in complex real-world environments [12].

This research paper aims to establish a benchmark for vehicle reidentification in a multi-camera tracking system. The benchmark harnesses transfer learning with pretrained models, specifically YOLOv8 for object detection [6] and ResNet-50 for feature extraction [7]. The primary objective is to design, develop, and evaluate a comprehensive multi-camera vehicle tracking and reidentification system. This system endeavors to address critical real-world challenges and complexities, ensuring accurate and reliable vehicle tracking across different cameras [8].

This paper is structured as follows: Section 2 provides a comprehensive literature review on vehicle re-identification. In Section 3, the dataset utilized in this research is described. Section 4 details the methodologies employed, followed by Section 5, which presents the experimentation process. Finally, Section 6 delves into the results of the study.

II. LITERATURE REVIEW

In the realm of intelligent transportation systems (ITS) and surveillance, vehicle reidentification (vehicle re-ID) has emerged as a critical area of research and development. The task of accurately identifying and tracking vehicles across multiple cameras has profound implications for traffic management, security, and urban planning [12]).

Vehicle reidentification, often referred to as "vehicle re-ID," plays an instrumental role in enhancing the efficiency of traffic management and surveillance systems [12]. In their work, Tan et al. (2012) introduced the concept of vehicle re-ID, emphasizing the importance of addressing challenges posed by vehicle appearance variations across cameras.

Deep learning-based methods have garnered substantial attention in recent years due to their capacity to capture intricate vehicle features. YOLO (You Only Look Once), a real-time object detection system, has gained prominence for its ability to efficiently identify and localize vehicles in images [6]. The integration of YOLO into vehicle reidentification pipelines has demonstrated significant improvements in tracking accuracy and speed. This approach aligns with the principles of transfer learning, where models pretrained on large datasets are fine-tuned for the specific task of vehicle re-ID [13].

Feature extraction forms a pivotal component of vehicle reidentification, facilitating the comparison of vehicle attributes. Residual Networks (ResNets) have emerged as prominent architectures for feature extraction [7]. ResNet-50, a 50-layer variant, has exhibited exceptional performance in extracting discriminative features for vehicle matching. By leveraging the high-level features captured by ResNet-50, researchers have achieved remarkable results in the reidentification of vehicles captured by distinct cameras.

The association of vehicles across cameras presents a substantial challenge in multi-camera tracking system. Tracking-by-detection, a prevalent methodology, employs object detection outputs to infer vehicle trajectories [14]. Additionally, data association techniques, such as the Hungarian algorithm, have been adapted to match vehicles across cameras, considering factors such as appearance and location [15].

Recent advancements in vehicle reidentification extend beyond traffic management and surveillance. Human reidentification, a concept inspired by vehicle re-ID, has garnered interest in applications such as public safety and urban planning [16]. The application of vehicle reidentification techniques to human tracking scenarios has shown promise in improving security and monitoring in crowded environments.

While previous works have made significant strides in vehicle reidentification, there exist certain limitations that need to be addressed. Previous methods often struggled with real-world challenges such as variations in lighting conditions, object occlusions, and the need for efficient real-time processing. Additionally, achieving high matching accuracy across cameras remains a complex task.

Considering these limitations, this paper proposes the development of a robust benchmark for vehicle reidentification in multi-camera tracking system. This benchmark aims to evaluate and advance the capabilities of vehicle reidentification techniques, particularly in challenging conditions. By leveraging the strengths of YOLOv8 for real-time object detection and ResNet-50 for feature extraction, this paper seeks to contribute to the ongoing evolution of vehicle reidentification, ultimately improving its applicability in real-world scenarios.

III. DATASET

The dataset was obtained from the AICity Challenge, specifically the AIC22 benchmark, also known as CityFlowV2 [17][18][19][20]. This dataset was captured by 46 surveillance cameras in a real-world traffic surveillance environment and contains annotations for a total of 880 vehicles across six different scenarios. The dataset is divided into training, validation, and testing sets, with a total of 215.03 minutes of video footage. The training set consists of 58.43 minutes of videos, the validation set contains 136.60 minutes, and the testing set includes 20.00 minutes of video data.

The dataset directory structure significantly enhances its utility for research and development by meticulously organizing subsets for training, validation, and testing, tailoring each to specific camera configurations and accompanying them with meticulous annotations of ground truths. Researchers can effectively leverage the provided videos to conduct dynamic analyses of vehicle movements, allowing for in-depth studies of traffic patterns and behaviours.

The inclusion of manual calibration results is a valuable addition, offering insights into the geometric calibration of cameras. This information contributes to more accurate tracking and evaluation, ensuring that algorithms can account for any distortions or variations in the camera setup.

To further facilitate research and ensure reproducibility, the dataset provides comprehensive information on camera locations, timestamps, and frame numbers. The precise GPS locations for key scenarios offer a crucial spatial context, enhancing the understanding of the real-world deployment of cameras. Additionally, detailed maps depicting camera locations are included, providing a visual reference for researchers to correlate with their findings and optimize their methodologies for multi-camera tracking scenarios.

For this work, only the testing portion of the dataset was utilized, and data from the first two cameras, specifically cameras 41 and 42, have been selected for experimentation. These cameras have been chosen to represent a real-world multi-camera tracking scenario.

The decision to use cameras 41 and 42 from the dataset was made with practical considerations in mind. Working with a smaller subset of cameras was more resource-efficient, given the computational demands of processing data from all six cameras. Cameras 41 and 42 have been selected to create a scenario representative of a real-world multi-camera tracking setup, aligning with the paper's objectives. This choice allowed for an initial exploration of methodologies and algorithms while maintaining a manageable scope. The testing portion of the dataset was chosen to ensure consistent evaluation.

The reason this dataset aligns perfectly with the chosen methodology lies in the temporal coherence of the video sequences. The videos captured by the cameras are temporally synchronized, meaning that the vehicles passing through the field of view of one camera are subsequently recorded by another camera. This temporal alignment is crucial for the methodology, ensuring that vehicles tracked through one camera can be consistently traced across multiple cameras. Moreover, the utilization of transfer learning models, such as YOLOv8 and ResNet-50, is well-suited for the dataset's characteristics. YOLOv8 excels in detecting objects in continuous video frames, and the extraction capabilities of ResNet-50 align seamlessly with the dataset's rich and diverse annotations. This aligns seamlessly with the proposed methodology, which involves matching objects across cameras and associating them based on human validation. The dataset's structure and content make it an ideal choice for assessing the effectiveness of the proposed system in a real-world, multi-camera surveillance setting. Figure 1 illustrates the camera setup in the AIC22 dataset, showcasing the surveillance cameras used for data collection.

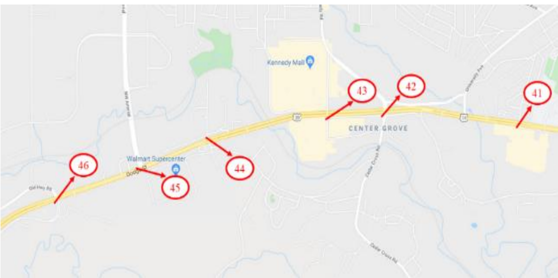

Figure 1:Camera Setup

IV. METHODOLOGY

The methodology for the multi-camera vehicle tracking and reidentification paper is organized into several key components and stages, each serving a specific purpose within the system's architecture.

1. Input Sources:

The research initiative commences by selecting video feeds from two distinct surveillance cameras, denoted as "Camera 41" and "Camera 42." These chosen sources serve as the primary inputs for the subsequent stages of the study, carefully curated to ensure a controlled and focused dataset. This deliberate selection is foundational to fostering a robust analytical framework for the study.

2. Object Detection (YOLOv8):

Employing the state-of-the-art YOLOv8 model, both "Camera 41 Video" and "Camera 42 Video" undergo an intricate process of object detection. This phase transcends conventional identification, incorporating real-time detection of vehicles with an emphasis on class determination. YOLOv8 further contributes by segmenting and storing objects in each frame based on their respective classes. The outcome is an intricately organized repository of cropped images, stratified by class—a strategic preparation for subsequent stages including feature extraction and matching.

3. Feature Extraction (ResNet-50):

Post-identification, the study harnesses the formidable ResNet-50 model for the extraction of salient features. Cropped images of identified vehicles from both cameras undergo meticulous processing to extract rich and discriminative features. These features play a pivotal role in the subsequent matching stage, elevating the system's discernment in accurately pairing and reidentifying vehicles across different perspectives.

4. Matching (Similarity > 90%):

The matching stage unfolds through a systematic comparison of the extracted features from both cameras. A stringent similarity threshold, set at 90% and above, rigorously filters highly similar features, signifying potential matches. Vehicles surpassing this threshold are designated as potential matches, indicating their likely appearance across divergent camera viewpoints.

5. Query Systems:

To refine and validate the matching results, two query systems are integrated:

a. Query System 1 (Object Association):

Confirmed matches from the preceding step traverse through Query System 1, introducing a human-in-the-loop methodology. This query system empowers users to interact and scrutinize matched objects, providing an essential layer of human validation to augment overall accuracy.

b. Query System 2 (Assign Unique IDs):

Objects affirmed as identical in Query System 1 proceed to Query System 2. In this phase, unique identification (ID) numbers are meticulously assigned to the matched objects. This critical step ensures the consistent tracking of vehicles across disparate cameras, constituting a fundamental aspect of the reidentification process.

6. Final ID Assignment:

Objects endowed with unique IDs represent the culmination of the study's endeavor, symbolizing the successful tracking and reidentification of vehicles. This final assignment underscores the system's efficacy in associating vehicles reliably across multiple cameras—an achievement of paramount significance within the context of the study.

In synthesis, Figures 2 and 3 visually articulate the systematic workflow of the study, highlighting the integration of cutting-edge models such as YOLOv8 and ResNet-50. Furthermore, the incorporation of user feedback through query systems imparts a layer of precision, ensuring accurate outcomes in the intricate realm of real-world, multi-camera tracking scenarios. The meticulous design of the methodology, coupled with advanced techniques, positions this study to make a substantive contribution to the field.

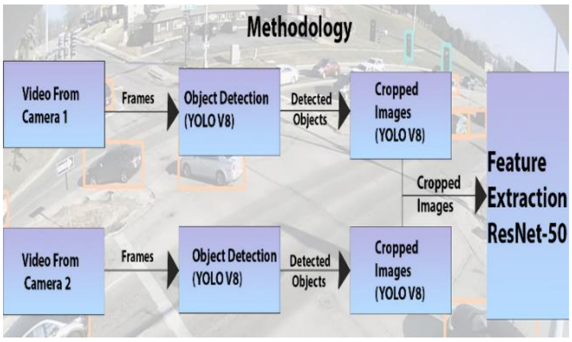

Figure 2:Re-ID Workflow

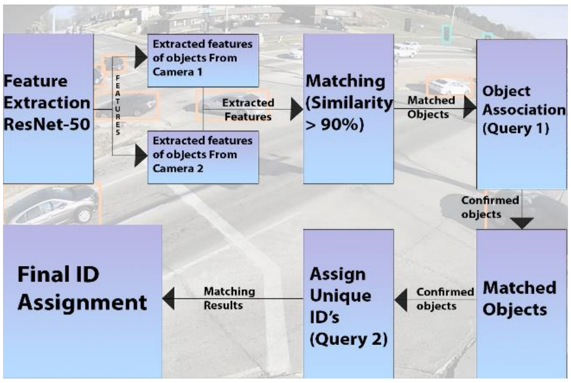

Figure 3: Re-ID Workflow continuation

V. Experimentation

1. Single Camera tracking:

For the single-camera tracking experiment, we harnessed the power of the YOLOv8 model independently on both "Camera 41" and "Camera 42" video feeds. Figure 4 provides a visual representation of this process, showcasing a frame extracted from "Camera 41" with detected vehicles denoted by bounding boxes. These bounding boxes not only outline the vehicles but also indicate their associated classes and unique IDs. As the YOLOv8 model efficiently tracks objects in each camera's feed, the detected vehicles are then cropped and saved based on their specific classes. This step lays the foundation for subsequent feature extraction, enabling further analysis and reidentification procedures.

Figure 4: Single Camera Tracking

2. Feature Extraction:

In this phase, the ResNet-50 model has been used to extract crucial features from images containing detected vehicles. These features are represented as sequences of numbers, capturing intricate details about the texture, shape, and distinctive attributes of the objects. Features are considered as unique fingerprints allowing us to differentiate one vehicle from another, even in challenging situations. Interestingly, the dataset provided calibration information, which we initially considered using to correct potential distortions in the vehicle images. Features serve as the cornerstone of the ability to distinguish and track objects effectively in a multi-camera environment.

3. Object Association:

In our pursuit of comprehensive tracking and analysis, associated objects are detected in both camera feeds by matching features extracted using ResNet-50. We extracted relevant object features, by calculating a similarity index, and applied a strict 90% similarity threshold for association. This process ensured precise tracking, focusing specifically on trucks, which was reflected in Figure 5, titled "Matched Vehicles," showcasing successful matching and tracking of trucks across different camera perspectives.

During this experimental phase, the primary goal was to establish the feasibility of object tracking across different camera perspectives. To achieve this, we devised a process for feature matching, utilizing a similarity index derived from ResNet-50 feature vectors. Notably, at this point, we have not yet calculated accuracy metrics for tracking performance.

Truck1	Truck2
c041_1001.jpg	c042_1191.jpg
c041_10685.jpg	c042_147.jpg
c041_1082.jpg	c042_14302.jpg
c041_1082.jpg	c042_1539.jpg
c041_1082.jpg	c042_1541.jpg
c041_1082.jpg	c042_1927.jpg
c041_1082.jpg	c042_1956.jpg
c041_1092.jpg	c042_1927.jpg
c041_1093.jpg	c042_1926.jpg
c041_1093.jpg	c042_1927.jpg
c041_1093.jpg	c042_1928.jpg
c041_1094.jpg	c042_1927.jpg
c041_1095.jpg	c042_1927.jpg
c041_1095.jpg	c042_1928.jpg
c041_11173.jpg	c042_1274.jpg
c041_11173.jpg	c042_1277.jpg
c041_11173.jpg	c042_869.jpg

Figure 5: Matched Vehicles

4. Query 1:

To enhance the accuracy of object matching across camera feeds, we introduced a Query and Feedback System, which played a pivotal role in displaying matched objects and gathering feedback to validate associations. Here's how it operated:

a. Matched Object Display:

Successfully matched objects based on feature similarity are presented to users, allowing them to compare objects from two camera perspectives.

b. Feedback Mechanism:

A feedback system prompted users to confirm whether the matched objects were indeed the same vehicle, with binary responses ("match" or "no match").

c. Learning and Validation:

User feedback was crucial for system validation and learning. It helped verify associations and contributed to improve matching accuracy over time. Figure 6 illustrates the functioning of the query system in action.

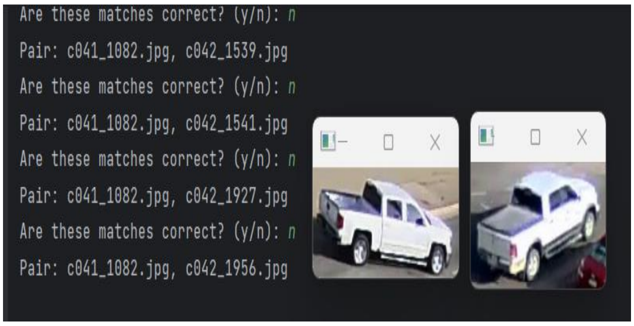

Figure 6: User Feedback

5. Query 2:

To ensure consistent tracking of vehicles across multiple frames within a video, we implemented a re-identification process following the initial object matching. This step aims to assign consistent identifiers to vehicles detected across different frames. The re-identification process is the following:

a. Matched Object Display:

Initially, the system displays the objects that are matched during the previous matching process, allowing users to view the matched objects requiring re-identification.

b. User Feedback:

A feedback mechanism is used to collect user input. Users are asked to confirm whether the displayed objects are indeed the same vehicle. This feedback aims to establish a consensus on object continuity.

c. Reassignment of IDs:

Based on the feedback received, the system reassigned consistent identifiers (IDs) to the matched objects. Objects confirmed to be the same vehicle were given identical IDs.

Vehicle re-identification holds significant implications for video tracking systems. It ensures consistent and accurate tracking of vehicles across frames in a video sequence, enhancing data continuity and accuracy. By assigning consistent identifiers and incorporating user feedback, it minimizes tracking errors, making it valuable for security, surveillance, traffic analysis, and urban planning applications. Figure 7 showcases the assigned IDs for matched objects in action.

Truck1	Truck2	AssignedID
:041_1082.jpg	c042_1956.jpg	1
:041_1414.jpg	c042_1948.jpg	2
:041_1414.jpg	c042_1956.jpg	3
:041_1424.jpg	c042_1948.jpg	4
:041_1424.jpg	c042_1956.jpg	4
:041_1455.jpg	c042_881.jpg	5
:041_1455.jpg	c042_882.jpg	6
:041_1456.jpg	c042_881.jpg	7
:041_1456.jpg	c042_882.jpg	7
:041_1456.jpg	c042_884.jpg	8
:041_1456.jpg	c042_885.jpg	7
:041_1457.jpg	c042_882.jpg	9
:041_1458.jpg	c042_881.jpg	10
:041_1458.jpg	c042_882.jpg	10
:041_1458.jpg	c042_885.jpg	10
:041_1460.jpg	c042_882.jpg	11

Figure 7: Assigned Id's

6. Traffic Monitoring and Management:

YOLOv8's advanced object tracking functionalities allowed for the generation of object tracks in a video. These object tracks effectively represent the movements of detected vehicles. To provide a clear snapshot of traffic patterns, a 30-second timeframe has been selected, and the tracks of objects within this window were visualized. This visualization, as demonstrated in Figure 8, offers insights into traffic flow and congestion points. Authorities and traffic controllers could utilize this information for real-time traffic management, making informed decisions to enhance road safety and traffic efficiency.

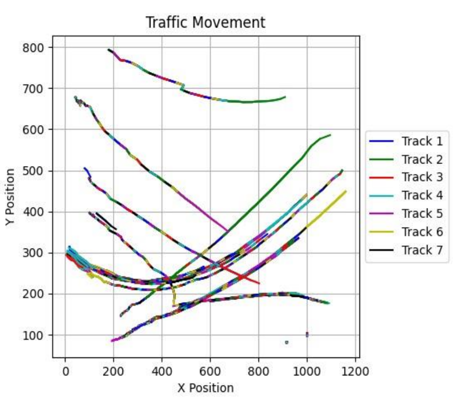

Figure 8: Traffic Plot

7. Assessing Model Robustness Through Data Augmentation:

To evaluate the robustness of the YOLOv8 model, we conducted experiments using data augmentation techniques. These techniques simulate real-world scenarios with varying conditions, such as changes in lighting and image compression. In the data augmentation process, we applied various modifications to a video from camera 41. These modifications introduce challenges like altered lighting conditions and image quality, as shown in Figure 9 ("Video frame after data augmentation"). Subsequently, we ran the object detection process on the augmented video to assess how well the model could identify and track objects under these challenging conditions. The performance comparison between object detection before and after data augmentation offered insights into the model's robustness. Specifically, we examined its ability to handle varying lighting conditions and image compression, both of which are critical factors in real-world applications where environmental conditions can fluctuate significantly. The results of this analysis, presented in the following section, provide valuable information for deploying the YOLOv8 model in practical scenarios.

Figure 9: Video Frame after data augmentation

VI. RESULTS

1. Vehicle Reidentification Evaluation:

In the evaluation of vehicle reidentification, the results highlight the discrepancy between the number of objects initially matched by the system using a 90% similarity threshold and the actual matches confirmed through user input. It's essential to clarify that this analysis specifically concentrated on vehicles detected in both camera 41 and camera 42 as matched objects.

Following user feedback, it became evident that only 16% of vehicles have been accurately matched. Notably, this matching process occurred on a frame-by-frame basis, resulting in a substantial number of mismatches. Many vehicles displayed similar characteristics but were moving in different directions, leading to erroneous associations. Consequently, all vehicles present in both video 1 and video 2 were included in the matching process, regardless of their travel direction.

2. Data Augmentation Results:

This section examines how YOLOv8, responds to challenging conditions like poor lighting and image compression. We focus on changes in detection counts after data augmentation, emphasizing the importance of robust models in real-world scenarios. The detail results are as follows.

a. **New Classes Introduced:**

Data augmentation introduces new classes in YOLOv8, indicating its struggle to adapt to significant changes in objects' appearance due to lighting variations and image compression.

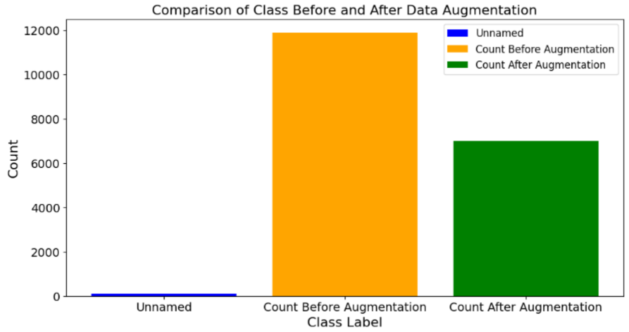

Figure 10: Class Count Comparison

b. **Reduced Counts for Specific Classes:**

YOLOv8 exhibits reduced counts for certain classes (e.g., Car and Truck) after data augmentation, revealing difficulties in detecting these objects accurately under adverse conditions.

c. **Misclassifications:**

Augmentation leads to misclassifications, especially for classes like Bicycle and Motorbike, as poor lighting blurs object boundaries and image compression introduces noise.

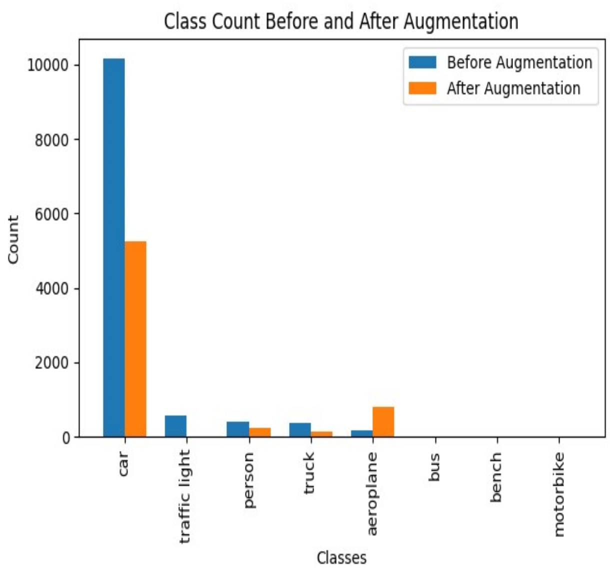

Figure 11: Increased Misclassification

d. **Challenges in Real-World Applications:**

These findings underscore the difficulties YOLOv8, and similar models may face in real-world scenarios with variable lighting and image quality. However, it's important to note that YOLOv8 represents a powerful and robust model, with potential for improvement in handling such challenges in future iterations.

e. **Continuous Advancements in YOLOv8:**

YOLOv8 is currently among the most potent models available, yet it still faces certain limitations, especially in scenarios with suboptimal lighting. However, ongoing advancements in YOLOv8 and its subsequent versions may address these challenges, making it even more effective in real-world applications.

VII. Conclusion

In this benchmarking endeavour of multi-camera vehicle tracking and reidentification we navigated the complex landscape. Leveraging the robust YOLOv8 model for object detection and ResNet-50 for feature extraction, we meticulously evaluated the precision and adaptability of these models. The paper unveiled crucial insights into the realm of vehicle reidentification and provided a benchmark for assessing model performance. Our findings underscore the potential and challenges of employing YOLOv8 in real-world scenarios, paving the way for future enhancements in traffic surveillance and management.

References

[1] Song, W., Shen, J., Lin, J., & Song, M. (2019). Deep multi-sensor vehicle detection and tracking for autonomous driving: datasets, sensors, and algorithms. arXiv preprint arXiv:1902.07857.

[2] Gao, J., Yang, Y., Li, C., & Hospedales, T. M. (2017). Vehicle re-identification with the space-time prior. Proceedings of the IEEE International Conference on Computer Vision.

[3] Li, X., Liu, H., & Wang, L. (2019). Vehicle re-identification with location-semantic attention. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

[4] Khan, S., Hayat, M., Bennamoun, M., Sohel, F., & Togneri, R. (2020). A survey of recent trends in one-shot learning. Artificial Intelligence Review, 53(6), 4029-4074.

[5] Cheng, H., Li, Z., & Gong, D. (2017). Learning to track: Online multi-object tracking by decision making. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[6] Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767.

[7] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition.

[8] Wu, Y., Lim, J. J., & Yang, M. H. (2018). Towards real-time vehicle detection in unmanned aerial vehicle imagery. Proceedings of the European conference on computer vision (ECCV).

[9] Zheng, M., Jiao, J., Shah, S. A., & Liu, J. (2019). Vehicle re-identification in urban surveillance by exploiting temporal information. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

[10] Zheng, Y., Wang, Y., & Yuan, J. (2017). Person re-identification in the wild. In Proceedings of the IEEE International Conference on Computer Vision.

[11] Gheissari, N., Sebastian, T. B., & Hartley, R. (2006). Person re-identification using spatiotemporal appearance. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition.

[12] Tan, X., Chen, S., Zhou, Z. H., Zhang, F., & Liu, S. (2012). Vehicle re-identification by signature based on Fisher vector encoding. In European conference on computer vision (ECCV) (pp. 201-214).

[13] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 248-255). IEEE.

[14] Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking with a deep association metric. 2016 IEEE International Conference on Image Processing (ICIP) (pp. 3464-3468).

[15] Kuhn, H. W. (1955). The Hungarian method for the assignment problem. Naval Research Logistics Quarterly, 2(1-2), 83-97.

[16] Wang, X., Doretto, G., Sebastian, T. B., Rittscher, J., & Tu, P. (2014). People re-identification by spectral classification of silhouettes. European Conference on Computer Vision (ECCV) (pp. 26-41).

[17] Tang, Z., Naphade, M., Liu, M., Yang, X., Birchfield, S., Wang, S., & Hwang, J. (2019). CityFlow: A city-scale benchmark for multi-

target multi-camera vehicle tracking and re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 8797-8806).

[18] Naphade, M., Tang, Z., Chang, M. C., Anastasiu, D. C., Sharma, A., Chellappa, R., & Lyu, S. (2019). The 2019 AI City Challenge. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (pp. 452-460).

[19] Naphade, M., Wang, S., Anastasiu, D. C., Tang, Z., Chang, M. C., Yang, X., & Chakraborty, P. (2020). The 4th AI City Challenge. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.

[20] Naphade, M., Wang, S., Anastasiu, D. C., Tang, Z., Chang, M. C., Yang, X., & Sclaroff, S. (2021). The 5th AI City Challenge. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (pp. 4263- 4273)